How Does Competition Affect Reputation Concerns? Theory and Evidence from Airbnb*

Rossi, Michelangelo†

March 13, 2019

PRELIMINARY AND INCOMPLETE: please do not circulate

Abstract

In this work, I show the existence of a trade-off between the extent of competition faced by Airbnb hosts and the effort they exert to build their reputation. To guide my empirical analysis, I develop a model of reputation inside a directed search framework, where the relative number of hosts and guests impacts the reputation return of hosts' effort. I verify this trade-off using a unique dataset of all Airbnb hosts present in San Francisco from May 2015 to December 2018. I consider the registration enforcement of Airbnb hosts effective since September 2017 in San Francisco as an exclusion restriction that affects the number of hosts active on the platform. With a shock-based IV design, I identify a negative and significant causal relationship between the number of competitors and host's effort. I reinforce my results with an estimation of hosts' effort taking advantage of the multiple tasks of the Airbnb review system and exploiting information about hosts' behavior.

Keywords: Reputation, Competition, Sharing Economy, Airbnb

JEL codes: D82, D83, L50, L81

*I am grateful to my advisors, Natalia Fabra and Matilde Pinto Machado, for their patient guidance and advice. I also thank Victor Aguirregabiria, Heski Bar-Isaac, Jesús Carro, Ambarish Chandra, Rahul Deb, Alberto Galasso, Ignatius J. Horstmann, Luigi Minale, Matthew Mitchell, Jan Stuhler and seminar participants at Universidad Carlos III de Madrid, and Rotman School of Management for helpful suggestions. Special gratitude to Elizaveta Pronkina for her enormous support and encouragement. All errors are mine.

†Universidad Carlos III, Department of Economics, Madrid, Spain. E-mail: mrossi@eco.uc3m.es.

1 Introduction

The rise of digital platforms such as Airbnb, Uber, and TaskRabbit has had a dramatic impact on many industries favoring the entry of new suppliers. Airbnb hosts are an example of service providers that recently entered the hospitality market thanks to the lower entry costs. The abatement of such costs is partially due to the absence of formal regulations that public authorities impose to traditional lodging services. With lax regulatory frameworks, digital platforms mainly rely on review systems to discipline users' behaviors and ensure the quality of their service. Reviews reveal past performances and form users' reputation. Thus, users' concerns for a good reputation are one of the key ingredients for the success of digital marketplaces and the quality of online transactions. Still, reputation concerns may endogenously depend on the structure of the market. In particular, the number of competitors could affect the users' willingness to exert effort and build their reputation since competition shapes the size of reputation returns. Yet, the relationship between competition and reputation concerns has been partially neglected by theory and empirical works.

Understanding how competition affects reputation concerns is a relevant question for policy recommendations regarding digital platforms. In recent years, many public authorities proposed to regulate online services imposing formal registers and restrictions. These policies are likely to change platforms' market structure and, potentially, the quality of services. My work sheds light over the latter connection and it clarifies the role of feedback systems to discipline agents' behavior in competitive settings.

In this work, I show the existence of a trade-off between the extent of competition faced by Airbnb hosts and the effort they exert to create a good reputation. To guide my empirical analysis, I develop a model of reputation in which the relative number of hosts and guests on the two sides of the market (the market tightness) impacts the reputation return of hosts' effort. The main testable prediction of the model regards variations in entry costs and their effects on hosts' effort. When entry costs increase, the number of new entrant hosts lowers, and the market tightness turns in favor of the host side. In equilibrium, this leads to more effort exerted by hosts.

In the empirical part of my work, I verify the relevance of the negative relationship between competition and reputation concerns established in the model. I use a unique dataset regarding all Airbnb hosts present in San Francisco from May 2015 to December 2018. With a shock-based IV design, I identify a negative and significant causal relationship between the number of active listings surrounding each host and host's ratings regarding effort. The exclusion restriction supporting my identification strategy comes from the change in regulation regarding the registration enforcement of Airbnb hosts signed in May 2017 and effective since September 2017 in San Francisco. The identification design closely follows the reasoning supporting the theoretical prediction. An increase in entry costs (registration) leads to a change in the number of hosts on the platforms. Reputation

incentives to exert efforts jump and a significant variation in hosts' ratings regarding effort is observed. I corroborate these findings with an estimation of hosts' effort, taking advantage of the multiple tasks of the Airbnb review system. Applying the same shock-based design, I continue to identify a negative and significant causal relationship between the number of active listings and the estimated host's effort.

The theoretical literature about reputation has seldom investigated the relationship between competition and users' incentives to exert effort. From the early start of this literature, reputation models have mainly studied the effort choices of a long-lived monopoly seller who meets short-lived buyers in every period. For a comprehensive review of the theoretical literature regarding reputation, see Bar-Isaac and Tadelis (2008). Few papers study reputation in competitive contexts, e.g. Hörner (2002). To the best of my knowledge, only Bar-Isaac (2005) explicitly investigates how variations over the extent of competition faced by agents affect their incentives to exert effort. In his analysis, competition has non-monotonic effects on the propensity to exert effort. A higher degree of competition may help to discipline agents, and, at the same time, it may erode reputational premia. In my work, the latter force outweighs the other since a reduction in the number of competitors does not change the disciplining force of the market. Relative to Bar-Isaac (2005), the extent of competition in my model is endogenously determined and it depends on hosts' entry costs. Moreover, I consider a directed search framework where the matching between hosts and guests is frictional. Up to my knowledge, reputation concerns have never been modeled inside a directed search framework. Accordingly, my model contributes to the existing literature of directed search with asymmetries of information. Wright et al. (2017) provide a comprehensive review of the most recent development on this topic. Fradkin (2017) applies a directed search framework to the Airbnb context to study the efficiency loss due to the frictions of guests' search. With respect to his framework, I introduce the hosts' decision about effort and the related reputation concerns.

The empirical part of my work contributes to the literature of digital platforms. Some recent studies analyze users' performances in several digital platforms. Klein et al. (2016) and Hui et al. (2016) take advantage of a variation in the eBay review system implemented in 2008 to study changes in eBay sellers' performance. The modification in the review system removed the potential bias of feedback due to the buyers' fear of retaliation. While Klein et al. (2016) claim that it improved market transparency and induced a disciplining effect on sellers' behavior, Hui et al. (2016) attribute the improvement to sellers' selection. Differently, Meng et al. (2018) investigate Uber drivers' effort studying driver routing choices. They compare drivers' choices with the ones chosen by traditional taxis and observe significant differences. In the Airbnb setting, Proserpio et al. (2018) show that members' reciprocity is at place and users can induce others to behave well by exerting more effort themselves. Still, up to my knowledge, I am the first to estimate users' changes in performance due to movements in the degree of competition they face inside the platform. Accordingly, my

empirical analysis also contributes to the literature on the relationship between competition and market outcomes in general. My identification strategy shares the approach of Dafny et al. (2012), Ashenfelter et al. (2015), and Chandra and Weinberg (2017). These papers study the relationship between market concentration and firms' behavior taking advantage of variations due to mergers. In line with their analyses, I propose a shock-based IV design to break the potential endogeneity of the relationship between the number of competitors and hosts' effort. My exclusion restriction is the registration enforcement of Airbnb hosts in September 2017 in San Francisco. As an instrumental variable, I propose the proportion of hosts that already display a license before September 2017 to predict the number of hosts that will remain on the platform after the registration enforcement.

The paper proceeds as follows. Section 2 describes the theoretical model and the testable predictions. In Section 3, first I provide some background context regarding Airbnb. Then, I illustrate the change in the institutional setting regarding Airbnb hosts regulation in San Francisco in September 2017. Finally, I present the dataset. I discuss my identification strategy in Section 4. Section 5 provides the empirical findings. I proceed with the estimation of hosts' effort and robustness check in Section 6. Section 7 concludes. All the proofs and additional tables are in Appendix.

2 Model

In this Section, I present the theoretical framework for my analysis. First, I describe the model environment. I show the agents' characteristics and payoffs; and I clarify the role of frictions with the assumptions regarding the matching function. Then, I present the timing of agents' interactions and the equilibrium concept. Finally, I characterize the equilibrium allocation and propose the main testable predictions of the model. All proofs are in Appendix.

2.1 Model Environment

The market lasts two periods. Hosts and guests populate the two sides of the market. Each guest (he) is willing to rent a house, whereas each host (she) owns a house and can rent it to one guest only.

In both periods there is an infinite population of hosts who can potentially enter the market. Hosts who enter in the first period stay in the market until the second period. To enter the market, hosts pay entry costs, f, in both periods. In each period, hosts post price p, and, in case of a match with a guest, decide whether to exert effort or not: $e = \{0, 1\}$. Hosts' cost of effort, c, is realized if a host is matched. The cost can take two values: $c = \{0, k\}$ with $k > 0$. Hosts draw $c = 0$ with

probability π. The cost is host private information, whereas the probability π is common knowledge for hosts and guests.

A unit mass of guests is present in the market in period 1 and measure G is present in period 2. Guests are homogeneous and the gross utility from a transaction, u, depends on the host effort and the price: $u = ae + b - p$, with $a, b > 0$ where b represents the benchmark utility that guests obtain from a transaction when hosts do not exert effort. The ex-post surplus of a transaction is defined by the sum of guest's utility and hosts' profit. If the host exerts effort, $e = 1$, the ex-post surplus is $(a + b - p) + (p - c) = a + b - c$. If the host does not exert effort, $e = 0$, the ex-post surplus is $(b-p)+p = b$. In order to guarantee the efficiency of exerting effort $e = 1$, I assume that $a - c > 0$ and that hosts always exert effort $e = 1$ if they draw $c = 0$.

The matching process between hosts and guests is frictional. In line with the directed competitive search literature, market frictions are captured by a matching function M. With a measure h of hosts and g of guests present in the market, a measure $M(h, g) \leq \min(h, g)$ of matches is formed. Assuming constant returns to scale in the matching function, the agents' probability of having a transaction can be determined as a function of the ratio between guests and hosts, denoted as the market tightness: $\theta = \frac{g}{h}$.

The probability of having a transaction for hosts when the market tightness is θ is defined as $\alpha(\theta) \equiv \frac{M(h,g)}{h}$. The probability of having a transaction for guests when the market tightness is θ is defined as $\frac{\alpha(\theta)}{\theta} \equiv \frac{M(h,g)}{g}$. I impose the following conditions on the function $\alpha(\theta)$:

Assumption 1. *For all $\theta \in [0, \infty)$:*

1. $\alpha(\theta) \in [0, 1]$ *and* $\frac{\alpha(\theta)}{\theta} \in [0, 1]$;
2. $\alpha(\theta)$ *is continuous, strictly increasing, twice differentiable, and strictly concave;*
3. $\alpha(\theta) - \theta\alpha'(\theta) > 0$;
4. $\alpha(\infty) = \alpha'(0) = 1$;
5. $\alpha(0) = \lim_{\theta\to\infty} \theta\alpha'(\theta) = 0$.

Assumption 1 is standard in the directed search literature.[1] In particular, $\alpha'(\theta) > 0$ and $\alpha(\theta) - \theta\alpha'(\theta) > 0$ state that, when the number of guests over hosts increases, the host matching probability strictly increases and the guest matching probability strictly decreases.

The expected payoffs of hosts and guests can be defined in terms of the host decisions regarding prices and effort, and the probability of having a transaction. In each period, the expected profits

[1] Delacroix and Shi (2013) and Shi and Delacroix (2018) extensively discuss the class of matching functions satisfying Assumption 1.

for hosts is:

$$\Pi = (p - ce)\alpha(\theta);$$

whereas the expected utility for guests is:

$$U = (ae + b - p)\frac{\alpha(\theta)}{\theta}.$$

The timing of the model is the following. In period 1:

1. Hosts decide to enter the market;
2. Hosts post prices: $p_1 \in \mathbb{R}^+$;
3. Guests form beliefs about hosts effort decision observing p_1, $\mu_1(p_1)$;
4. Guests choose where to direct their search and matches are formed;[2]
5. Hosts matched with a guest draw their cost of effort c;
6. Hosts choose whether or not to exert effort: $e_1(c, p_1)$;
7. Transactions occur.

At the end of period 1, a history h is formed for each host and it is public information. If a host was matched with a guest, her history is a couple composed by the price posted in period 1 and the effort, $h = (p_1, e_1(c))$. If the host was not matched, her history is composed by the price and the information that the host had no guests: $h = (p_1, \emptyset)$. Hosts who enter in period 2 have a blank history $h = (\emptyset)$. After observing histories, guests form interim beliefs $\bar{\mu}_2(h)$.

In period 2, the same timing applies. Still, guests may update their interim beliefs about hosts' effort observing current prices:

1. Hosts decide to enter the market;
2. Hosts post prices: $p_2(c, h) \in \mathbb{R}^+$;
3. Guests update interim beliefs about hosts effort decision observing $p_2(c, h)$, $\mu_2(p_2(c, h), h)$
4. Guests choose where to direct their search and matches are formed;

[2] I do not explicitly model the search process by guests. Depending on how the market is organized, different matching functions (all satisfying Assumption 1) can be micro-founded. For further details, see Peters (1991), Burdett et al. (1995) and Burdett et al. (2001).

5. Hosts matched with a guest who were not matched in period 1 draw their cost of effort c;
6. Hosts choose whether or not to exert effort: $e_2(c, p_2(c, h), h)$;
7. Transactions occur.

2.2 Equilibrium Characterization

The equilibrium concept used is symmetric perfect Bayesian equilibrium with pure strategies in prices. In this setting, posted prices play two separate functions. First, prices "direct" guests search as they affect the number of guests who are willing to be matched with hosts. Moreover, prices posted in period 2 can be a signal for hosts' cost of effort. I limit my analysis imposing some assumptions regarding these two tasks of prices.

In line with the directed search literature, I assume that, in each period, the ex-ante guests' utility U_t cannot be affected by the price posted by a single host:

$$U_t = (a\bar{\mu}_t + b - p_t)\frac{\alpha(\theta_t)}{\theta_t}, \tag{2.1}$$

where $\bar{\mu}_t$ defines guests' beliefs about hosts' effort choice. Accordingly, changes in price p_t that do not affect guests' beliefs $\bar{\mu}_t$ are fully compensated by changes in tightness θ_t: if a host chooses a lower price, more guests direct their search towards her until the tightness increases and the probability to have a transaction for guests decreases. Equation 2.1 characterizes guests' beliefs about tightness levels for every price, even for those prices that are not posted in equilibrium. This approach is denoted in the directed search literature as the "market utility" approach (Wright et al., 2017).

In this setting, prices in period 2 can be a signal for hosts' cost of effort since they can affect guests' beliefs $\bar{\mu}_t$. After a host is matched with a guest in period 1, her cost of effort is realized and it is private information. Hosts' cost of effort is relevant for guests' utility: while hosts with cost $c = 0$ always exert effort, hosts with positive cost $c = k > 0$ strategically choose whether to exert effort or not.

Yet, prices in period 2 are not the only variable signaling hosts' cost of effort. Hosts' histories are observed by guests in period 2 and they may be informative of hosts' cost. When a host's history reports $e_1 = 0$, guests in period 2 know with certainty that she has positive cost of effort (hosts with zero cost always choose to exert effort) and she does not exert effort in period 2: $\bar{\mu}_2 = 0$. Differently, histories reporting $e_1 = 1$ can sustain positive guests' beliefs about hosts' effort in period 2 ($\bar{\mu}_2 \geq \pi$).

The signaling functions of prices and histories are related. If prices fully solve the asymmetry of information between hosts and guests, histories' signal of hosts' cost of effort is ineffective. In

particular, if hosts with different cost of effort have separate pricing strategies in period 2, then guests perfectly infer hosts' costs and, in equilibrium, hosts with zero cost exert effort $e_1(0) = e_2(0) = 1$, whereas hosts with positive cost do not exert effort $e_1(k) = e_2(k) = 0$. I restrict my analysis over a class of equilibria where histories provide effective signals about hosts' costs, and I denote these equilibria as *reputational equilibria.*[3] Focusing on reputational equilibria is motivated by two reasons. Empirical evidence suggests that prices do not fully reveal users' private type. Histories (reviews) are important to reduce the asymmetry of information in digital platforms.[4] Moreover, outside the class of reputational equilibria, hosts who draw a positive cost of effort in period 1 do not exert effort in any of the two periods ($e_1(k) = e_2(k) = 0$). Differently, in reputational equilibria, hosts who draw a positive cost may exert effort in period 1 ($e_1(k) = 1$) in order to mimic hosts with $c = 0$ and get a price premium in period 2. Thus, since exerting effort is efficient ($a > c$), reputational equilibria are Pareto superior in terms of the ex-post surplus of transactions relative to other non-reputational equilibria.

The class of reputational equilibria is characterized by pooling strategies in prices for hosts with the same history in period 2. In period 1, all hosts post the same price since the cost of effort is drawn after matches are formed. Accordingly, guests in both periods cannot infer hosts' costs directly from prices in period 1. After transactions occur, hosts have different histories depending on the reported effort, which affect guests' beliefs $\bar{\mu}_2$ about hosts' effort choice in the future. In period 2, hosts with the same history post the same price. In particular, hosts who were not matched in period 1 and new entrants post the same price since their cost of effort is drawn after matches. The case is similar for hosts who were matched in period 1. By pooling in prices, hosts with $c = k > 0$ obtain a price premium in period 2 if they exert effort in period 1. It constitutes the reputational benefits (the "carrot") of having exerted effort. Conversely, if hosts with $c = k > 0$ do not exert effort, they cannot pool in period 2 and their cost is fully disclosed (the "stick"). Price pooling is vital to implement the "carrot-stick strategy" that characterizes reputational equilibria. Multiple prices can sustain these equilibria and a continuum of equilibria is present in this class. In the main text, I restrict my analysis to the price profile that implements the constrained efficient equilibrium allocation and maximizes hosts' profits. To do so, I consider guests' beliefs that disregard the additional signaling role of prices in period 2: for any posted price, guests in period 2 do not update their beliefs about hosts' cost of effort (formed observing the host's history). This restriction is not necessary since a wide range of guests' beliefs sustains the constrained efficient equilibrium allocation. Disregarding the signaling from prices in period 2 is justified by the following observation. Independently of their cost of effort, hosts with the same history in period 2 have the same profit function: hosts with

[3] In Appendix, I discuss non-reputational equilibria and I show that their existence and stability relies on further assumptions regarding model's parameters.

