The Cost Of The "Cold-Start" Problem On Airbnb

Florian Dendorfer and Regina Seibel*

August 26, 2024

Abstract

In digital markets with peer-to-peer reviews, new products encounter the so-called "cold-start" problem: Little-known products are bought too rarely and remain little known. While consumers benefit from observing reviews written by others, they do not account for the benefits of trying new products, thereby revealing the products' quality to everyone else, in their own purchase decision. In this paper, we investigate the inefficiency arising from such insufficient social learning on Airbnb. We estimate a dynamic structural model of demand and supply for Airbnbs in Manhattan, New York, from 2016 to 2019. We then implement counterfactual tax-subsidy regimes, which boost demand for new listings relative to incumbent listings and address the cold-start problem. In our main counterfactual, we find that, on average, the rental rate of unreviewed listings should be 30% lower, leading to 25% higher demand. As a result, in the current equilibrium, the number of unreviewed listings is too large by 23%, though in total there are 2% too few listings. According to our estimates, the cold-start problem reduces social welfare by $177,241 per month, a harm that is entirely driven by a reduction in consumer surplus. In a second counterfactual, we explore tax-subsidy regimes which may also be beneficial for reasons beyond the cold-start problem.

Keywords: Cold-start problem, digital platforms, experimentation, market dynamics, product reviews, social learning.

JEL: L11, L15, L83, L86, L88, D83

1 Introduction

A key characteristic of digital platforms is their virtually unlimited shelf space, which provides consumers with an extensive range of products. Amid this abundance, peer-to-peer reviews have become crucial, allowing consumers to learn from the experiences of previous buyers to inform their purchasing decisions. For instance, Reimers and Waldfogel (2021) estimate that

*Dendorfer: Department of Economics, University of Toronto. Seibel: Rotman School of Management, University of Toronto. Email: florian.dendorfer@utoronto.ca, and regina.seibel@rotman.utoronto.ca. We thank Stefan Bühler, El Hadi Caoui, Gregory Crawford, Luise Eisfeld, Brett Hollenbeck, Armin Schmutzler, and conference participants at CRESSE 2023, IIOC 2023, MaCCI 2023, the Swiss IO Day 2023 and the 6th Doctoral Workshop on the Economics of Digitization for helpful comments.

the existence of peer-to-peer reviews are ten times more valuable to consumers than traditional review outlets. However, consumers often hesitate to purchase unreviewed products, which might be of poor quality, preferring reviewed alternatives. At the same time, the marginal review of a little-known product provides more information about its quality than that of a well-established product. Therefore, its social value is relatively larger, but this is not reflected in consumer purchase decisions. This behavior results in the "cold-start" problem (Che and Hörner, 2018; Kremer, Mansour and Perry, 2014), where new products are discovered at an insufficient rate initially. In other words, consumers do not internalize the informational externalities of their purchases, and reviews, on future consumers at the time of purchase.

The cold-start problem also has significant implications for the supply side. Knowing that their demand will be very low initially, new entrant sellers may be discouraged from entering or remaining in the market. For incumbent's however, the cold-start problem may serve to protect their position in the market and they might be less likely to exit. One should note that sellers take social learning into account in their price decisions and therefore may partially counteract the cold-start problem. Entrant sellers may lower their prices to increase the likelihood of being reviewed and to gain a competitive advantage in expectation, which is a motive reminiscent to learning-by-doing (Cabral and Riordan, 1994, 1997). Yet, their products' prices may still be too high from a welfare perspective. Sellers only consider their future profit rather than the overall welfare in their pricing decisions. Conversely, established incumbent sellers may set inefficiently low prices.

In this paper, we examine the impact of the cold-start problem on consumers, sellers, and overall welfare in an online marketplace, specifically analyzing the Airbnb market in Manhattan, New York. We estimate an empirical model of the Airbnb market where listing quality is initially unknown, and peer-to-peer reviews result from past purchases. Utilizing this model, we explore counterfactual per-booking taxes/subsidies on reviewed "incumbent" listings and unreviewed "entrant" listings that would eliminate the welfare loss from the cold-start problem, i.e., from insufficient social learning. First, we consider a tax-subsidy scheme which allows us to focus on the cold-start problem alone by imposing that , on average, taxes and subsidies offset each other such that the average tax level is zero. We thereby avoid that our welfare results is a compound effect of both the tax difference between entrants and incumbent and the overall tax level which affects welfare for reasons other than the cold-start problem. This policy results in a welfare increase of $177,241 per month. In a second counterfactual, we allow the average tax level to be non-zero, which leads to an overall higher welfare increase of $453,191 per month, around 7% of which can be attributed to alleviating the cold-start problem.

To illustrate the cold-start problem, we start by analyzing a simple two-period model of social learning involving two firms: an incumbent and an entrant whose product's quality is a priori unknown. From a social surplus perspective, the entrant's price in the first period should always be lower than the incumbent's price. It turns out that in the competitive equilibrium, dynamic incentives result in a price difference between entrant and incumbent which is higher in absolute values compared to the outcome of static prices. Hence hosts have an incentive to partially counteract the cold-start problem. However, if the prior belief about the entrant's quality is not much worse than the incumbent's quality, this is not enough to overcome the cold-start problem. In equilibrium, a consumer's incentive to purchase the entrant's product is still inefficiently low. This suggests that the existence of the cold-start problem depends on the prior beliefs about product quality.