[4] Cabral and Hortaçsu (2010), Fan et al. (2016), and Jolivet et al. (2016) show evidence regarding the significant impact of reviews on sellers' profitability in several online marketplaces.

$c = k > 0$ do not exert effort in period 2 and their expected profits are $p_2\alpha(\theta_2)$; similarly, hosts with $c = 0$ do exert effort (that is costless for them) and get $p_2\alpha(\theta_2)$. Accordingly, the optimal pricing strategy is aligned for both hosts' types and guests may not update their beliefs after observing prices in period 2. Furthermore, thanks to the profit function alignment for hosts with different costs of effort, reputational pooling equilibria are not eliminated by refinements such as the intuitive criterion by Cho and Kreps (1987).

Before providing a formal definition of the equilibrium, I characterize hosts' decisions proceeding by backward induction.[5] The effort decision in period 2 is straightforward.

Lemma 1 (Effort Decision in Period 2)**.** *In equilibrium, hosts who are matched with a guest in period 2 exert effort if and only if they have zero cost of effort $c = 0$.*

Lemma 1 directly follows from the assumption that hosts with cost $c = 0$ always exert effort. Differently, hosts with cost $c = k > 0$ always exert $e_2(k) = 0$ since effort is costly for them and they cannot commit to exert positive effort since guests direct their search without knowing hosts' effort decision.

In period 2, hosts post prices to match with guests. Hosts with the same history who were matched with guests in period 1 post the same price. Hosts who were not matched with guests in period 1 post the same price with new entrants since no information pertaining their cost of effort is revealed.

Proposition 1 (Pooling in Prices in Period 2)**.** *In equilibrium, hosts who were matched with a guest in period 1 and have history h pool in prices in period 2. Given guests' interim beliefs $\bar{\mu}_2(h)$ and the expected utility U_2, hosts post prices $p_2^{pool}(h)$ and guests direct their search so as to form tightness $\theta_2^{pool}(h)$:*[6]

$$\alpha'(\theta_2^{pool}(h)) = \frac{U_2}{a\bar{\mu}_2(h) + b}$$

$$p_2^{pool}(h) = a\bar{\mu}_2(h) + b - \frac{\theta_2^{pool}(h)}{\alpha(\theta_2^{pool}(h))}U_2,$$

if $a\bar{\mu}_2(h) + b \geq U_2$. Otherwise, $\theta_2^{pool}(h) = 0$ and $p_2^{pool}(h) = 0$. Hosts who were not matched with a guest in period 1 and new entrants post the same price $p_2^{\emptyset}$ and guests direct their search so as to

[5] In Appendix, I illustrate the constrained efficient allocation and I discuss the Hosios (1990) conditions that characterize the equilibrium (proposed in the main text) implementing this allocation.

[6] To simplify the notation, in the remaining part of the analysis, superscripts denote hosts' costs. If hosts who draw different cost of effort may play the same strategy, superscript "pool" is present. If hosts who have not yet drawn the cost of effort play a strategy, the superscript $\emptyset$ is present.

form tightness $\theta_2^{\emptyset}$*:*

$$\alpha'(\theta_2^{\emptyset}) = \frac{U_2}{a\pi + b}$$
$$p_2^{\emptyset} = a\pi + b - \frac{\theta_2^{\emptyset}}{\alpha(\theta_2^{\emptyset}}U_2,$$

if $a\pi + b \geq U_2$*. Otherwise,* $\theta_2^{\emptyset} = 0$ *and* $p_2^{\emptyset} = 0$*.*

Proposition 1 establishes a relationship between the price posted by hosts in period 2 and the effort exerted in period 1. If hosts do not exert effort, guests realize that they have positive cost $c = k > 0$ and they do not exert effort in period 2: $\bar{\mu}_2(h) = 0$. Conversely, if hosts exert effort, then guests can only partially guess their cost of effort and $\bar{\mu}_2(h) > \pi$. Accordingly, exerting effort in period 1 rises hosts' prices $p_2^{pool}(h)$ and the probability to have a transaction $\alpha(\theta_2^{pool}(h))$.

In period 1, hosts who draw a positive cost of effort $c = k > 0$ choose whether to exert effort or not. Their decision is reported in their history and it changes the expected profits in period 2 according to Proposition 1.

Proposition 2 (Effort Decision in Period 1)**.** *In equilibrium, hosts who are matched with a guest in period 1 always exert effort if they have zero cost of effort,* $c = 0$*:* $e_1(0) = 1$*. If their cost of effort is positive,* $c = k > 0$*, they exert effort with probability* $\omega \in [0, 1]$*.* ω *is unique and it depends on the values of* a, b, π*, the cost of effort* k*, and the discount factor* β*.*

In period 1, hosts have not yet drawn their cost of effort. Accordingly, the optimal pricing in period 1 is established in a condition of symmetric information between hosts and guests. It is uniquely derived as follows.

Proposition 3 (Prices in Period 1)**.** *In equilibrium, given guests' expected utility for a match* U_1*, hosts post prices* $p_1^{\emptyset}$ *and guests direct their search so as to form tightness* $\theta_1^{\emptyset}$*:*

$$\alpha'(\theta_1^{\emptyset}) = \frac{U_1}{a(\pi + (1-\pi)\omega) + b - k(1-\pi)\omega + \beta\Delta\Pi}$$
$$p_1^{\emptyset} = a(\pi + (1-\pi)\omega) + b - \frac{\theta_1^{\emptyset}}{\alpha(\theta_1^{\emptyset})}U_1,$$

where $\Delta\Pi$ *represents the hosts' value of a transaction in terms of reputation updating. It is defined as follows:*

$$\Delta\Pi = \Pi_2\big(a\frac{\pi}{\pi + (1-\pi)\omega} + b\big)(\pi + (1-\pi)\omega) + (1-\pi)(1-\omega)\Pi_2(b) - \Pi_2(a\pi + b). \qquad (2.2)$$

If $a\pi + b + \beta\Delta\Pi < U_1$*,* $\theta_1^{\emptyset} = 0$ *and* $p_1^{\emptyset} = 0$*.*

Proposition 3 establishes that the value of a transaction in period 1 is not only related to the guests' expected utility $(a(\pi + (1-\pi)\omega) + b)$ and the cost of effort $(k(1-\pi)\omega)$, but it embeds a reputational value for hosts. If hosts draw $c = 0$, then they strictly benefit from having a transaction since they get, with zero cost, expected profits $\Pi_2(a\frac{\pi}{\pi+(1-\pi)\omega} + b)$ in period 2 with $\Pi_2(\frac{\pi}{\pi+(1-\pi)\omega}) \geq \Pi_2(a\pi + b)$. Conversely, if hosts draw $c = k > 0$, then having a transaction is not necessarily beneficial in terms of reputation updating. In particular, if $\omega = 0$, hosts with $c = k > 0$ get $\Pi_2(b) \leq \Pi_2(a\pi + b)$.

In the remaining part of this Section, I provide a formal definition of reputational equilibria and I analyze their existence and uniqueness. To simplify the notation, I denote histories that appear in equilibrium with positive probability as follows:

$$h^1 = (p_1^\emptyset, e_1 = 1);$$
$$h^0 = (p_1^\emptyset, e_1 = 0);$$
$$h^\emptyset = (p_1^\emptyset, \emptyset) = (\emptyset).$$

The same notation $h^\emptyset$ can be used to denote histories $(p_1^\emptyset, \emptyset)$ and $(\emptyset)$. The two histories are equivalent since guests' interim beliefs in period 2 are the same for both histories as well as the hosts' pooling pricing decision.

Definition 1 (Reputational Equilibrium). *A Reputational Equilibrium is defined by the following elements:*

- $n_1, p_1, \mu_1(p_1), U_1$*: the number of hosts who enter the market, the pricing decision, the guest' beliefs about hosts' effort, and the guests' expected utility for a match in period 1;*
- $e_1(c, p_1)$*: the effort decision by hosts with cost of effort* $c = 0$ *and* $c = k > 0$ *in period 1;*
- $n_2(h), p_2(c, h), \bar{\mu}_2(h), \mu_2(h, p_2(c, h)), g_2(h, p_2(c, h)), U_2$*: the number of hosts with history h present in the market, the hosts' pricing decision for each cost and history, the guests' interim and updated beliefs about hosts' effort, the number of guests who direct the search to hosts with certain history and price, and the guests' expected utility for a match in period 2.*
- $e_2(c, p_2(c, h))$*: the effort decision by hosts with cost of effort* $c = 0$ *and* $c = k > 0$ *in period 2.*

The following conditions are satisfied in equilibrium:

1. *The free-entry condition does not allow positive profits for hosts who enter the market in period*

1:

$$
\begin{aligned}
(p_1^{\emptyset} - c((1-\pi)\omega))\alpha(\theta_1^{\emptyset}) \\
+(1-\pi)(1-\omega)\beta\alpha(\theta_1^{\emptyset})p_2^{pool}(h^0)\alpha(\theta_2^{pool}(h^0)) \\
+(\pi+(1-\pi)\omega)\beta\alpha(\theta_1^{\emptyset})p_2^{pool}(h^1)\alpha(\theta_2^{pool}(h^1)) \\
+\beta(1-\alpha(\theta_1^{\emptyset}))p_2^{\emptyset}\alpha(\theta_2^{\emptyset}) \leq f;
\end{aligned}
$$

2. *Hosts post prices in period 1 according to Proposition 3;*

3. *Guests' beliefs about hosts' effort in period 1 are derived as follows:*

$$\mu_1(p_1^{\emptyset}) = \pi + (1-\pi)\omega;$$

4. *The market tightness in period 1 is defined as* $\theta_1^{\emptyset} = \frac{1}{n_1}$*;*

5. *Guests' expected utility for a match in period 1 is defined as follows:*

$$(a(\pi+(1-\pi)\omega) + b - p_1^{\emptyset})\frac{\alpha(\theta_1^{\emptyset})}{\theta_1^{\emptyset}} = U_1$$

6. *Hosts exert effort in period 1 depending on their cost of effort according to Proposition 2;*

7. *The measures of hosts with history* h^1 *and* h^0 *in period 2 depend on the measure of hosts who entered in period 1, the probability* π*, and the probability to exert effort by hosts with* $c = k > 0$*,* ω*:*

$$
\begin{aligned}
n_2(h^1) &= (\omega n_1(1-\pi) + \pi n_1)\alpha(\theta_1^{\emptyset}) \\
n_2(h^0) &= (1-\omega)n_1(1-\pi)\alpha(\theta_1^{\emptyset});
\end{aligned}
$$

8. *The free-entry condition does not allow positive profits for hosts who enter the market in period 2:*

$$p_2^{\emptyset}\alpha(\theta_2^{\emptyset}) \leq f;$$

9. *Guests' interim beliefs about hosts' effort in period 2 are formed applying Bayes formula when*

possible:

$$\bar{\mu}_2(h^1) = \frac{\pi}{\pi + (1-\pi)\omega}$$
$$\bar{\mu}_2(h^{\emptyset}) = \pi$$
$$\bar{\mu}_2(h) = 0, \ \forall h \neq h^1, h^{\emptyset};$$

10. *Hosts post prices in period 2 according to Proposition 1;*

11. *Guests (do not) update interim beliefs observing the price posted in period 2. Given a history h:*

$$\mu_2(h, p_2) = \bar{\mu}_2(h)$$

12. *Guests in period 2 are assigned to different sets of hosts characterized by the couple formed by history and price. Tightness levels are such that $\theta_2(h) = \frac{g_2(h, p_2^{pool}(h))}{n_2(h)}$ and:*

$$\sum_h g_2(h, p_2^{pool}(h)) = G;$$

13. *Guests' expected utility for a match in period 2 is the same across hosts active in the market:*

$$(a\frac{\pi}{\pi + (1-\pi\omega)} + b - p_2^{pool}(h^1))\frac{\alpha(\theta_2^{pool}(h^1))}{\theta_2^{pool}(h^1)} = U_2$$
$$(a\pi + b - p_2^{\emptyset})\frac{\alpha(\theta_2^{\emptyset})}{\theta_2^{\emptyset}} \leq U_2$$
$$(b - p_2^{pool}(h^0))\frac{\alpha(\theta_2^{pool}(h^0))}{\theta_2^{pool}(h^0)} \leq U_2$$

14. *Hosts exert effort in period 2 depending on their cost of effort according to Lemma 1.*

After having defined the equilibrium, I proceed with the theorem regarding its existence and uniqueness.

Theorem 1 (Existence and Uniqueness with Entry)**.** *If the number of guests active in period 2 is greater than a threshold value $\bar{G}$, then reputational equilibrium exists and it is unique. In this equilibrium, a positive mass of hosts enters in both periods.*

2.3 Testable Predictions

Here I propose two predictions of the model that can be directly tested using data from Airbnb. The first one regards the effort dynamics and it directly relates to Lemma 1 and Proposition 2. The second prediction is a comparative statics exercise and it investigates movements in hosts' effort strategy in equilibrium when entry costs change.

Proposition 4 (Effort Dynamics)**.** *Assume that the measure of guests present in the market in period 2 is big enough to allow hosts' entry in both periods. Then, in the reputational equilibrium, hosts' effort weakly decreases from period 1 to period 2.*

From Lemma 1, only hosts with cost $c = 0$ exert effort in period 2. Differently, in period 1 hosts with positive cost of effort $c = k > 0$ exert effort with probability $\omega \in [0, 1]$ as shown in Proposition 2. Proposition 4 directly follows with a weakly greater proportion of hosts exerting effort in period 1 relative to period 2.

Evidence in line with Proposition 4 are presented in Subsection 6.1. Extracting information from the review system implemented by Airbnb, I estimate the effort exerted by Airbnb hosts over time. In accordance with the previous result, a negative and significant relationship is present between hosts' effort and the number of reviews that hosts already received.

Proposition 5 (Entry Costs and Effort)**.** *Consider two levels of entry costs f and f' with $f' > f > 0$ and assume that the measure of guests present in the market in period 2 is big enough to allow hosts' entry in both periods for f and f'. Then, in the reputational equilibrium, the probability to exert effort for hosts with $c = k > 0$ in period 1 is weakly higher with f' relative to f: $\omega' \geq \omega$.*[7]

I provide here a heuristic proof for the above proposition.[8] If entry costs increase, the number of hosts who enter the market in period 2 decreases. The market is now tighter for guests in period 2 and guests' expected utility U_2 decreases. Conversely, hosts' expected profits increase and they increase more for hosts with higher beliefs to have $c = 0$. Thus, in period 1, hosts who draw $c = k > 0$ have stronger incentives to exert effort since the benefits of exerting effort - having a better reputation in period 2 - are higher.

The heuristics of the proof relies on the positive relationship between the tightness of the market is period 2 and the incentives to exert effort in period 1. In line with this reasoning, the empirical results in Section 5 regard the effect of a change in competition, due to a variation in entry costs, over the effort exerted by hosts on Airbnb.

The identification strategy described in Section 4 proposes an instrumental variable that follows the channel highlighted in the proof of Proposition 5. Hosts anticipate the movement

[7] The superscript " $'$ " denotes the equilibrium values associated with f'.

[8] The interested reader may find the complete proof in Appendix.

in tightness due to an exogenous change in entry costs. Thus, comparing hosts located in different areas, hosts exert more effort where the number of competitors drops more significantly: in a less competitive framework, exerting effort leads to greater reputational benefits.

3 Empirical Setting and Dataset

In this Section, I introduce the empirical part of my work. First, I present the Airbnb setting. Then, I describe the short-term rentals regulation in the city of San Francisco and I focus on the settlement agreement signed in May 2017 between the San Francisco City Council and Airbnb. Finally, I describe the unique dataset used for my analysis: I provide descriptive statistics regarding the population of Airbnb listings and the impact of the agreement signed in May 2017 in terms of hosts' selection.

3.1 Airbnb

Airbnb is one of the leading digital platforms in the hospitality industry. It works in more than 60,000 cities and it offers to its members the possibility to arrange and offer lodging and other tourism experiences. Airbnb receives a commission fee for every transaction and it does not own any real estate listed on the platform. I restrict my analysis over the lodging services and I denote the Airbnb members who arrange and offer accommodations as guests and hosts, respectively. To be an Airbnb member, a digital registration procedure is required. Airbnb guests need to insert personal information such as the email, phone number, and a scan of an identity document. The procedure to become an Airbnb hosts is different. It requires hosts to provide additional information and take photos of the listing; choose the days when they are willing to host; and set prices.[9] Further requirements are necessary for hosts due to local laws and regulations.

After being registered, guests and hosts appear on the Airbnb platform with a personal webpage. Guests can search for hosts selecting the location and the period of their stay. Furthermore, other advanced filters are available to restrict the guests' search, such as price range and listings' characteristics. Guests can select hosts and visit their webpages. Then, they can choose to book the listing. If hosts accept guests' requests, their listings are officially booked.

After the guest's stay, host and guest have 14 days to review each other. Guests feedback consists of four elements:

1. A written comment;

[9] For more information regarding the registration procedure for Airbnb hosts, see the official Airbnb guide to become a host at www.airbnb.com/b/hosting_checklist.

2. Private comments to the host;

3. A one-to-five star rating about the overall experience;

4. Six specific ratings regarding the following categories:

   – The accuracy of the listing description;

   – The check-in process at the beginning of the stay;

   – The cleanliness of the listing;

   – The communicativeness of the host;

   – The listing location;

   – The "value-for-money" of the stay.

Similarly, host can review guests answering whether or not she would recommend the guest; writing a comment; and rating the guest considering the communicativeness, the cleanliness and how well the guest respected the rules of the house.

Not all these elements are published on the platform and, for what concerns the guest feedback, only written comments are directly published on hosts' webpages. The other ratings are not displayed singularly with the comments: only the rounded average of the score and subscores are published on the listing and the host webpages. In the same way, only the comment written by the host is published on the guest webpage.

3.2 Institutional Setting

Airbnb and other online marketplaces have had a sizable impact on the hospitality industry and many city councils have tried to regulate rentals on digital platforms. I restrict my analysis over the city of San Francisco and I report here a synthetic chronology of the regulations adopted by the San Francisco City Council starting from the San Francisco Short-Term Rentals Regulation enacted at the end of 2014.

3.2.1 The San Francisco Short-Term Rentals Regulation (February 2015)

With an ordinance signed in October 2014 and effective from February 2015, San Francisco City Council legalized short-term rentals in the city and imposed a regulation. Before this ordinance, San Francisco banned short-term rentals in residential buildings. Rentals are considered "short-term" if the properties are rented for less than 30 consecutive nights at a time. Short-term rentals constitute the great majority of transactions occurring on hospitality digital platforms. Still, listings present

on Airbnb can be exempt from the regulation if they only accept guests for periods of 30 or more days; or in case they are professional structures such as hotels and B&B. The regulation is mainly composed by the following parts:[10]

- Only San Francisco permanent residents who own or rent single-family dwellings in the city are eligible to engage in short-term rentals. In particular, hosts must reside in their dwellings for at least 275 days per year;
- Resident tenants must notify their landlords before engaging in short-term rentals. If the contract between tenant and landlord prohibits subletting, the landlord may evict the tenant. Moreover, tenants cannot charge more rent than they are paying to the landlord and rent control laws must be respected;
- Only the primary residence can be used for short-term rentals;
- When a host is absent, the dwelling can be rented for a maximum of 90 days per year;
- Hosts must obtain a permit and register at the Office of Short Term Rental. Every two years, they must pay a $250 fee. Moreover, hosts are required to obtain a city business license;
- The San Francisco hotel tax must be collected from renters and paid to the city. For Airbnb hosts, the platform automatically collects and pays such a tax for its hosts;
- Hosts must be covered by a insurance with a coverage of at least $500,000. Airbnb provides hosts with 1 million in coverage. Compliance to city building code requirements is necessary.

This regulation introduces several limitations on who can offer lodging service on Airbnb. To be legally present on the platform, hosts have to face additional costs and respect extra requirements.

In the first years after the introduction of the regulation, the enforcement of part of the law has proven to be difficult. In particular, regulators could not enforce the rules regarding hosts residence since registration rates at the Office of Short Term Rental were very low and digital platforms did not disclose to the authorities any personal information regarding their hosts. Because of the difficulties regarding the law enforcement, San Francisco City Council enacted an additional ordinance in June 2016 that required digital platforms to list on their websites only legal listings with an official registration. Airbnb filed a suit against the City Council and, after a U.S. judge rejected the suit and postponed the enforcement of the new rules, an agreement was found in May 2017.

[10] For a comprehensive analysis of all the regulation's requirements, see the Short-Term Residential Rental Starter Kit provided by the San Francisco Office of Short Term Rental at https://businessportal.sfgov.org/start/starter-kits/short-term-rental, and the official text of the ordinance at https://sfgov.legistar.com/View.

3.2.2 The Settlement Agreement with Airbnb and HomeAway (May 2017)

The agreement clarifies the role of digital platforms in the hosts registration process for short-term rentals. It has been signed, together with the San Francisco City Council, by Airbnb and another hospitality platform, HomeAway. The main resolutions are the following:[11]

- From September 2017, new hosts willing to arrange a short-term rental on Airbnb or HomeAway have to "input their city Office of Short-Term Rental registration number (or application pending status) to post a listing";
- From September 2017, a "pass-through registration" system is implemented by Airbnb and HomeAway for hosts who are already registered on the platforms to send applications directly to the Office of Short Term Rental for consideration. If the platforms receive notice of an invalid registration, they will cancel future stays and deactivate the listings;
- From January 2018, all hosts present on Airbnb and HomeAway are required to be registered. If some listings are not registered at this date, the platform will cancel future stays and deactivate the listings until a registration number (or application pending status) is provided;
- Airbnb and HomeAway will regularly provide the city with information about San Francisco listings to allow for an effective enforcement of the law.

3.3 InsideAirbnb Dataset

The dataset for this study comes from information on InsideAirbnb, a website that tracks all the Airbnb listings present in specific locations over time.[12] In my analysis, the dataset is formed by forty snapshots of all the Airbnb listings present in San Francisco at forty different dates from May 2015 to December 2018. Data scraping is performed at the beginning of each month with some months missing in 2015 and some multiple snapshots per month at the beginning of 2018.[13] I combine all the snapshots to form an unbalanced panel dataset composed by 28,012 listings and 298,556 listing observations over time. In each snapshot, listings are observed if they appear on the Airbnb website

[11] All quotes are from the official announcement of the San Francisco City Attorney, available at https://www.sfcityattorney.org.

[12] All data are publicly available on InsideAirbnb. InsideAirbnb is "an independent, non-commercial set of tools and data that allows you to explore how Airbnb is really being used in cities around the world" and all scraped data are available under a Creative Commons CC0 1.0 Universal (CC0 1.0) license.

[13] The list of all snapshots follows: May 2015, September 2015, November 2015, December 2015, February 2016, April 2016, May 2016, June 2016, July 2016, August 2016, September 2016, October 2016, November 2016, December 2016, January 2017, February 2017, March 2017, April 2017, May 2017, June 2017, July 2017, August 2017, September 2017, October 2017, November 2017 (two snapshots), December 2017 (two snapshots), January 2018 (two snapshots), February 2018, March 2018, April 2018, May 2018, July 2018, August 2018, September 2018, October 2018, December 2018.

at the snapshot date. Accordingly, for each Airbnb listing in the dataset, entry, exit and inactivity periods are identified.[14] When a listing is observed, several listing's characteristics are displayed. Some are time-invariant such as the listing's location (longitude, latitude and neighborhood), and dwelling's characteristics. Some others update at each snapshot such as the number of guests' reviews and average star ratings, the price charged for one night at the snapshot date, the number of nights in which the listing is available after the snapshot, whether or not the listing displays the Office of Short-Term Rental registration number and whether the registration is necessary for the listing.

Descriptive statistics are reported in Table 1. Panel A presents the characteristics of all listings observed in the panel data from May 2015 to December 2018. All the reported variables regard the last snapshot in which listings are observed. The average amount of time in which listings are present on Airbnb is approximately one year. The total number of reviews presents a skewed distribution with more than half of listings having less than 5 reviews before exiting the platform. High variability characterizes the price per night and the number of nights in which the listing is available after the snapshot, implying that the listings' performances on Airbnb widely vary across listings. In contrast, the variation of the average rating is much lower. The average rating regarding the overall experience is 93.70 with standard deviation 9.04, that corresponds to an average of almost 5 stars with an extremely limited variation.[15] This result is in line with Zervas et al. (2015) who observe that almost 95 percent of Airbnb listings have an average rating greater or equal than 4.5 stars. The percentages regarding the number of hosts engaging in short-term rentals and displaying a registration number confirm two elements highlighted in the previous Subsection regarding Airbnb in San Francisco. First, the short-term rentals constitute the great majority of transactions occurring on Airbnb (more than 80%). Second, before the Settlement Agreement, the regulation imposed by the San Francisco City Council was largely neglected.

Panel B shows listings information regarding the number of reviews written between two consecutive snapshots and the averages of the ratings associated to these reviews. All the variables are constructed starting from the original variables shown in Panel A. I call these variables, the number of reviews per snapshot and the average ratings per snapshot. The number of reviews per snapshot is derived from the difference between the total number of reviews displayed in a snapshot and in the next ones ($n_{i,t+1} - n_{i,t}$). Similarly, the average ratings per snapshot are computed using the average rating and the total numbers of reviews. I denote with $n_{i,t}$ and $\bar{R}^k_{i,t}$ the total number of reviews displayed for listing i at snapshot t and the average rating displayed for listing i at snapshot t

[14] Airbnb hosts can decide to remove their listings from Airbnb for a period of time and then re-enter with the same listing profile.

[15] On the Airbnb platform guests can choose in a range of stars between 1 and 5. Still, the scraped variable regarding the average rating for the overall experience varies from 0 to 100. All other scraped ratings varies from 0 to 10.

for the category k, respectively. Then, the average rating per snapshot, $\bar{r}^k_{i,t}$, for listing i at snapshot t and category k where $k \in \{$overall, accuracy, check-in, cleanliness, communication, location, value$\}$ can be computed as follows:[16]

$$\bar{r}^k_{i,t} = \frac{\bar{R}^k_{i,t+1} n_{i,t+1} - \bar{R}^k_{i,t} n_{i,t}}{n_{i,t+1} - n_{i,t}}.$$

The number of reviews per snapshot varies for listings and snapshots. The average number of review per snapshot equals 1.62 with standard deviation 2.93. More limited variations are present for the average ratings per snapshots. For all ratings, the averages are higher than 9 with standard deviations lower than 1.2.

3.3.1 The Settlement Agreement: Hosts' Selection

The Short-Term Rental Regulation has been effective since February 2105. Still, as highlighted in Subsection 3.2, the enforcement of listings' registration at San Francisco Office of Short-Term Rental has proven to be difficult. The Settlement Agreement, effective from September 2017, solved the enforcement difficulties of registration. It implemented a resolution that forced every eligible Airbnb listings to be registered before January 2018. Figure 1 reports the percentage of listings that displayed a registration number at each snapshot. Before September 2017, less than 15% of all listings displays registration numbers. Conversely, at the beginning of 2018 the percentage of listings with registration numbers reaches more than 80%. Even though registration is mandatory after January 2018, I observe a group of listings without registration number since not all Airbnb listings need a registration: those who are not involved in short-term rentals, and those who are officially registered hotels and B&B are exempt from the regulation. The decrease in the percentage of registered listings during 2018 (in December 2018 less than 70% of listings have a registration) is due to the greater presence of exempted listings in the platform. After the agreement, the presence of short-term lodgings in residential buildings reduces in favor of professional services (hotels and B&B) and rental periods longer than 30 days.

Figure 2 captures the change in the total population of Airbnb listings in San Francisco at each snapshot. After a fast growth from May 2015 to December 2016, the number of Airbnb listings remains between 8000 and 9000 units until September 2017, when the "pass-through registration" system started to be in place. After September 2017, the number of listings sharply drops with less than 5000 units being present in February 2018, when all eligible Airbnb listings should be

[16] Since $\bar{R}^k_{i,t}$ are rounded averages, the procedure is likely to be affected by measurement errors. In order to reduce these errors, I substitute values of $\bar{r}^k_{i,t}$ lower than 0 or greater than 10 with 0 and 10, respectively. These substitutions regard the 0.15 percent of the sample. Moreover, I drop from the estimating sample snapshots with a number of reviews per snapshot greater than 26. I treat these snapshots as outliers due to the scrapping method. They account for the 0.01 percent of the sample.

Table 1: Summary Statistics

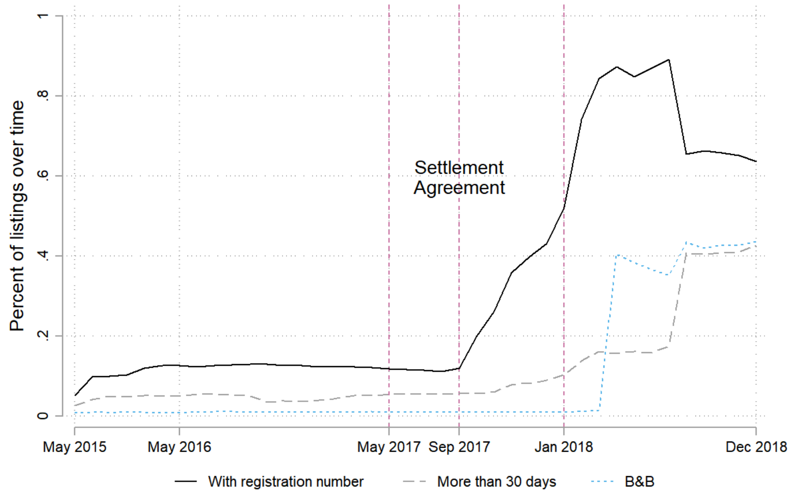
	Mean	SD	N	Min	Max
Panel A					
Days in Airbnb	352.51	379.56	28,012	0.00	1,312.00
Total number of reviews	20.79	45.87	28,012	0.00	649.00
Percent of the listing population:					
Less than 5 reviews	57%	-	28,012	-	-
Between 5 and 10 reviews	10%	-	28,012	-	-
Between 10 and 20 reviews	10%	-	28,012	-	-
Between 20 and 50 reviews	11%	-	28,012	-	-
Between 50 and 100 reviews	6%	-	28,012	-	-
More than 100 reviews	5%	-	28,012	-	-
No short-term rentals	16%	-	28,012	-	-
Registration displayed or not required	41%	-	28,012	-	-
Price per night	243.76	548.26	28,012	0.00	30,000.00
Availability next 30 days	10.22	11.41	28,012	0.00	30.00
Availability next 60 days	22.91	23.09	28,012	0.00	60.00
Availability next 90 days	37.48	35.05	28,012	0.00	90.00
Minimum nights required	6.40	10.56	28,012	1.00	100.00
Panel B					
Average rating: overall	93.80	9.15	19,641	20.00	100.00
Number of reviews per snapshot	1.62	2.93	22,676	0.00	26.00
Average rating per snapshot: overall	93.14	8.84	13,741	0.00	100.00
Average rating per snapshot: accuracy	9.51	0.92	13,732	0.00	10.00
Average rating per snapshot: check-in	9.67	0.80	13,721	0.00	10.00
Average rating per snapshot: cleanliness	9.30	1.12	13,736	0.00	10.00
Average rating per snapshot: communication	9.67	0.82	13,732	0.00	10.00
Average rating per snapshot: location	9.40	0.94	13,720	0.00	10.00
Average rating per snapshot: value-for-money	9.12	1.03	13,719	0.00	10.00

Note: Panel A refers to every single listings present in the panel data combining the snapshots from May 2015 to December 2018. All the statistics refer to the last snapshot in which the listing is observed. The variable "Days in Airbnb" is derived considering the difference between the last and the first snapshot in which the listing is observed. The "Percent of the listing population" groups listings in different categories. Listings are divided by the number of reviews that are displayed in their last snapshot; whether or not they engage in short-term rentals; and whether or not they display a registration number. Panel B refers to the variables constructed from the original dataset about the number of reviews written between two consecutive snapshots and the averages of the ratings associated to these reviews. The statistics in Panel B do not refer to the last snapshot, but they are the average of all values observed per each listing. Missing data regarding the variables "Average rating" are due to the high presence of listings with no reviews.

registered. The number of listings rises again at the end of 2018 (7000 units in December 2018) without reaching the pre-agreement levels.

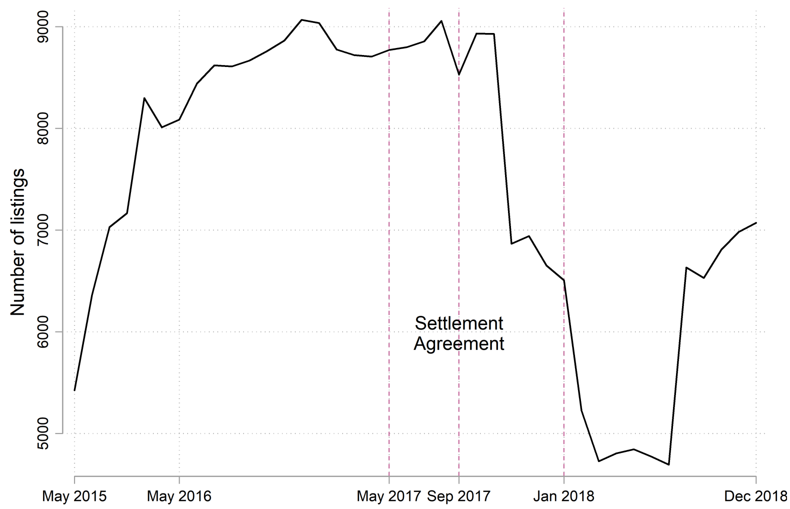
Accordingly, the Settlement Agreement determined a selection in the type of listings that continued to be present on the platform after the implementation of the registration requirements.

Figure 1: Percent of Airbnb Listings over Time



Note: The figure plots three time series regarding percentages of Airbnb listings in San Francisco over time from May 2015 to December 2018. The solid line represents the percent of listings that displayed registration number. The other two dotted lines represent the percent of listings that do not offer short-term lodging and the percent of B&B, respectively. Both categories are exempted from the Short-Term Rental Regulation. The three vertical lines regard the Settlement Agreement between the San Francisco City Council and Airbnb. The agreement was signed in May 2017 and it has been effective since September 2017. According to the resolution, from January 2018 all eligible Airbnb listings should be registered.

Figure 2: Airbnb Listings Population over Time



Note: The figure plots the total population of Airbnb listings in San Francisco over time (different snapshots) from May 2015 to December 2018. The three vertical lines regards the Settlement Agreement between the San Francisco City Council and Airbnb. The agreement was signed in May 2017 and it has been effective since September 2017. According to the resolution, from January 2018 all eligible Airbnb listings should be registered.

In Table 2, I present some summary statistics to characterize this selection process. Listings are divided in four groups: Group A contains all listings that exit the platform before September 2017,

when the implementation of the Settlement Agreement has not yet started. Differently, Group B contains all listings that enter the platform after September 2017. Listings in Group C are those that enter the platform before September 2017 and exit after January 2018, when the implementation of the Settlement Agreement was completed. Finally, Group D contains all listings that enter the platform before September 2017 and exit between September 2017 and January 2018. Accordingly, only listings belonging to Group C "survived" the Settlement Agreement and, in Section 5, I focus on this group in order to study the change in the degree of competition faced by each listing due to the regulation implementation.

Panel A compares listings that have not been affected by the Settlement Agreement (Group A) with listings that entered after September 2017 (Group B). This latter group of listings tends to engage in significantly longer rentals relative to Group A. In particular, hosts in Group B require guests to stay and rent their house for at least 15 consecutive nights, on average; whereas hosts in Group A require, on average, less than 4 consecutive nights. Accordingly, listings who enter after the Settlement Agreement are much less likely to engage in short-term rentals than those listings that were active before September 2017. The difference in the duration of the lodging services across groups may explain other differentials in terms of prices and total number of reviews. The price per night charged by listings in Group A is significantly higher than the one charged by listings in Group B: shorter rentals tend to be more expensive. In the same line, longer stays mechanically produce a lower stream of reviews over time.

Moreover, listings in Group B tend to have significantly higher ratings than listings in Group A: this may be due to the different service duration, or to an improvement in the service quality provided by hosts.

A similar differential in the listing profiles is present in Panel B where listings that survived the Settlement Agreement (Group C) are compared with those who enter before September 2017 and exit during the implementation of the new regulation (Group D). Survivors requires guests to stay for more consecutive nights relative to listings in Group D and they charge lower prices. Still, they have a greater turnover since the number of reviews per snapshot is higher for Group C than Group D. Moreover, listings that stay after January 2018 have significantly higher ratings relative to those that exit before apart from the rating regarding location. In this sense, listings in Group C seem to be selected among those who were present on Airbnb before the Settlement Agreement. Table 11 (in Appendix) confirms the presence of listings' selection. Here I report statistics measured in September 2017 for listings in the two groups. In September 2017, listings in Group C have, on average, almost five times more reviews than listings in Group D, and enter the platform almost 65 days before. Moreover, listings in Group C charge significantly lower prices than listings in Group D and have higher ratings (apart from the rating regarding location).

Table 2: Summary Statistics: the Settlement Agreement and Hosts' Selection

	Group A		Group B			
	Mean	SD	Mean	SD	Δ	$p-value$
Panel A						
Days in Airbnb	159.14	193.31	117.33	112.21	41.81	0.00
Total number of reviews	11.16	25.00	7.85	16.72	3.30	0.00
Number of reviews per snapshot	1.82	2.82	1.32	2.21	0.50	0.00
Price per night	250.16	660.79	216.49	389.61	33.67	0.00
Average rating per snapshot: overall	91.33	10.37	94.85	8.96	-3.52	0.00
Average rating per snapshot: accuracy	9.32	1.09	9.66	0.84	-0.34	0.00
Average rating per snapshot: check-in	9.49	0.99	9.77	0.72	-0.28	0.00
Average rating per snapshot: cleanliness	9.08	1.28	9.45	1.17	-0.37	0.00
Average rating per snapshot: communication	9.53	1.01	9.75	0.76	-0.22	0.00
Average rating per snapshot: location	9.29	1.09	9.48	0.93	-0.19	0.00
Average rating per snapshot: value-for-money	8.96	1.17	9.22	1.10	-0.25	0.00
Minimum nights required	3.38	5.25	15.70	16.94	-12.31	0.00
No short-term rentals	3%	-	44%	-	-0.41	-
Registration displayed or not required	34%	-	51%	-	-0.16	-
Number of listings	12,980	-	5,706	-	-	-

	Group C		Group D			
	Mean	SD	Mean	SD	Δ	$p-value$
Panel B						
Days in Airbnb	964.74	304.34	567.10	247.90	397.64	0.00
Total number of reviews	71.77	81.72	12.58	27.68	59.19	0.00
Number of reviews per snapshot	2.50	4.31	0.66	1.27	1.84	0.00
Price per night	223.60	356.34	279.35	524.17	-55.75	0.00
Average rating per snapshot: overall	94.39	6.10	92.94	8.91	1.45	0.00
Average rating per snapshot: accuracy	9.66	0.64	9.48	0.94	0.18	0.00
Average rating per snapshot: check-in	9.80	0.54	9.67	0.76	0.13	0.00
Average rating per snapshot: cleanliness	9.51	0.76	9.20	1.16	0.31	0.00
Average rating per snapshot: communication	9.78	0.58	9.67	0.79	0.11	0.00
Average rating per snapshot: location	9.47	0.73	9.44	0.92	0.03	0.15
Average rating per snapshot: value-for-money	9.25	0.77	9.15	1.02	0.10	0.00
Minimum nights required	6.90	9.31	2.96	3.52	3.94	0.00
No short-term rentals	14%	-	1%	-	0.14	-
Registration displayed or not required	81%	-	7%	-	0.75	-
Number of listings	4,654	-	4,672	-	-	-

Note: The two panels of the table show and compare the profile of listings before and after the Settlement Agreement. All the statistics refer to the last snapshot in which the listing is observed apart from the variables "Average rating per snapshot". Listings are divided in four groups: Group A contains all listings who exit the platform before September 2017, when the implementation of the Settlement Agreement has not yet started. Group B contains all listings who enter the platform after September 2017. Group C contains all listings who enter the platform before September 2017 and exit after January 2018, when the implementation of the Settlement Agreement was completed. Group D contains all listings who enter the platform before September 2017 and exit before January 2018. The last two columns provide the differences between the statistics' averages and the $p-value$ of the difference.

4 Identification Strategy

In this Section I discuss the identification strategy of the causal relationship between listing competition in Airbnb and the hosts' propensity to exert effort. My approach exploits the effect of the Settlement Agreement over the degree of competition that each listing faces due to its geographical location. This strategy is similar of the approach used in recent studies of mergers such as Dafny et al. (2012), Ashenfelter et al. (2015) and Chandra and Weinberg (2017). In these works, variations in the market concentration index due to a merger are used to study how competition affects firms' behavior. In my work, the variation in competition is due to the change in the entry conditions established by the Settlement Agreement: it regards all listings in San Francisco. Still, it affected differently listing concentration in certain areas of the city. Accordingly, the change in competition that each listing faced because of the Settlement Agreement varies across listings. From this perspective, the identification approach is similar to the design by Stevenson et al. (2010) and Ahern and Dittmar (2012) where a policy change over time is combined with a cross-sectional measure of exposure.

Moreover, the channel of the research design closely follows the heuristics of Proposition 5 in Section 2: variations in the entry costs change market tightness (the proportion of guests and hosts) and affect the equilibrium hosts' effort in period 1.

The main estimating regression to capture the causal relationship between competition and hosts' effort is the following:

$$\ln(r_{i,t}^{effort}) = \alpha_i + \gamma_t + \beta \ln(NL_{i,t}^{j}) + \varepsilon_{i,t}, \tag{4.1}$$

where α_i and γ_t are the full set of dummy variables for each listing and snapshot.

$\bar{r}_{i,t}^{effort}$ is the measure of host' effort for listing i at snapshot t. In Section 5, I use two rating categories as proxies for hosts' effort: check-in and communication. From now on, I denote the average rating per snapshot for listing i, snapshot t and category check-in and communication with $\bar{r}_{i,t}^{check}$ and $\bar{r}_{i,t}^{comm}$, respectively. I use $\bar{r}_{i,t}^{effort}$ to simultaneously refer to both average ratings. The focus on these two categories is justified by a principal component analysis performed on all the rating categories (average rating per snapshot). In Appendix, Figure 4 plots the loadings of all categories over the first two components. Check-in and communication are the most correlated ratings and their loadings separate from all others. In Section 6, I provide an estimation of the effort exerted by hosts using a control function approach to account for reviews' confounding factors related to guests' characteristics.

$NL_{i,t}^{j}$ represents the degree of competition faced by listing i at snapshot t. It is defined as the sum of listings present at snapshot t that are distant to listing i less than j kilometers. In my

analysis I use three values for j: 0.5 kilometer, 1 kilometer, and 2 kilometers.[17]

With ordinary least squares (OLS), correlation between $NL^j_{i,t}$ and $\varepsilon_{i,t}$ produces inconsistent estimates of β. The main potential threat of endogeneity regards the presence of omitted variables concerning the demand side. A high number of competitors is a signal of the attractiveness of the area and high demand. Thus, regressing $\bar{r}^{effort}_{i,t}$ over $NL^j_{i,t}$ may partially capture the impact of the demand over hosts' effort.

To tackle the endogeneity issues, I implement an instrumental variables (IV) strategy exploiting the Settlement Agreement between the San Fransisco City Council, Airbnb and HomeAway. In the same spirit of Dafny et al. (2012) Ashenfelter et al. (2015) and Chandra and Weinberg (2017), I propose a measure p^j_i of the predicted change, due to the registration enforcement, in the sum of listings distant to listing i less than j kilometers. The measure p^j_i takes advantage of the sum of listings that display a registration number on their webpages (denoted with $NRL^j_{i,t}$) few days before the Settlement Agreement became effective.[18] It is defined as follows:

$$p^j_i = \frac{NRL^j_{i,Sept2017}}{NL^j_{i,Sept2017}},$$

where $NRL^j_{i,Sept2017}$ and $NL^j_{i,Sept2017}$ are the sum of listings with registration numbers and the total sum of listings, respectively, present at the beginning of September 2017 and distant less than j kilometers to listing i. With a value of p^j_i close to one, the competition for listing i is not expected to change much since a high number of listings already displays a license. Conversely, with low value of p^j_i the expected change of competition for listing i due to the Settlement Agreement is likely to be more relevant.

In this sense, the instrumental variable is formed by the product between p^j_i and $post_{Nov2017}$: a dummy variable that takes value 1 for each snapshot after November 2017 and it is zero otherwise.[19]

The quality of this instrument depends on the strong correlation between p^j_i and $NL^j_{i,t}$ and the assumption regarding the exclusion restriction.

In line with Dafny et al. (2012) and Chandra and Weinberg (2017), I show here to which extent the drop in the number of competitors of each listing can be predicted by p^j_i. To do this, I consider the following lead-lag model in which the degree of competition $NL^j_{i,t}$ is regressed over the product between p^j_i and a full set of dummy variables for each snapshot from September 2016 (one

[17] These variables are created using the information regarding latitude and longitude of each listing.

[18] The snapshot in September 2017 was scrapped on September 2, 2017, whereas the new registration process started September 6, 2017. See http://www.sfexaminer.com/airbnb-launches-new-registration-system/.

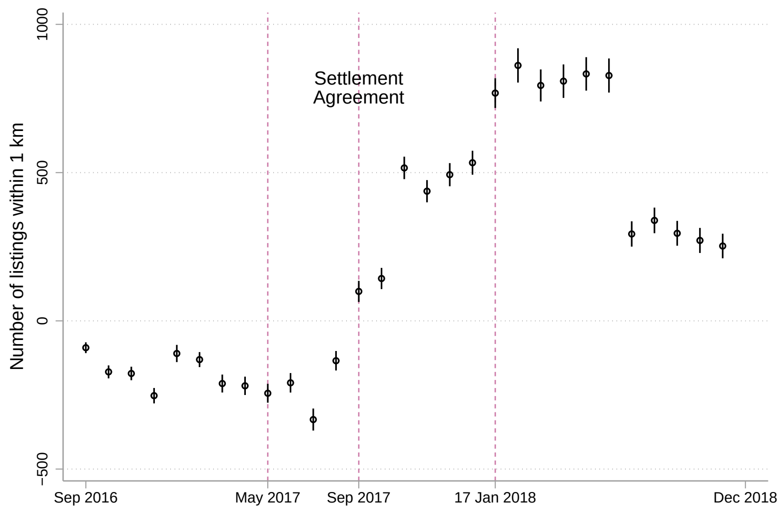
[19] From Figure 2, November 2017 results to be the first snapshot with a significant drop in the number of listings in the platform.

year before the registration enforcement):

$$\ln(NL^j_{i,t}) = \alpha_i + \gamma_t + \sum_{\tau=Sept2016}^{Dec2018} \beta_\tau p^j_i \times 1(t=\tau) + \varepsilon_{i,t}. \tag{4.2}$$

I present the results of the OLS estimates of Equation 4.2 with an event study graph. In Figure 3 I plot the estimated $\hat{\beta}_\tau$ over the snapshot dates considering the number of competitors between 1 kilometer distance. The decrease in the number of competitors due to the registration is observable in the first snapshots after September 2017 and the first ones of 2018. The impact decreases over time because of the entry of new listings (as it is suggested by Figure 1). Moreover, from Figure 3, it is possible to reject the presence of pre-trends in the number of listings correlated with the predicted impact of the regulation enforcement, p^j_i: there is no clear trend of the coefficients $\hat{\beta}_\tau$ before May 2017.

Figure 3: Estimated Coefficients from Equation 4.2



Note: In line with Equation 4.2, $\ln(NL^j_{i,t})$ is regressed on listing and snapshot fixed effects, and on the products between p^j_i and snapshot dummies. Standard errors are cluster by listing. The graph plots the estimated coefficients on these products.

For what regards the exclusion restriction, I provide here a list of arguments in favor of this assumption. There is no evidence that the San Francisco Short-Term Rental Regulation and the Settlement Agreement were motivated by concerns over the quality of the service on hospitality platforms.[20] Accordingly, other than reducing the number of listings present on the platform, the regulation enforcement of the Settlement Agreement may affect hosts' effort through the selection of hosts who stay after the enforcement of the registration. In particular, Tables 2 and 11 show

[20]The City Attorney, Dennis J. Herrera, never mentions the quality of the Airbnb service and the hosts' effort in his announcement of the Settlement Agreement, available at https://www.sfcityattorney.org.

that the implementation of the registration requirements leads to a positive selection over listings: those who stay have more reviews and better ratings. In order to tackle this issue, I restrict my analysis on those listings that enter before September 2017 and exit after January 2018, when the registration enforcement is completed (Group C in Section 3.3.1). For this sample, the identification strategy excludes the presence of unobserved factors that affect hosts' effort and are correlated with how the regulation changes the number of listings in different areas of San Francisco. Still, external validity concerns may be at place restricting the sample on those listings that survive the registration enforcement. The estimated effects of competition over hosts' effort are based on a selected part of the population. In particular, they do not take into account movements in the hosts' effort for listings that exit between September 2017 and January 2018 (Group D in Section 3.3.1). In Section 6 I address this issue providing evidence regarding the behavior of those listings that exit before January 2018 and those who enter after September 2017 (Group D and Group B in Section 3.3.1, respectively).

5 Results

I present now the main empirical results. I start estimating the OLS panel regressions that relate hosts' effort to the degree of competition as represented in Equation 4.1. Table 3 presents the results regarding hosts' ratings. For each rating, three regressions are performed: the independent variables vary depending on the distance used to delimit the competition faced by listings. The results suggest a not significant relationship between effort and competition measured as the sum of competitors within 0.5, 1 and 2 kilometers to each listing.

As described in Section 4, the OLS panel regressions are likely to be affected by the presence of omitted determinants of demand: the higher is the number of Airbnb hosts in a specific area, the greater is the area attractiveness for guests. Because of this, causality cannot be inferred from the OLS panel model. Accordingly, I take advantage of the variation in the degree of competition due to the Settlement Agreement to estimate the effect of competition over hosts' effort.

The "first stage" of the IV design documents a positive and significant relationship between the actual movement of the number of competitors for each listing and how the registration enforcement was expected to change the degree of competition. This evidence is in line with the event study in Figure 3. The estimating equation of the "first stage" is the following:

$$\ln(NL_{i,t}^{j}) = \alpha_i + \gamma_t + \beta p_i^j \times post_{Nov2017} + \varepsilon_{i,t}, \tag{5.1}$$

where the endogenous variable $\ln(NL_{i,t}^{j})$ is regressed over the expected change in competition due to the Settlement Agreement. The results with listings and snapshot fixed effects are in Table

Table 3: OLS Estimates of the Impact of Competition on Hosts' Ratings regarding Effort

	$\ln(\bar{r}_{i,t}^{comm})$			$\ln(\bar{r}_{i,t}^{check})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(NL_{i,t}^{0.5})$	0.000978			-0.00284		
	[0.00177]			[0.00193]		
$\ln(NL_{i,t}^{1})$		0.00179			-0.00113	
		[0.00267]			[0.00296]	
$\ln(NL_{i,t}^{2})$			0.0000261			-0.00162
			[0.00344]			[0.00334]
Constant	2.261***	2.255***	2.264***	2.279***	2.274***	2.278***
	[0.00774]	[0.0146]	[0.0230]	[0.00843]	[0.0162]	[0.0223]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00492	0.00637	0.00663	0.00183	0.00200	0.00198
N	145053	145098	145134	145053	145098	145134

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4. The expected movement in the number of competitors, p_i^j, is a good predictor for the actual change in competition occurring after November 2017: the higher is the value of p_i^j, the lower is the expected negative effect of the Settlement Agreement over the hosts' population surrounding listing i. For each distance, all coefficients are positive and significant with a F-statistics above the standard threshold to detect the presence of weak instruments.

Table 4: Impact of the Settlement Agreement on Competition (First Stage)

	$\ln(NL_{i,t}^{0.5})$	$\ln(NL_{i,t}^{1})$	$\ln(NL_{i,t}^{2})$
$p_i^{0.5} \times post_{Nov2017}$	0.715***		
	[0.0458]		
$p_i^{1} \times post_{Nov2017}$		1.002***	
		[0.0370]	
$p_i^{2} \times post_{Nov2017}$			1.173***
			[0.0309]
Constant	4.148***	5.418***	6.665***
	[0.00769]	[0.00534]	[0.00386]
Listing FE	YES	YES	YES
Snap FE	YES	YES	YES
F-test	584.5	905.2	1360.4
R-squared	0.0502	0.0839	0.124
N	296164	297671	297980

Note: Standard errors clustered by listing are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Furthermore, Table 4 shows the economic significance of p_i^j in predicting the number of com-

petitors faced by each listing after November 2017. In particular, when p_i^j passes from 0 to 1, the number of competitors within 0.5 kilometers increases by more than 70%, those within 1 kilometer by more than 100%, and those within 2 kilometers by almost 120%.

Before presenting the results regarding the IV estimates, I show the effect of the expected change in competition due to the regulation (the instrument) over the hosts' ratings about effort. The estimating equation presents the same functional form of Equation 4.1:

$$\ln(\bar{r}_{i,t}^{effort}) = \alpha_i + \gamma_t + \beta p_i^j \times post_{Nov2017} + \varepsilon_{i,t}, \tag{5.2}$$

This equation constitutes the "reduced form" of the IV estimates. Alternatively, it is possible to interpret it as a difference-in-difference design with a continuous control (the variable p_i^j) that defines the extent to which the listing is affected by the regulation, i.e. the listing propensity to be treated by the shock. Table 5 presents the results. For every specification and every rating, a negative and significant relationship between the instrument $p_i^j \times post_{Nov2017}$ and hosts' effort is observed. Accordingly, lower values of p_i^j, that predict a greater drop in the number of competitors for each listing, are associated with higher hosts' effort after November 2017. In this sense, a lower number of competitors is beneficial for hosts' ratings about effort.

Table 5: Impact of the Settlement Agreement on Hosts' Ratings regarding Effort (Reduced Form)

	$\ln(\bar{r}_{i,t}^{comm})$			$\ln(\bar{r}_{i,t}^{check})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$p_i^{0.5} \times post_{Nov2017}$	-0.0301***			-0.0375***		
	[0.00831]			[0.00912]		
$p_i^1 \times post_{Nov2017}$		-0.0475***			-0.0530***	
		[0.0114]			[0.0122]	
$p_i^2 \times post_{Nov2017}$			-0.0432***			-0.0569***
			[0.0131]			[0.0142]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00475	0.00458	0.00493	0.00353	0.00342	0.00342
N	86063	86185	86202	86063	86185	86202

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The parameters are also economically significant: when $p_i^{0.5}$ passes from 0 to 1, $\bar{r}_{i,t}^{comm}$ and $\bar{r}_{i,t}^{check}$ decrease by more than 3%. Drops by more than 4.5% are associated with longer distances p_i^1 and p_i^2. It is important to note that the distributions of $\bar{r}_{i,t}^{effort}$ (presented in Table 2 for listings in Group C) are extremely concentrated and the magnitude of these changes roughly accounts for two third of a standard deviation.

Finally, I turn to the IV estimates. In Equation 4.1 there is one endogenous variable, $\ln(NL_{i,t}^j)$,

and only one IV is derived to predict the impact of the regulation, $p_i^j \times post_{Nov2017}$, the two-stage least squares parameters correspond to the ratio between the coefficients derived before for the "reduced form" and the "first stage" regressions (Equations 5.2 and 5.1, respectively). The estimates are in Table 6. The results show a significant and negative effect of the number of competitors over hosts' effort in line with the parameters of the "reduced form". The negative and significant impact of the IV is in contrast with the OLS estimates where the confounding factors due to demand side lead to inconclusive results.

Moreover, the negative impact of the competition over hosts' effort is in line with the main prediction of the model in Proposition 5. In a less competitive setting, hosts' effort is more beneficial and reputation concerns become more important.

In particular, a 50% decrease in the number of competitors within 0.5 kilometers leads to a increase of more than 1.5% in the rating $\bar{r}_{i,t}^{comm}$, and of more than 2.2% in the rating $\bar{r}_{i,t}^{check}$.[21] As commented before, the distributions of $\bar{r}_{i,t}^{effort}$ are very concentrated and a change of 1% accounts for one fourth of the standard deviation. Interestingly, the magnitude of the parameters does not monotonically increase with the distance; the greatest parameters for both ratings are associated with a distance of 1 kilometers.

Table 6: IV Estimates of the Impact of Competition on Hosts' Ratings regarding Effort

	$\ln(\bar{r}_{i,t}^{comm})$			$\ln(\bar{r}_{i,t}^{check})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(NL_{i,t}^{0.5})$	-0.0362***			-0.0448***		
	[0.0103]			[0.0113]		
$\ln(NL_{i,t}^{1})$		-0.0421***			-0.0468***	
		[0.0104]			[0.0110]	
$\ln(NL_{i,t}^{2})$			-0.0316***			-0.0416***
			[0.00970]			[0.0105]
Constant	2.421***	2.498***	2.482***	2.458***	2.526***	2.550***
	[0.0422]	[0.0555]	[0.0642]	[0.0460]	[0.0590]	[0.0692]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00322	0.00149	0.000595	0.000199	0.000329	0.000479
N	86056	86185	86202	86056	86185	86202

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[21] Figure 2 shows that the Airbnb hosts population dropped by almost 50% during the months in which the Settlement Agreement was implemented.

6 Extensions and Robustness Checks

Here I examine the robustness of the IV estimates. The Section is divided in two parts. In the first part, I provide an estimation of hosts' effort and I show the negative impact of the competition over the estimated hosts' effort. In the second part, I repeat the analysis of Section 5 varying the date of the shock and the timing of the outcome variable. All tables for the second part of the Section are in Appendix.

6.1 Effort Estimation

Submitting reviews, guests answer several questions about their stay. Many dimensions of the lodging service are part of the guests' feedback, and not all regards the effort exerted by hosts during the stay. In Section 5, two rating categories are used as proxies for hosts' effort: check-in and communication. Still, although guests' feedback may be informative about hosts' effort, reviews are also affected by other factors related to guests' characteristics. To account for such confounding factors, I provide here a hosts' effort estimation using a control function approach.

I denote with $\bar{r}_{i,t}^{location}$ the average rating per snapshot for listing i, snapshot t and the category location. Taking advantage of the fact that location should not depend on hosts' effort, in contrast with check-in and communication, I propose the following functional forms:

$$\bar{r}_{i,t}^{location} = \theta_i + guest_{i,t}^{location} \tag{6.1}$$

$$\bar{r}_{i,t}^{effort} = e_{i,t} + guest_{i,t}^{effort}, \tag{6.2}$$

where θ_i is the fixed quality of listing i; $e_{i,t}$ is the effort exerted by the host of listing i at snapshot t; $guest_{i,t}^{location}, guest_{i,t}^{effort}$ account for the guests' specific characteristics about the location and effort such as guests' attitude, tastes or generosity. The control function approach relies on the following equation:

$$guest_{i,t}^{effort} = \alpha + \beta guest_{i,t}^{location} + \epsilon_{i,t}, \tag{6.3}$$

with $E(guest_{i,t}^{location}\epsilon_{i,t}) = 0$. Equation 6.3 assumes a common linear relationship between guests' characteristics for all ratings in the dataset. It allows guests to have different values of $guest_{i,t}^{location}$ and $guest_{i,t}^{effort}$, but a common linear relationship is always present up to the orthogonal error $\epsilon_{i,t}$.[22]

Plugging Equation 6.3 into the previous system of equations, I derive the following fixed effect

[22] The common relationship can be relaxed allowing the parameter β to change over neighborhoods. Still, all results presented in this Section do not change when I allow for different values of β.

panel regression:

$$\begin{aligned}
\bar{r}_{i,t}^{effort} &= e_{i,t} + guest_{i,t}^{effort} \\
&= e_{i,t} + \alpha + \beta guest_{i,t}^{location} + \epsilon_{i,t} \\
&= e_{i,t} + \alpha + \beta(\bar{r}_{i,t}^{location} - \theta_i) + \epsilon_{i,t} \\
\bar{r}_{i,t}^{effort} &= \alpha - \beta\theta_i + \beta\bar{r}_{i,t}^{location} + e_{i,t} + \epsilon_{i,t}.
\end{aligned} \tag{6.4}$$

In Equation 6.4, $\bar{r}_{i,t}^{effort}$ is regressed on $\bar{r}_{i,t}^{location}$ with constant and listing fixed effect accounting for $\alpha - \beta\theta_i$. Accordingly, the host effort $e_{i,t}$ can be estimated from the residuals of fixed effect panel regression with noise $\epsilon_{i,t}$. To have a consistent estimate of β (and unbiased measures of effort), the following orthogonality conditions need to hold:

$$E[\bar{r}_{i,t}^{location}\epsilon_{i,t}|\theta_i] = 0 \tag{OC_1}$$

$$E[\bar{r}_{i,t}^{location}e_{i,t}|\theta_i] = 0. \tag{OC_2}$$

Condition OC_1 directly follows from the assumption 6.3 and the orthogonality of the error $\epsilon_{i,t}$ with $guest_{i,t}^{location}$. Differently, condition OC_2 imposes hosts' effort to not be correlated with deviations of $\bar{r}_{i,t}^{location}$ from the fixed quality θ_i.

I provide empirical evidence supporting condition OC_2 studying the relationship between the effort measures $e_{i,t}^{check}$, $e_{i,t}^{comm}$, the location rating $\bar{r}_{i,t}^{location}$ and a different proxy for hosts' effort present in the dataset: hosts' response rate. This variable represents the percent of new requests the host responded to within 24 hours in the past thirty days.[23] In case of hosts with multiple listings, the variable does not adjust and it considers all messages received by a host. To account for this, I restrict the analysis to single listings, i.e. listings whose hosts do not manage multiple properties on Airbnb.[24] I regress the variable hosts' response rate over the effort measures $e_{i,t}^{check}$, $e_{i,t}^{comm}$, and the location rating $\bar{r}_{i,t}^{location}$ controlling for listing and snapshot fixed effects (see Table 7). The results support condition OC_2: hosts' response rate is not significantly correlated with deviations of $\bar{r}_{i,t}^{location}$, whereas it is positively and significantly correlated with the effort dimensions.

In the remaining part of the Subsection, I study hosts' effort showing the previous results using the effort measures $e_{i,s}^{check}$, $e_{i,s}^{comm}$ estimated as residuals of the regression in Equation 6.4. Still, before studying the relationship between hosts' effort and competition, I present here a result regarding hosts' effort dynamics.

[23]For more information regarding how the response time is computed, see the official Airbnb webpage at www.airbnb.com/help/article/430/what-is-response-rate-and-how-is-it-calculated.

[24]At each snapshot I observe the amount of listings managed by each host. Single listings constitute the 63 percent of total amount of Airbnb listings in the dataset.

Table 7: Evidence Supporting Assumption OC_2: Response Rate, $\bar{r}_{i,s}^{location}$, $e_{i,s}$

	Response rate	Response rate	Response rate
$\bar{r}_{i,t}^{location}$	0.000135		
	[0.000255]		
$e_{i,t}^{comm}$		0.00162***	
		[0.000584]	
$e_{i,t}^{check}$			0.00164**
			[0.000718]
Constant	0.970***	0.971***	0.971***
	[0.00242]	[2.74e-08]	[0.000000562]
Listing FE	YES	YES	YES
R-squared	.0000752	.0002651	.0001957
N	68371	68371	68371

Note: Standard errors clustered by listing are in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.1.1 Effort Dynamics

A common feature of reputation models regards the decreasing pattern of effort exerted by agents over time. Proposition 4 points out this result in the model presented in Section 2: the probability to exert effort for non-commitment hosts weakly decreases from the first to the second period. To support this proposition, I study the relationship between hosts' effort and the number of reviews. In Table 8 I present panel regressions with listings and snapshot fixed effects of $\bar{r}_{i,t}^{comm}$ and $e_{i,t}^{comm}$ over $n_{i,t}$, the total number of reviews written until snapshot t. To remove selection effects, I run three regressions for each dependent variable with different listing samples. The samples are formed by listings with less than 5, 10, and 20 reviews before exiting the platform. The results show a negative and significant relationship between effort and number of reviews in line with Proposition 4. The decreasing pattern is present for the two measures of effort and in every sample. More negative and significant parameters are present for listings with a low number of reviews before the exit. In Appendix, the same regressions are present for $\bar{r}_{i,t}^{check}$ and $e_{i,t}^{check}$ with similar results.

6.1.2 The Impact of Competition over Hosts' Effort

The identification strategy presented in Section 4 can be replicated using $e_{i,t}^{comm}$ and $e_{i,t}^{check}$ as proxies for hosts' effort. Relative to the previous design, the logarithm transformation is not feasible for the estimated effort measures since they are defined as residuals of Equation 6.4 and they may take negative values. Table 13 (in Appendix) presents the OLS panel regressions of hosts' effort and the number of competitors as shown in Equation 4.1. These regressions show non-significant results, similar to the case of ratings $\bar{r}_{i,t}^{effort}$ (Table 3). Demand-driven confounding factors are

Table 8: Effort Dynamics over The Total Number of Reviews

	$\bar{r}_{i,t}^{comm}$			$e_{i,t}^{comm}$		
	>5 reviews	>10 reviews	>20 reviews	>5 reviews	>10 reviews	>20 reviews
$n_{i,t}$	-0.0715***	-0.0275***	-0.00468***	-0.0608***	-0.0234***	-0.00355***
	[0.00834]	[0.00329]	[0.00105]	[0.00804]	[0.00323]	[0.00104]
Constant	9.909***	9.860***	9.810***	0.0985***	0.0775***	0.0216**
	[0.0165]	[0.0152]	[0.0104]	[0.0159]	[0.0149]	[0.0104]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.0167	0.00993	0.00244	0.0135	0.00844	0.00134
N	4469	7789	14342	4469	7789	14342

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

likely at place and endogeneity issues affect the regressions' coefficients. Thus, I estimate the effect of competition over hosts' effort considering variations in the number of competitors due to the Settlement Agreement.

Table 9: Impact of the Settlement Agreement on Hosts' Effort (Reduced Form)

	$e_{i,t}^{comm}$			$e_{i,t}^{check}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$p_i^{0.5} \times post_{Nov2017}$	-0.192**			-0.0962		
	[0.0806]			[0.0869]		
$p_i^1 \times post_{Nov2017}$		-0.247**			-0.150	
		[0.101]			[0.110]	
$p_i^2 \times post_{Nov2017}$			-0.213*			-0.156
			[0.116]			[0.126]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00475	0.00458	0.00493	0.00353	0.00342	0.00342
N	86063	86185	86202	86063	86185	86202

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

First, I present results about the "reduced form" of the IV estimates considering the functional form of Equation 5.2. Table 9 shows the results. The negative relationship between the instrument $p_i^j \times post_{Nov2017}$ and hosts' effort holds for both measures. Still, it is statistically significant only for $e_{i,t}^{comm}$. The economic significance of the relationship also continue to be present. When $p_i^{0.5}$ passes from 0 to 1, $e_{i,t}^{comm}$ decreases by almost 0.2 units; and longer distances, p_i^1 and p_i^2, are associated with drops greater than 0.2 units. It is important to recall that, because of its nature of residuals, $e_{i,t}^{comm}$ has a zero sample mean with standard deviation equal to 0.64. Thus, a change of 0.2 accounts for almost one third of the standard deviation.

Table 10: IV Estimates of the Impact of Competition on Hosts' Effort

	$e_{i,t}^{comm}$			$e_{i,t}^{check}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(NL_{i,t}^{0.5})$	-0.230**			-0.115		
	[0.0973]			[0.104]		
$\ln(NL_{i,t}^{1})$		-0.218**			-0.133	
		[0.0893]			[0.0977]	
$\ln(NL_{i,t}^{2})$			-0.156*			-0.114
			[0.0855]			[0.0925]
Constant	0.873**	1.103**	0.962*	0.386	0.630	0.674
	[0.396]	[0.479]	[0.565]	[0.426]	[0.523]	[0.611]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00322	0.00149	0.000595	0.000199	0.000329	0.000479
N	86056	86185	86202	86056	86185	86202

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Similar results characterize the IV estimates, presented in Table 10.[25] While $e_{i,t}^{check}$ is not statistically affected by the degree of competition, a negative relationship between the number of competitors and hosts' effort is present for $e_{i,t}^{comm}$. In terms of economic significance, a decrease of 50% in the number of competitors within 0.5 kilometers leads to an increase in $e_{i,t}^{comm}$ of more than 0.11 units.

Accordingly, the negative relationship between number of competitors and hosts' effort partially holds even after removing confounding factors due to guests' effect. In particular, the statistical significance of this relationship is limited to the dimension regarding communication.

6.2 Robustness Checks

The first robustness check enlarges the sample used for the analysis. In Section 3 I claim that the registration enforcement due to the Settlement Agreement could have change the profile of Airbnb hosts after November 2017. Table 2 confirms it: listings that exit before the Settlement Agreement differ in many characteristics relative to those that are present on the platform before and after the change in regulation. Thus, the results in Section 5 regards only listings that stay on the platform after the end of the process of registration enforcement in January 2018. Table 14 shows the IV estimates enlarging the sample to all listings that are present on the platform after September 2017 (Group B, Group C, and Group D in Section 3). Results are robust with a negative and significant effect of competition over hosts' effort. The size of the impact is larger when all the sample of

[25] As before, the two-stage least squares parameters correspond to the ratio between the coefficient derived for the "reduced form" and the "first stage".

listings is considered. This suggests that the Settlement Agreement positively selects hosts in terms of their propensity to exert effort.

The second robustness dimension that I investigate regards the time lag that exists between the host's effort and the guest's reviews. The Airbnb review system allows users to write feedback in a period of 15 days after the end of the guest's stay. The reviews (comments and ratings) are displayed on the users' webpages as soon as both parties review the other in order to reduce potential retaliation from the hosts' side (Fradkin et al., 2018). Accordingly, reviews describe hosts' effort exerted days before the date of display on the webpages. This lag should not be of a particular concern since hosts may anticipate changes in the degree of competition due to the regulation. Results in Section 5 show that hosts increase effort with a certain degree of anticipation. Table 15 presents the IV estimates of the impact of the degree in competition at snapshot t and the hosts' effort displayed at snapshot $t+1$. In this sense, the hosts' effort could be contemporaneous with the observed degree in competition. Results are robust with negative and significant effect of competition over hosts' effort. The magnitude of the parameters is similar. This is in line with the hosts anticipating the change and exerting more effort.

The final robustness check is about the choice of snapshot in which the Settlement Agreement starts to be effective. In the main analysis I consider November 2017 as the starting snapshot since Figure 2 shows that the number of listings starts to decrease from this month. Still, at the snapshot of October 2017, the Settlement Agreement was already in place, while further drops occur the next months. In order to show the robustness of the IV estimates regarding the starting snapshot of the regulation, I present in Table 16 the results of "first stage" regressions in which the starting snapshot for the registration enforcement is moved in October 2017 and in December 2017. In both cases, the results are robust. Then, Tables 17 and 18 present the IV estimates with the dummy variable *post* taking value 1 since October 2017 and since December 2017, respectively. Results are robust with negative and significant effect of competition over hosts' effort. The size of the effect is larger if I consider October 2017, suggesting again the anticipation behavior by hosts regarding the future movement in the number of competitors.

7 Conclusion

In this work, I provide theoretical and empirical evidence regarding the relationship between competition and reputation concerns. The existence of a trade-off between the number of competitors and the incentives to exert effort is shown using a model of reputation concerns. First, I develop a reputation model in a directed search framework where movements in entry costs impact the number of agents in the market and their incentives to exert effort. Then, using a unique dataset of

Airbnb, I identify the causal relationship between the number of competitors on the platform and hosts' effort. To do this, I consider a change in the regulation regarding the registration enforcement of Airbnb hosts in San Francisco in September 2017. I obtain a negative and significant effect of the extent of competition over hosts' effort. All the empirical results are in line with the main predictions of the model.

This work is still preliminary and incomplete. In the next months, I will focus on the impact of the Settlement Agreement over the hosts' selection. In particular, I will provide more details on the differences among hosts who continue to be present on Airbnb after January 2018 and those who exit during the period of the registration enforcement (from September 2017 to January 2018). Moreover, I will study how the behavior of Airbnb hosts is affected by changes in the number of competitors with different ratings.

The main limitation of my work regards the structure of the dataset and the available pieces of information concerning transactions and effort. All the proxies that I use to estimate hosts' effort are extracted from the Airbnb feedback system. In this sense, my analysis considers reported effort. Similarly, only part of the total transactions are studied since my dataset includes only transactions that are reviewed by guests. When a guest decides to not report any feedback, the effort exerted for that transaction cannot be estimated.

A second limitation is due to the nature of the main variable of interest. I observe the average ratings and not the rating of each single review. This leads to measurement errors, that only partially are removed thanks to the IV identification strategy. In the future, I will tackle this problem by analyzing all the guests' comments displayed on the hosts' webpages. I am currently working on a sentiment analysis of the comments that could significantly amplify the ratings' variation and overcome the limitation due to the measurement errors of the averages.

From a policy perspective, the results of my work are in favor of regulations that partially restrict the market access. Yet, my analysis assumes that entry costs do not completely block entry and the disciplining force of competition is not altered. Further investigation regarding the optimal size of entry costs is necessary to evaluate proper regulatory policies.

References

Ahern, K. R. and Dittmar, A. K. (2012). The Changing of the Boards: The Impact on Firm Valuation of Mandated Female Board Representation. *The Quarterly Journal of Economics*, 127(1):137–197.

Ashenfelter, O. C., Hosken, D. S., and Weinberg, M. C. (2015). Efficiencies brewed: Pricing and consolidation in the US beer industry. *RAND Journal of Economics*, 46(2):328–361.

Bar-Isaac, H. (2005). Imperfect competition and reputational commitment. *Economics Letters*, 89(2):167–173.

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

Burdett, K., Coles, M., Kiyotaki, N., and Wright, R. (1995). Buyers and sellers: should I stay or should I go? *The American Economic Review*, 85(2):281–286.

Burdett, K., Shi, S., and Wright, R. (2001). Pricing and Matching with Frictions. *Journal of Political Economy*, 109(5):1060–1085.

Cabral, L. and Hortaçsu, A. (2010). The Dynamics of Seller Reputation: Evidence from Ebay. *The Journal of Industrial Economics*, LVIII(1):54–78.

Chandra, A. and Weinberg, M. (2017). How Does Advertising Depend on Competition? Evidence from U.S. Brewing.

Cho, I.-K. and Kreps, D. M. (1987). Signaling Games and Stable Equilibria. *The Quarterly Journal of Economics*, 102(2):179–221.

Dafny, B. L., Duggan, M., and Ramanarayanan, S. (2012). Paying a Premium on Your Premium ? Consolidation in the US Health Insurance Industry. *American Economic Review*, 102(2):1161–1185.

Delacroix, A. and Shi, S. (2013). Pricing and signaling with frictions. *Journal of Economic Theory*, 148(4):1301–1332.

Fan, Y., Ju, J., and Xiao, M. (2016). International Journal of Industrial Organization Reputation premium and reputation management : Evidence from the largest e-commerce platform in China R. 46:63–76.

Fradkin, A. (2017). Search, Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb.

Fradkin, A., Grewal, E., and Holtz, D. (2018). The Determinants of Online Review Informativeness: Evidence from Field Experiments on Airbnb.

Hörner, J. (2002). Reputation and competition. *American Economic Review*, 92(3):644–663.

Hosios, A. J. (1990). On the Efficiency of Matching and Related Models of Search and Unemployment. *The Review of Economic Studies*, 57(2):279–298.

Hui, X., Saeedi, M., Shen, Z., and Sundaresan, N. (2016). Reputation and Regulations: Evidence from eBay. *Management Science*, 62(12):3604–3616.

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

Klein, T. J., Lambertz, C., and Stahl, K. O. (2016). Market Transparency, Adverse Selection, and Moral Hazard. *Journal of Political Economy*, 124(6):1677–1713.

Meng, L., Brynjolfsson, E., and Dowlatabadi, J. (2018). Do Digital Platforms Reduce Moral Hazard? The Case of Taxis and Uber.

Peters, M. (1991). Ex ante price offers in matching games non-steady states. *Econometrica: Journal of the Econometric Society*, pages 1425–1454.

Proserpio, D., Xu, W., and Zervas, G. (2018). You get what you give: theory and evidence of reciprocity in the sharing economy. *Quantitative Marketing and Economics*, 16(4):371–407.

Shi, S. and Delacroix, A. (2018). Should Buyers or Sellers Organize Trade In a Frictional Market? *The Quarterly Journal of Economics*, 1:44.

Stevenson, B., Freeman, R., Goldin, C., Minter, C., Katz, L., Madrian, B., and Wolfers, J. (2010). Beyond the Classroom: Using Title IX to Measure the Return to High School Sports. *The Review of Economics and Statistics*, 92(May):284–301.

Wright, R., Kircher, P., Julîen, B., and Guerrieri, V. (2017). Directed Search: A Guided Tour. Technical report.

Zervas, G., Proserpio, D., and Byers, J. (2015). A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average.

Appendix

A Model

Here I provide proofs of the Propositions and Theorems discussed in Section 2. Before doing that, I briefly discuss the hosts optimal pricing if the cost of effort becomes public information after being drawn by hosts. I show that this allocation may not be sustained when the cost of effort is hosts' private information. Then, in the context of asymmetry of information, I characterize non-reputational equilibria with separating strategies in prices pointing out the additional conditions that are necessary for their existence. Finally, I characterize the constrained efficient reputational equilibrium allocation and I proceed with the proofs.

A.1 Perfect Information

With public information about hosts' cost of effort, effort exerted in period 1 and he price posted in period 2 do not impact guests' beliefs. In period 2, the problem for hosts who draw $c = 0$ is defined as follows:

$$\begin{aligned} \max_{p_2} \ & p_2 \alpha(\theta_2) \\ s.t. \ & (a + b - p_2)\frac{\alpha(\theta_2)}{\theta_2} = U_2, \end{aligned}$$

where U_2 is the guests' expected utility for a match with hosts in period 2. Accordingly, the optimal price and tightness for hosts with $c = 0$, p_2^0, θ_2^0 are defined in terms of U_2. If $a + b < U_2$, then $\theta_2^0 = 0$ and $p_2^0 = 0$. If $a + b \geq U_2$:

$$\begin{aligned} \alpha'(\theta_2^0) &= \frac{U_2}{a+b} \\ p_2^0 &= a + b - \frac{\theta_2^0}{\alpha(\theta_2^0)} U_2. \end{aligned} \tag{A.1}$$

Thus, the expected profit with public information for hosts with $c = 0$ is defined as follows:

$$\Pi_2(a + b) = (a + b)(\alpha(\theta_2^0) - \alpha'(\theta_2^0)\theta_2^0), \tag{A.2}$$

where θ_2^0 is derived by Equation A.1. The expected profit is increasing in the guests' surplus of transactions $(a+b)$ if $U_2 \le a+b$:

$$\begin{aligned}
\frac{\partial \Pi_2(a+b)}{\partial(a+b)} &= \alpha(\theta_2^0) - \alpha'(\theta_2^0)\theta_2^0 + (a+b)\frac{\partial(\alpha(\theta_2^0) - \alpha'(\theta_2^0)\theta_2^0)}{\partial\theta_2^0}\frac{\partial\theta_2^0}{\partial(a+b)} \\
&= \alpha(\theta_2^0) - \alpha'(\theta_2^0)\theta_2^0 - (a+b)\alpha''(\theta_2^0)\theta_2^0\frac{\partial\theta_2^0}{\partial(a+b)} \\
&= \alpha(\theta_2^0) - \alpha'(\theta_2^0)\theta_2^0 + (a+b)\alpha''(\theta_2^0)\theta_2^0\frac{1}{\alpha''(\theta_2^0)}\frac{U_2}{(a+b)^2} \\
&= \alpha(\theta_2^0) > 0,
\end{aligned}$$

where the third passage directly follows from the properties of the derivative of the inverse function.[26] Conversely, the expected profit is decreasing in U_2 if $U_2 \le a+b$:

$$\begin{aligned}
\frac{\partial \Pi_2(a+b)}{\partial U_2} &= (a+b)\frac{\partial(\alpha(\theta_2^0) - \alpha'(\theta_2^0)\theta_2^0)}{\partial\theta_2^0}\frac{\partial\theta_2^0}{\partial U_2} \\
&= -(a+b)\alpha''(\theta_2^0)\theta_2^0\frac{\partial\theta_2^0}{\partial U_2} \\
&= -(a+b)\alpha''(\theta_2^0)\theta_2^0\frac{1}{\alpha''(\theta_2^0)}\frac{1}{(a+b)} \\
&= -\theta_2^0 < 0.
\end{aligned}$$

Similarly, in period 2, the problem for hosts who draw $c = k > 0$ is defined as follows:

$$\begin{aligned}
&\max_{p_2}\ p_2\alpha(\theta_2) \\
&s.t.\ (b - p_2)\frac{\alpha(\theta_2)}{\theta_2} = U_2.
\end{aligned} \tag{A.3}$$

If $b < U_2$, then the optimal price and tightness for hosts with cost of effort $c = k > 0$ p_2^k, θ_2^k are $\theta_2^k = 0$ and $p_2^k = 0$. If $b \ge U_2$:

$$\begin{aligned}
\alpha'(\theta_2^k) &= \frac{U_2}{b} \\
p_2^k &= b - \frac{\theta_2^k}{\alpha(\theta_2^k)}U_2.
\end{aligned} \tag{A.4}$$

Thus, the expected profit with public information for hosts with cost of effort $c = k > 0$ is defined as follows:

$$\Pi_2(b) = b(\alpha(\theta_2^k) - \alpha'(\theta_2^k)\theta_2^k), \tag{A.5}$$

[26]Recall that the first derivative of the fuction α is invertible by Assumption 1.

where θ_2^k is defined by Equation A.4. Similarly to the case of hosts with cost of effort $c = 0$, the expected profit is increasing in b and decreasing in U_2 if $b \geq U_2$:

$$\frac{\partial \Pi_2(b)}{\partial(b)} = \alpha(\theta_2^k) > 0$$
$$\frac{\partial \Pi_2(b)}{\partial U_2} = -\theta_2^k < 0.$$

Hosts who did not match with guests in period 1 do not draw their cost of effort and, together with new arrivals solve the following problem in period 2.

$$\max_{p_2} \; p_2\alpha(\theta_2) \qquad \text{(A.6)}$$
$$s.t. \;\; (a\pi + b - p_2)\frac{\alpha(\theta_2)}{\theta_2} = U_2.$$

The ex-ante guests' utility from a transaction with this class of hosts is $a\pi + b$ since, with probability π hosts draw zero cost of effort and then they exert effort $e_2 = 1$. Otherwise, with probability $1 - \pi$ they draw positive cost of effort and they have no incentives to exert effort: $e_2 = 0$. Accordingly, if $a\pi + b < U_2$, the optimal price and tightness for these hosts, $p_2^{\emptyset}, \theta_2^{\emptyset}$ are $\theta_2^{\emptyset} = 0$ and $p_2^{\emptyset} = 0$. If $a\pi + b \geq U_2$:

$$\alpha'(\theta_2^{\emptyset}) = \frac{U_2}{a\pi + b} \qquad \text{(A.7)}$$
$$p_2^{\emptyset} = a\pi + b - \frac{\theta_2^{\emptyset}}{\alpha(\theta_2^{\emptyset})}U_2.$$

Thus, the expected profit for hosts who did not match with guests in period 1 and for new arrivals is defined as follows:

$$\Pi_2(a\pi + b) = (a\pi + b)(\alpha(\theta_2^{\emptyset}) - \alpha'(\theta_2^{\emptyset})\theta_2^{\emptyset}), \qquad \text{(A.8)}$$

where $\theta_2^{\emptyset}$ is defined by Equation A.7. In period 1, hosts have not yet drawn their cost of effort and their problem is the following:

$$\max_{p_1} \; p_1\alpha(\theta_1) + \beta\alpha(\theta_1)(\pi\Pi_2(a + b) + (1 - \pi)\Pi_2(b)) + \beta(1 - \alpha(\theta_1))\Pi_2(a\pi + b) \qquad \text{(A.9)}$$
$$s.t. \;\; (a\pi + b - p_1)\frac{\alpha(\theta_1)}{\theta_1} = U_1.$$

The ex-ante guests' utility from a transaction in period 1 is $a\pi + b$ since hosts who draw positive cost of effort have no incentives to exert effort: their cost of effort is public information and they cannot commit to exert effort in period 2. Thus, if $a\pi + b + \beta(\pi\Pi_2(a + b) + (1 - \pi)\Pi_2(b) - \Pi_2(a\pi + b)) < U_1$,

the optimal price and tightness for these hosts, $p_1^{\emptyset}, \theta_1^{\emptyset}$ are $\theta_1^{\emptyset} = 0$ and $p_1^{\emptyset} = 0$. Otherwise:

$$\alpha'(\theta_1^{\emptyset}) = \frac{U_1}{a\pi + b + \beta(\pi\Pi_2(a+b) + (1-\pi)\Pi_2(b) - \Pi_2(a\pi + b))} \tag{A.10}$$
$$p_1^{\emptyset} = a\pi + b - \frac{\theta_1^{\emptyset}}{\alpha(\theta_1^{\emptyset})}U_1.$$

If the cost of effort is hosts' private information, the equilibrium above may not be sustained. Hosts are better-off posting p_2^0 relative to p_2^k. This follows since U_2 is the same for hosts with different cost of effort. Thus, $\alpha'(\theta_2^k) < \alpha'(\theta_2^0)$ from Equations A.1 and A.4. Then, by the concavity of α, $\theta_2^0 > \theta_2^k$ and $\alpha(\theta_2^0) > \alpha(\theta_2^k)$: hosts with $c = 0$ have higher chances to be matched with guests relative to hosts with $c = k$. Hence, hosts are better-off posting p_2^0 with expected profits equal to $p_2^0\alpha(\theta_2^0)$:

$$\begin{aligned} p_2^0\alpha(\theta_2^0) &= (a+b)(\alpha(\theta_2^0) - \alpha'(\theta_2^0)\theta_2^0) \\ &> b(\alpha(\theta_2^0) - \alpha'(\theta_2^0)\theta_2^0) > b(\alpha(\theta_2^k) - \alpha'(\theta_2^k)\theta_2^k) = p_2^k\alpha(\theta_2^k), \end{aligned}$$

where the inequality in the last passage in due to Assumption 1: $\alpha(\theta) - \alpha'(\theta)\theta > 0$ $\forall\theta$ and $\frac{\partial(\alpha(\theta)-\alpha'(\theta)\theta)}{\partial\theta} = -\alpha''(\theta)\theta > 0$.

Accordingly, the perfect information equilibrium may not be sustained if the cost of effort is hosts' private information. In particular, if the following condition holds, hosts who draw cost $c = k > 0$ are willing to exert effort, incur in cost k and obtain future expected profits $p_2^0\alpha(\theta_2^0)$:

$$\beta(p_2^0\alpha(\theta_2^0) - p_2^k\alpha(\theta_2^k)) \geq c. \tag{A.11}$$

If the condition in A.11 does not hold, then the perfect information equilibrium allocation can be sustained. In the next Section, I will show how this allocation is a particular case of reputational equilibrium.

A.2 Non-Reputational Equilibria

In this paper I focus on reputational equilibria where the information provided by hosts' histories is not made ineffective by the prices posted in period 2. Here I briefly discuss non-reputational equilibria.

As mentioned in the main text, in non-reputational equilibria, hosts with different cost of effort play separate pricing strategies in period 2. Accordingly, guests can perfectly infer hosts' cost of effort observing period 2 prices irrespectively of hosts' histories. In equilibrium, hosts who draw cost $c = k > 0$ post in period 2 the perfect information price p_2^k. Differently, hosts who

draw cost $c = 0$ post in period 2 price $p_2^{sep} > 0$ such that hosts with cost $c = k > 0$ are better-off posting p_2^k. In this sense, the existence of non-reputational equilibria relies on the fact that the profit $p_2^k \alpha(\theta_2^k)$ is strictly positive, that is $b > U_2$. If this condition holds, then the following two incentive compatibility constraints have to be satisfied:

$$
\begin{aligned}
p_2^k \alpha(\theta_2^k) &\geq p_2^{sep} \alpha(\theta_2^{sep}) \\
p_2^{sep} \alpha(\theta_2^{sep}) &\geq p_2^k \alpha(\theta_2^k).
\end{aligned}
$$

The ex-ante utility for guests who are matched with hosts posting p_2^k is b; whereas, the utility for those matched with hosts posting p_2^{sep} is $a + b$:

$$
(b - p_2^k)\frac{\alpha(\theta_2^k)}{\theta_2^k} = U_2 = (a + b - p_2^{sep})\frac{\alpha(\theta_2^{sep})}{\theta_2^{sep}}. \tag{A.12}
$$

From the incentive compatibility constraints it results that, when host with $c = 0$ separate, they do not increase their expected profits since $p_2^k \alpha(\theta_2^k) = p_2^{sep} \alpha(\theta_2^{sep})$. In particular, from Equation A.12, $p_2^{sep} > p_2^k$ and $\theta_2^k > \theta_2^{sep}$. Accordingly, the existence of this equilibrium relies on hosts' willingness to separate even when their expected profits do not increase after separating.

A.3 Reputational Equilibria

Here I provide the proofs of the Propositions and Theorems discussed in Section 2. At the same time, I illustrate the constrained efficient allocation and I show that the prices posted by hosts in the equilibrium respect the Hosios (1990) conditions.

The constrained efficient allocation is the allocation that a benevolent social planner would choose taking as given the following elements:

- the frictions that characterize the matching between hosts and guests;
- the hosts' entry cost f;
- the hosts' private information concerning the cost of effort.

Accordingly, the social planner aims at allocating guests to hosts in order to implement the efficient hosts' entry and effort provision.

In line with the main text, I start my analysis from period 2.

A.3.1 Period 2

In order to implement the efficient hosts' entry in period 2, the social planner faces the following problem:

$$\max_{\theta_2} \ (a\pi + b)\frac{\alpha(\theta_2)}{\theta_2} - \frac{f}{\theta_2}.$$

The factor $(a\pi + b)\frac{\alpha(\theta_2)}{\theta_2}$ represents the expected surplus from a transaction for each guest, whereas $\frac{f}{\theta_2}$ defines the hosts' entry costs for each guest. The optimal θ_2^* that maximizes the social planner objective function is such that:

$$(a\pi + b)(\alpha(\theta_2^*) - \alpha'(\theta_2^*)\theta_2^*) = f. \tag{A.13}$$

It is possible to note that the optimal price posted by hosts who enter the platform in period 2, $p_2^\emptyset$ implements the efficient entry condition of period 2 when the hosts' free entry condition is binding. The optimal expected profits for new entrant hosts is defined by Equation A.8 and it equals the LHS of Equation A.13. Accordingly, the latter condition equalizes the optimal expected profits for new entrant hosts to the entry costs f: i.e. it imposes a binding free entry condition for entrant hosts in period 2. The rule proposed by Hosios (1990) states that hosts' entry is constrained efficient when the two sides of the market share the ex-ante surplus of transactions (in this case $a\pi + b$) according to the elasticity of the matching function with respect to the tightness. In fact, in this case the expected profits for new entrant hosts, $\Pi_2(a\pi + b)$, and the guests' expected utility, U_2, are defined as follows:

$$\Pi_2(a\pi + b) = p_2^\emptyset \alpha(\theta_2^\emptyset) = (a\pi + b)(1 - \epsilon_2^\emptyset)\alpha(\theta_2^\emptyset)$$
$$U_2 = (a\pi + b)\epsilon_2^\emptyset \frac{\alpha(\theta_2^\emptyset)}{\theta_2^\emptyset},$$

where $\epsilon_2^\emptyset = \alpha'(\theta_2^\emptyset)\frac{\theta_2^\emptyset}{\alpha(\theta_2^\emptyset)}$ denotes the elasticity of the matching function with respect to the tightness calculated at $\theta_2^\emptyset$.

I analyze the constrained efficient allocation for hosts who enter in period 1 in the next session. In period 2 they post prices to maximize their profits given guests' beliefs $\bar{\mu}_2(h)$ and U_2.

Proof of Proposition 1. Assuming that guests do not update their beliefs about hosts' cost of effort after observing prices in period 2, hosts who were matched in period 1 and with history h solve the

following problem in period 2:

$$\max_{p_2} \ p_2\alpha(\theta_2) \tag{A.14}$$
$$s.t. \ (a\bar{\mu}_2(h) + b - p_2)\frac{\alpha(\theta_2)}{\theta_2} = U_2.$$

If $a\bar{\mu}_2(h) + b < U_2$, then the optimal price and tightness $p_2^{pool}(h), \theta_2^{pool}(h)$ are $\theta_2^{pool}(h) = 0$ and $p_2^{pool}(h) = 0$. If $a\bar{\mu}_2(h) + b \geq U_2$:

$$\alpha'(\theta_2^{pool}(h)) = \frac{U_2}{a\bar{\mu}_2(h) + b} \tag{A.15}$$
$$p_2^{pool}(h) = a\bar{\mu}_2(h) + b - \frac{\theta_2^{pool}(h)}{\alpha(\theta_2^{pool}(h))}U_2.$$

Similarly, hosts who were not matched in period 1 and new entrants solve the problem in Equation A.6 and their optimal price and tightness $p_2^{\emptyset}, \theta_2^{\emptyset}$ are reported in Equation A.7. □

It is possible to note that the optimal prices for hosts who enter in period 1 follow Hosios (1990) conditions since hosts and guests share the ex-ante surplus $a\bar{\mu}_2(h) + b$ according to the elasticity of the matching function. In particular:

$$\Pi_2(a\bar{\mu}_2(h) + b) = p_2^{pool}(h)\alpha(\theta_2^{pool}(h)) = (a\bar{\mu}_2(h) + b)(1 - \epsilon_2^{pool}(h))\alpha(\theta_2^{pool}(h))$$
$$U_2 = (a\bar{\mu}_2(h) + b)\epsilon_2^{pool}(h)\frac{\alpha(\theta_2^{pool}(h))}{\theta_2^{pool}(h)},$$

where $\epsilon_2^{pool}(h) = \alpha'(\theta_2^{pool}(h))\frac{\theta_2^{pool}(h)}{\alpha(\theta_2^{pool}(h))}$ denotes the elasticity of the matching function with respect to the tightness calculated at $\theta_2^{pool}(h)$. Accordingly, the ex-ante surplus of transactions in period 2 is greater for hosts who exert effort in period 1. Furthermore, hosts who exert effort in period 1 get a greater share of the surplus since the elasticity $\epsilon_2^{pool}(h)$ is decreasing in the tightness and $\theta_2^{pool}(h) \geq \theta_2^k \ \forall \bar{\mu}_2(h) > 0$. In this sense, in order to increase the effort provision (in period 1) and obtain the efficient hosts' entry in period 2, the social planner may commit to allocate guests to hosts in period 2 such that the tightness levels $\theta_2^{pool}(h), \theta_2^{\emptyset}$ are formed.

A.3.2 Period 1

In line with the analysis in the main text, I start with the proof of Proposition 2 regarding the effort provision in period 1. Then, I provide the proof for Proposition 3 and I characterize the constrained efficient allocation in period 1.

Proof of Proposition 2. The effort strategy by hosts with $c = k > 0$ in period 1 realizes the interest of these hosts to mimic hosts with $c = 0$, post higher prices and attract more guests as it has been observed in Proposition 1. In particular, exerting $e_1 = 1$, hosts with $c = k > 0$ pool together with hosts with $c = 0$ in period 2, posting $p_2^{pool}(p_1^{\emptyset}, e_1 = 1)$. Exerting $e = 0$, with $c = k > 0$ cannot pool anymore since their history is fully revealing their costs. The value of pooling depends on the guests' interim beliefs. They are derived by the Bayes formula when possible:

$$\bar{\mu}_2(p_1^{\emptyset}, e_1 = 1) = \frac{\pi}{\pi + (1-\pi)\omega} \tag{A.16}$$

$$\bar{\mu}_2(p_1^{\emptyset}, \emptyset) = \mu_2(\emptyset) = \pi \tag{A.17}$$

$$\bar{\mu}_2(p_1^{\emptyset}, e_1 = 0) = 0, \tag{A.18}$$

where $\omega \in [0, 1]$ is the probability to exert effort $e_1 = 1$ by hosts with $c = k > 0$ in equilibrium in period 1. In this sense, the discounted marginal benefits of exerting effort for non-commitment types are defined as follows:

$$MB = \beta(p_2^{pool}(p_1^{\emptyset}, e_1 = 1)\alpha(\theta_2^{pool}(p_1^{\emptyset}, e_1 = 1)) - (p_2^{pool}(p_1^{\emptyset}, e_1 = 0)\alpha(\theta_2^{pool}(p_1^{\emptyset}, e_1 = 0))).$$

Recalling the function $\Pi_2(.)$ introduced in the previous Section A.1, the discounted marginal benefits can be defined as follows:

$$MB = \beta\Big(\Pi_2\big(a\frac{\pi}{\pi + (1-\pi)\omega} + b\big) - \Pi_2(b)\Big),$$

where the function $\Pi_2(.)$ is weakly increasing in the value of $a\mu_2(h) + b$. Hosts with $c = k > 0$ compare MB with the cost of effort k. The following algorithm characterizes the equilibrium level of ω:

1. Consider the case $\omega = 1$ and calculate the MB. If the MB is greater than k:

   $$\beta(\Pi_2\big(a\pi + b\big) - \Pi_2(b) \geq k,$$

   then, in equilibrium, hosts with $c = k > 0$ exert effort in period 1 with probability $\omega = 1$;

2. If the inequality above does not hold true, then consider the case $\omega = 0$ and calculate again the MB. If the MB is lower than k:

   $$\beta(\Pi_2\big(a + b\big) - \Pi_2(b)) \leq k,$$

   then, in equilibrium, hosts with $c = k > 0$ exert effort in period 1 with probability $\omega = 0$;

3. If the two inequalities above do not hold true, then derive $\omega \in (0,1)$ such that the following equality holds:

$$\beta\Big(\Pi_2\big(a\frac{\pi}{\pi+(1-\pi)\omega}+b\big)-\Pi_2(b)\Big)=k. \tag{A.19}$$

Since the *LHS* of Equation A.19 is strictly decreasing in ω, Equation A.19 admits only one solution.

$\square$

Proof of Proposition 3. Hosts who enter in period 1 have not yet drawn their cost of effort. Thus, their problem in period 1 is the following:

$$\begin{aligned}
\max_{p_1}\ & (p_1-k(1-\pi)\omega)\alpha(\theta_1)+\beta\alpha(\theta_1)(\pi\Pi_2\big(a\frac{\pi}{\pi+(1-\pi)\omega}+b\big)+(1-\pi)(1-\omega)\Pi_2(b))\\
&+\beta(1-\alpha(\theta_1))\Pi_2(a\pi+b)\\
s.t.\ & (a(\pi+(1-\pi)\omega)+b-p_1)\frac{\alpha(\theta_1)}{\theta_1}=U_1.
\end{aligned}$$

The ex-ante guests' utility from a transaction in period 1 is $a(\pi+(1-\pi)\omega)+b$ since hosts who draw positive cost of effort exert effort in period 1 with probability ω. Thus, if $a(\pi+(1-\pi)\omega)+b-k(1-\pi)\omega+\beta\Delta\Pi<U_1$, the optimal price and tightness for these hosts, $p_1^{\emptyset},\theta_1^{\emptyset}$ are $\theta_1^{\emptyset}=0$ and $p_1^{\emptyset}=0$. Otherwise:

$$\begin{aligned}
\alpha'(\theta_1^{\emptyset}) &= \frac{U_1}{a(\pi+(1-\pi)\omega)+b-k(1-\pi)\omega+\beta\Delta\Pi}\\
p_1^{\emptyset} &= a(\pi+(1-\pi)\omega)+b-\frac{\theta_1^{\emptyset}}{\alpha(\theta_1^{\emptyset})}U_1.
\end{aligned}$$

In this sense, if $a(\pi+(1-\pi)\omega)+b-k(1-\pi)\omega+\beta\Delta\Pi>U_1$, the expected profits for new entrants in period 1 is defined as follows:

$$\Pi_1=(a(\pi+\omega(1-\pi))+b-k(1-\pi)\omega+\beta\Delta\Pi)(\alpha(\theta_1^{\emptyset})-\alpha'(\theta_1^{\emptyset})\theta_1^{\emptyset})+\beta\Pi_2(a\pi+b). \tag{A.20}$$

$\square$

The constrained efficient allocation in period 1 implies the efficient hosts' entry and effort provision in period 1. Accordingly, the social planner commits to allocate guests to hosts in period 2 in order to form the tightness levels $\theta_2^{pool}(h),\theta_2^{\emptyset}$. In this sense, hosts who draw cost $c=k>0$ in period 1 have incentives to exert effort with probability ω in line with Proposition 2. Therefore,

the social planner solves the following problem in period 1:

$$\max_{\theta_1} \; (a(\pi+\omega(1-\pi))+b-k(1-\pi)\omega)\frac{\alpha(\theta_1)}{\theta_1}+\frac{R}{\theta_1}-\frac{f}{\theta_1}.$$

Similarly to period 2, the factor $(a(\pi+\omega(1-\pi))+b-k(1-\pi)\omega)\frac{\alpha(\theta_1)}{\theta_1}$ represents the expected surplus from a transaction for each guest, whereas $\frac{f}{\theta_1}$ defines the hosts' entry costs for each guest. Yet, in period 1, an additional element forms the surplus of a transaction. The factor R captures the value of a transaction in updating hosts' reputation and changing the ex-ante surplus of transactions in period 2:

$$R=\beta\alpha(\theta_1)(\pi\Pi_2\big(a\frac{\pi}{\pi+(1-\pi)\omega}+b\big)+(1-\pi)(1-\omega)\Pi_2(b))+\beta(1-\alpha(\theta_1))\Pi_2(a\pi+b).$$

The optimal θ_1^* that maximizes the social planner objective function is such that:

$$(a(\pi+\omega(1-\pi))+b-k(1-\pi)\omega+\beta\Delta\Pi)(\alpha(\theta_1^*)-\alpha'(\theta_1^*)\theta_1^*)+\beta\Pi_2(a\pi+b)=f. \tag{A.21}$$

It is possible to note that the optimal price posted by hosts who enter the platform in period 1, $p_1^{\emptyset}$ implements the efficient entry condition of period 1 when the hosts' free entry condition is binding. The optimal expected profits for new entrant hosts is defined by Equation A.20 and it equals the LHS of Equation A.21. Accordingly, the latter condition equalizes the optimal expected profits for new entrant hosts to the entry costs f: i.e. it imposes a binding free entry condition for entrant hosts in period 1.

A.3.3 Existence and Uniqueness

The proof of Theorem 1 has the following structure: first, I assume that a positive measure of hosts enter in period 2. With this assumption, I show the existence and the uniqueness of the equilibrium and I derive the threshold level $\bar{G}$ such that there is entry in the second period for $G>\bar{G}$.

Proof of Theorem 1. With a positive measure of hosts entering in period 2, the free entry condition for hosts in period 2 holds with equality. From the free entry condition and the optimal pricing for new entrants derived by Proposition 1, it is possible to uniquely determine $p_2^{\emptyset}, \theta_2^{\emptyset}$ and U_2. Accordingly, the free entry condition can be written in terms of $\theta_2^{\emptyset}$:

$$(a\pi+b)(\alpha(\theta_2^{\emptyset})-\theta_2^{\emptyset}\alpha'(\theta_2^{\emptyset}))=f. \tag{A.22}$$

From Equation A.22, the equilibrium value of $\theta_2^{\emptyset}$ can be uniquely derived. Recall that $a\pi+b>0$, $\alpha''(\theta)<0$, and $\alpha(\theta)-\theta\alpha'(\theta)>0\;\forall\theta$. Moreover, $\alpha(\theta)-\theta\alpha'(\theta)$ is strictly increasing in θ. Then,

the LHS of Equation A.22 is strictly increasing in θ. For $\theta = 0$, LHS is zero, and the equilibrium value of $\theta_2^{\emptyset}$ is unique and strictly positive with $f > 0$. Using $\theta_2^{\emptyset}$, the equilibrium values of $p_2^{\emptyset}$ and U_2 can be uniquely derived from Equation A.7.

With U_2, the values of $p_2^{pool}(h^0)$ and $\theta_2^{pool}(h^0)$ can be uniquely derived by Proposition A.4:

$$\alpha'(\theta_2^{pool}(h^0)) = \alpha'(\theta_2^k) = \frac{U_2}{b}$$
$$p_2^{pool}(h^0) = p_2^k = b - \frac{\theta_2^k}{\alpha(\theta_2^k)} U_2,$$

if $b \geq U_2$. Otherwise $\theta_2^{pool}(h^0) = 0$ and $p_2^{pool}(h^0) = 0$. Similarly, the values of $p_2^{pool}(h^1)$ and $\theta_2^{pool}(h^1)$ can be uniquely derived in terms of ω:

$$\alpha'(\theta_2^{pool}(h^1)) = \frac{U_2}{a\frac{\pi}{\pi+(1-\pi)\omega} + b}$$
$$p_2^{pool}(h^1) = a\frac{\pi}{\pi + (1-\pi)\omega} + b - \frac{\theta_2^{pool}(h^1)}{\alpha(\theta_2^{pool}(h^1))} U_2.$$

As I showed early $\theta_2^{pool}(h^{\emptyset}) > 0$, then $a\pi + b > U_2$. Still, since $\frac{\pi}{\pi+(1-\pi)\omega} \geq \pi$, then we have that $a\frac{\pi}{\pi+(1-\pi)\omega} + b > U_2$ and $\theta_2^{pool}(h^1) > 0$. By Proposition 2, ω can be uniquely determined. It follows that also $\theta_2^{pool}(h^1)$ and $p_2^{pool}(h^1)$ are uniquely determined. Accordingly, the equilibrium system of equations determines all terms regarding period 2.

The expected profits for entrant hosts in period 1 can be rewritten as follows:

$$(p_1^{\emptyset} - k(1-\pi)\omega)\alpha(\theta_1^{\emptyset}) + \beta\alpha(\theta_1^{\emptyset})\Delta\Pi + \beta\Pi_2(a\pi + b),$$

where $\Delta\Pi$ is defined in Proposition 3 and denotes the value of a transaction in terms of reputation updating. Then, by Proposition 3, with $\theta_1^{\emptyset} > 0$:

$$(p_1^{\emptyset} - k(1-\pi)\omega)\alpha(\theta_1^{\emptyset}) + \beta\alpha(\theta_1^{\emptyset})\Delta\Pi = [a(\pi + (1-\pi)\omega) + b + \beta\Delta\Pi - k\omega(1-\pi)](1 - \epsilon_1^{\emptyset}),$$

where $\epsilon_1^{\emptyset} = \alpha'(\theta_1^{\emptyset})\frac{\theta_1^{\emptyset}}{\alpha(\theta_1^{\emptyset})}$ denotes the elasticity of the matching function with respect to the tightness calculated at $\theta_1^{\emptyset}$.

Thus, the free entry condition in period 1 has the following structure:

$$[a(\pi + (1-\pi)\omega) + b + \beta\Delta\Pi - k\omega(1-\pi)](1 - \epsilon_1^{\emptyset}) + \beta\Pi_2(a\pi + b) = f. \tag{A.23}$$

Then, the optimal value of $\epsilon_1^{\emptyset}$ and $\theta_1^{\emptyset}$ can be uniquely derived by Equation A.23. It is possible to note that: $a(\pi + (1-\pi)\omega) + b + \beta\Delta\Pi - k\omega(1-\pi) \geq 0$ for all values of θ_1; the value of ϵ_1 is strictly

decreasing in θ_1; and $\Pi_2(a\pi + b) \leq f$ by the free-entry condition in period 2. Accordingly, Equation A.23 uniquely characterizes $\theta_1^{\emptyset}$ with $\theta_1^{\emptyset} > 0$. Knowing $\theta_1^{\emptyset}$, I obtain $p_1^{\emptyset}$ and U_1 by Proposition 3, and the measure of entrants in period 1, n_1, by $\theta_1^{\emptyset} = \frac{1}{n_1}$. With n_1, ω, and $\theta_1^{\emptyset}$ the measures of hosts who entered in period 1 and have with histories h^1, h^0, and $h^{\emptyset}$ in period 2 are derived as follows:

$$\begin{aligned} n_2(h^1) &= (\omega n_1(1-\pi) + \pi n_1)\alpha(\theta_1^{\emptyset}) \\ n_2(h^0) &= (1-\omega)n_1(1-\pi)\alpha(\theta_1^{\emptyset}) \\ n_2(h^{\emptyset}) &= n_1(1-\alpha(\theta_1^{\emptyset})). \end{aligned}$$

Then, with $\theta_2^{pool}(h^1), \theta_2^{pool}(h^0)$, and $\theta_2^{(h^1)}$, the measures of guests who direct their search to hosts with histories h^1, h^0, and $h^{\emptyset}$ posting $p_2^{pool}(h^1), p_2^{pool}(h^0)$ and $p_2^{\emptyset}$, are respectively the following:

$$\begin{aligned} g_2(h^1) &= \theta_2^{pool}(h^1)n_2(h^1) \\ g_2(h^0) &= \theta_2^{pool}(h^0)n_2(h^0) \\ g_2(h^{\emptyset}) &= G - g_2(h^1) - g_2(h^0). \end{aligned}$$

Finally, the number of new entrants in period 2 is the difference between the total measure of hosts with history $h^{\emptyset}$ and $n_2(h^{\emptyset})$:

$$\frac{g_2(h^{\emptyset})}{\theta_2^{\emptyset}} - n_2(h^{\emptyset}). \tag{A.24}$$

I started the proof assuming that a positive measure of hosts enter in period 2. Still, for some G, the value in Equation A.24 can be negative. In this sense, the proof of the existence and the uniqueness of the equilibrium relies on a value of $G \geq \bar{G}$, with $\bar{G}$:

$$\bar{G} = g_2(h^1 + g_2(h^0) + n_2(h^{\emptyset})\theta_2^{\emptyset}. \tag{A.25}$$

□

A.3.4 Testable Predictions

Proof of Proposition 5. The measure of guests present in the market in period 2 is assumed to be big enough to allow hosts' entry in period 2 for both equilibria. Accordingly, the free entry condition is binding for f and f'. Then, $\theta_2'^{\emptyset} > \theta_2^{\emptyset}$ recalling that the expected profits for new entrants is strictly increasing in θ. Moreover, directly from the relationship established in the Proposition 1 between the tightness $\theta_2^{\emptyset}$ and the level of U_2, it results that $U_2 > U_2'$. Accordingly, $\theta_2'^{pool}(h^1) > \theta_2^{pool}(h^1)$, and $\theta_2'^{pool}(h^0) \geq \theta_2^{pool}(h^0)$ from the Equations in Proposition 1. In this sense, higher entry costs reduce

the number of hosts who enter the market in period 2; thus, the tightness for all hosts increases and the guests' expected utility from the matches decreases. The derivative of the profits over U_2 has already been discussed in Section A.1 using the definition of the function $\Pi_2(.)$ In particular:

$$\frac{\partial \Pi_2(a\bar{\mu}_2 + b)}{\partial U_2} = \frac{\partial \Pi_2(a\bar{\mu}_2 + b)}{\partial \theta_2}\frac{\partial \theta_2}{\partial U_2} = -\theta_2 \alpha''(\theta_2)\left[\alpha'^{-1}\left(\frac{U_2}{b}\right)\right]' = -\theta_2,$$

if $U_2 \leq b$. The last passage directly follows from the properties of the derivative of the inverse function. If $U_2 > b$, the derivative is equal to zero. Accordingly, a decrease in U_2 has a greater, positive impact on the expected profits for those hosts with a higher value of θ_2. Taking advantage of this result, it is possible to show that $\omega' \geq \omega$ with the same algorithm used in the proof of Proposition 2:

1. Consider the case in which hosts with cost $c = k > 0$ exert effort with probability $\omega = 1$ when entry costs are f. Then, in equilibrium:

$$\beta(\Pi_2\big(a\pi + b - c\big) - \Pi_2(b)) \geq k. \tag{A.26}$$

   From the previous results about the derivative of profits in period 2, the LHS of Equation A.26 is greater with f'. Then, hosts with cost $c = k > 0$ exert effort with probability 1 also with entry costs f': $\omega' = 1$;

2. Consider the case $\omega = 0$ when entry costs are f. Then, in equilibrium:

$$\beta(\Pi_2\big(a + b - c\big) - \Pi_2(b)) \leq k. \tag{A.27}$$

   As before, the LHS of Equation A.27 is greater with f'. Then, Equation A.27 may not be satisfied with f' and, in equilibrium $\omega' \geq 0$;

3. Finally, consider the case in which $\omega \in (0, 1)$ when entry costs are f, such that Equation A.19 is satisfied. With f' the LHS of Equation A.19 increases if $\omega' = \omega$. To restore the equality, the value of ω' has to increase (if possible) so as to decrease the reputation of hosts with history h^1. Thus $\omega' \geq \omega$.

$\square$

B Empirical Setting and Dataset

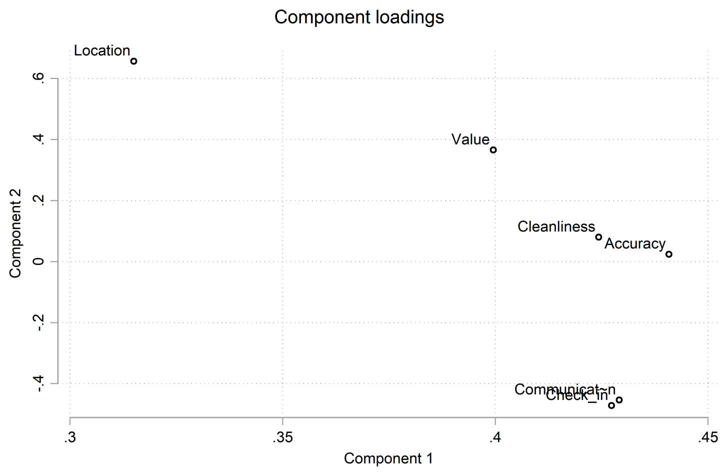
Table 11: Summary Statistics: the Settlement Agreement and Hosts' Selection in September 2017.

	Group C		Group D			
	Mean	SD	Mean	SD	Δ	$p - value$
Days in Airbnb (before September 2017)	544.05	283.28	480.27	253.52	63.78	0.00
Total number of reviews	49.81	63.82	10.86	25.54	38.95	0.00
Number of reviews per snapshots	2.47	2.93	0.59	1.59	1.89	0.00
Price per night	219.26	323.31	284.24	552.97	-64.99	0.00
Average rating per snapshot: overall	94.32	9.11	92.37	10.72	1.96	0.00
Average rating per snapshot: accuracy	9.69	1.11	9.45	1.38	0.23	0.00
Average rating per snapshot: check-in	9.87	0.68	9.66	1.12	0.22	0.00
Average rating per snapshot: cleanliness	9.61	1.01	9.29	1.32	0.32	0.00
Average rating per snapshot: communication	9.83	0.85	9.72	0.88	0.12	0.00
Average rating per snapshot: location	9.48	1.14	9.41	1.43	0.07	0.15
Average rating per snapshot: value-for-money	9.33	1.31	9.10	1.56	0.23	0.00
Minimum number of nights	5.98	10.69	3.07	4.50	2.91	0.00
No short-term rentals	11%	-	1%	-	0.10	-
Registration displayed or not required	22%	-	3%	-	0.19	-
Number of listings	3,958		4,229			

Note: The table compare the profile of listings before and after the Settlement Agreement. All the statistics refer to the snapshot regarding September 2017. Listings are divided in two groups: Group C contains all listings who enter the platform before September 2017 and exit after January 2018, when the implementation of the Settlement Agreement was completed. Group D contains all listings who enter the platform before September 2017 and exit before January 2018. The last two columns provide the differences between the statistics' averages and the $p - value$ of the difference. The numbers of listings in the two groups are not equal to the ones shown in Table 2 since not all listings in the two groups were active (present on the platform) at the date of the snapshot regarding September 2017.

C Identification Strategy

Figure 4: (PCA) Loading of All Rating Categories over the First Two Components



Note: The figure plots the loading of all categories over the first two components of a PCA performed over the rating per snapshot of all the categories. The ratings are all very correlated and the first component already explain more than 30 percent of the ratings variations. Ratings regarding check-in and communication correlate the most and their loadings are distant from all the others. The rating regarding location moves separately, whereas all the other dimensions tend to have similar loadings. These three results are robust to many specifications of principal components and factor analyses.

D Robustness Checks

Table 12: Effort Dynamics over The Total Number of Reviews

	$\bar{r}_{i,t}^{check}$			$e_{i,t}^{check}$		
	>5 reviews	>10 reviews	>20 reviews	>5 reviews	>10 reviews	>20 reviews
$n_{i,t}$	-0.0637***	-0.0355***	-0.00513***	-0.0533***	-0.0315***	-0.00404***
	[0.00831]	[0.00366]	[0.00115]	[0.00803]	[0.00364]	[0.00114]
Constant	9.863***	9.884***	9.798***	0.0782***	0.118***	0.00848
	[0.0164]	[0.0169]	[0.0114]	[0.0158]	[0.0168]	[0.0114]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.0175	0.0145	0.00311	0.0130	0.0122	0.00133
N	4469	7789	14342	4469	7789	14342

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: OLS Estimates of the Impact of Competition on Hosts' Effort

	$e_{i,t}^{comm}$			$e_{i,t}^{check}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(NL_{i,t}^{0.5})$	0.0123			-0.00154		
	[0.0170]			[0.0165]		
$\ln(NL_{i,t}^{1})$		-0.00576			0.0114	
		[0.0230]			[0.0259]	
$\ln(NL_{i,t}^{2})$			-0.00841			-0.00539
			[0.0291]			[0.0313]
Constant	-0.104	-0.0238	0.00102	-0.0560	-0.124	-0.0264
	[0.0714]	[0.124]	[0.193]	[0.0712]	[0.141]	[0.209]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00492	0.00637	0.00663	0.00183	0.00200	0.00198
N	145053	145098	145134	145053	145098	145134

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: IV Estimates of the Impact of Competition on Hosts' Ratings regarding Effort (All Sample)

	$\ln(\bar{r}_{i,t}^{comm})$			$\ln(\bar{r}_{i,t}^{check})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(NL_{i,t}^{0.5})$	-0.0369***			-0.0473***		
	[0.0107]			[0.0116]		
$\ln(NL_{i,t}^{1})$		-0.0425***			-0.0501***	
		[0.0108]			[0.0115]	
$\ln(NL_{i,t}^{2})$			-0.0318***			-0.0449***
			[0.0103]			[0.0111]
Constant	2.417***	2.494***	2.476***	2.463***	2.538***	2.566***
	[0.0441]	[0.0585]	[0.0686]	[0.0480]	[0.0621]	[0.0737]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00205	0.00118	0.000660	0.000104	0.000155	0.000210
N	143796	144853	145035	143796	144853	145035

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: IV Estimates of the Impact of Competition on Hosts' Ratings regarding Effort (Lagged Effort)

	$\ln(\bar{r}^{comm}_{i,t+1})$			$\ln(\bar{r}^{check}_{i,t+1})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(NL^{0.5}_{i,t})$	-0.0325***			-0.0431***		
	[0.0104]			[0.0111]		
$\ln(NL^{1}_{i,t})$		-0.0377***			-0.0450***	
		[0.00994]			[0.0105]	
$\ln(NL^{2}_{i,t})$			-0.0291***			-0.0401***
			[0.00918]			[0.0101]
Constant	2.410***	2.480***	2.470***	2.449***	2.515***	2.539***
	[0.0425]	[0.0536]	[0.0610]	[0.0453]	[0.0566]	[0.0668]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00220	0.000645	0.0000903	0.0000738	0.0000892	0.000130
N	80073	80086	80101	80073	80086	80101

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Impact of the Settlement Agreement on Competition (First Stage): October and December 2017

	October 2017			December 2017		
	$\ln(NL^{0.5}_{i,t})$	$\ln(NL^{1}_{i,t})$	$\ln(NL^{2}_{i,t})$	$\ln(NL^{0.5}_{i,t})$	$\ln(NL^{1}_{i,t})$	$\ln(NL^{2}_{i,t})$
$p^{0.5}_i \times post_{Oct2017}$	0.641***					
	[0.0439]					
$p^{1}_i \times post_{Oct2017}$		0.913***				
		[0.0351]				
$p^{2}_i \times post_{Oct2017}$			1.074***			
			[0.0293]			
$p^{0.5}_i \times post_{Dec2017}$				0.687***		
				[0.0465]		
$p^{1}_i \times post_{Dec2017}$					0.976***	
					[0.0373]	
$p^{2}_i \times post_{Dec2017}$						1.138***
						[0.0308]
Constant	4.148***	5.419***	6.665***	4.148***	5.418***	6.665***
	[0.00769]	[0.00533]	[0.00385]	[0.00769]	[0.00534]	[0.00385]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
F-test	541.9	834.2	1292.1	572.5	908.6	1355.1
R-squared	0.0500	0.0824	0.122	0.0527	0.0866	0.126
N	296164	297671	297980	296164	297671	297980

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: IV Estimates of the Impact of Competition on Hosts' Ratings regarding Effort (October 2017)

	$\ln(\bar{r}_{i,t}^{comm})$			$\ln(\bar{r}_{i,t}^{check})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(NL_{i,t}^{0.5})$	-0.0316***			-0.0366***		
	[0.0113]			[0.0128]		
$\ln(NL_{i,t}^{1})$		-0.0357***			-0.0359***	
		[0.0111]			[0.0121]	
$\ln(NL_{i,t}^{2})$			-0.0229**			-0.0305***
			[0.0104]			[0.0118]
Constant	2.402***	2.464***	2.424***	2.424***	2.467***	2.477***
	[0.0460]	[0.0598]	[0.0687]	[0.0523]	[0.0649]	[0.0778]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00302	0.00137	0.000494	0.000163	0.000299	0.000437
N	86056	86185	86202	86056	86185	86202

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 18: IV Estimates of the Impact of Competition on Hosts' Effort (December 2017)

	$\ln(\bar{r}_{i,t}^{comm})$			$\ln(\bar{r}_{i,t}^{check})$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(NL_{i,t}^{0.5})$	-0.0323***			-0.0469***		
	[0.0106]			[0.0113]		
$\ln(NL_{i,t}^{1})$		-0.0376***			-0.0488***	
		[0.0105]			[0.0110]	
$\ln(NL_{i,t}^{2})$			-0.0284***			-0.0428***
			[0.00968]			[0.0104]
Constant	2.405***	2.474***	2.461***	2.467***	2.537***	2.558***
	[0.0433]	[0.0562]	[0.0641]	[0.0459]	[0.0587]	[0.0687]
Listing FE	YES	YES	YES	YES	YES	YES
Snap FE	YES	YES	YES	YES	YES	YES
R-squared	0.00314	0.00154	0.000697	0.000186	0.000340	0.000521
N	86056	86185	86202	86056	86185	86202

Note: Standard errors clustered by listing are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$