Building on this, we introduce our empirical model of the Airbnb market. Airbnb guests make individually optimal booking choices among the available listings based on their expectation about a listing's quality. When deciding which listing to book, guests do not take into account the impact of the review that they might write on future guests. Guests hold prior beliefs that they update according to Bayes' rule based on the number of reviews and the rating. On the supply side, there is a pool of hosts, some of whom are inactive. Active hosts set rental rates while considering the potential effect on future demand through the accumulation of reviews. Hosts also make entry and exit decisions; active hosts decide whether to remain active or exit the market and save the operating cost, while inactive hosts can enter at the beginning of each period at an entry cost. Given the large number of Airbnb hosts, we approximate the symmetric Markov equilibrium of this game using the oblivious equilibrium concept introduced by Weintraub, Benkard and Van Roy (2008), where hosts assume that the distribution of competitors remains fixed and equal to the steady state.

We estimate the demand and supply sides of our model separately. To recover the demand parameters, we employ the Generalized Method of Moments, utilizing appropriate instrumental variables. Our estimates underscore the significant impact that ratings have on both guest and host behavior. On average, a good review increases a listing's occupancy rate by 2.6%. For unreviewed listings, this effect is more than twice as large. We estimate a mean quality corresponding to 4.15 stars (on a one-to-five-star scale) in the listing population, whereas the average rating in the data is 4.51 stars. This discrepancy arises because only high-quality listings tend to survive in the market.

We obtain the supply parameters by utilizing a nested maximum likelihood routine to match the observed distribution of listing types and reviews. Our estimates indicate that hosts face substantial entry costs, averaging about half of their lifetime profit. Based on our estimates of hosts' operating costs, we calculate an average profit margin of 24.5%.

To evaluate model fit, we use our model and the parameter estimates to simulate data and compare it to our sample. The simulated data aligns well with the actual data, particularly in terms of generating a similar rating distribution, rental rates and occupancy rates. However, the model predicts larger exit rates, overestimating the dynamism on Airbnb. Nonetheless, we demonstrate that reviews influence host and consumer behavior in a manner broadly consistent with the observed data.

In the counterfactual analysis section of this paper, we induce rate adjustments through subsidies and taxes on unreviewed and reviewed listings to maximize long-run equilibrium welfare. For both reviewed and unreviewed listings, we effectively implement two-part tariffs: Hosts of taxed listings are reimbursed for the tax burden, and hosts of subsidized listings for the subsidy cost through lump-sum payments such that the tax-subsidy scheme is revenue neutral. With our approach, we effectively force hosts to set a price that is different from their individually optimal price and bear the burden associated with that. To make sure that the welfare effects we find are due to a change in the relative attractiveness of entrants for consumers instead of a change in the overall attractiveness of Airbnb, we impose that the average tax/subsidy must be zero. In a second counterfactual, we search over a more general tax/subsidy scheme on reviewed and unreviewed listings, which allows the average tax burden to be non-zero. The optimal tax-subsidy scheme in this counterfactual also incorporates welfare gains from changing the profitability of being a host on Airbnb for *all* hosts and as well as the average price level consumers face, reasons which are unrelated to social learning. To quantify the contribution of alleviating the cold-start problem, we suggest a welfare decomposition.

In the first counterfactual, we find that welfare is maximized with per-booking subsidy of 54.6% and a per-booking tax of 8.1% of the rental rates for unreviewed and reviewed listings, respectively. This leads to a welfare increase amounting to $177,241 per month which is exclusively driven by an increase in consumer surplus. Most of the welfare improvement comes from previously unreviewed listings now having at least one review. Specifically, we observe 49 fewer unreviewed listings but 76 more reviewed ones in the counterfactual equilibrium. When we allow the tax level to be non-zero in the second counterfactual, we find that a per-booking subsidy of 0.7% and a per-booking tax of 22.2% of the rental rates for unreviewed and reviewed listings is optimal, hence a social planner would want to implement a tax-subsidy scheme that leads to a positive average tax of 19.3% of pre-tax prices and increases overall price levels. We find that this counterfactual leads to a larger welfare increase of $453,191 per month. Of this, alleviating the cold-start problem accounts for $31,349 per month.

Our analysis indicates that the cold-start problem leads to a significant welfare loss on Airbnb and potentially other markets with similar characteristics, though the extent of the cold-start problem generally depends on the prior expectation of entrants' quality relative to incumbents'. This paper demonstrates that unreviewed listings are priced too high relative to reviewed ones, resulting in infrequent bookings. Consequently, there is an inefficiently low amount of information about listing quality available to consumers. Although the cold-start problem has been extensively studied theoretically (Che and Hörner, 2018; Kremer et al., 2014; Vellodi, 2022), we contribute to the literature by providing novel, empirical results about the scale and scope of the cold-start problem. To our knowledge, we are the first to estimate an empirical model of social learning and assess the welfare loss from the cold-start problem.

The paper is organized as follows. In Section 2, we discuss the relevant literature. To motivate the paper, we characterize the cold-start problem in a simple model in Section 3. We introduce the empirical model in Section 4 and present the data in Section 5. Section 6 and Section 7 contain the estimation procedure and results for the demand and supply side of the model respectively. We describe the model fit in Section 8. The description and results of the counterfactual analysis are found in Section 9. Section 10 concludes.

2 Literature

There is an extensive literature on social learning (Acemoglu, Makhdoumi, Malekian and Ozdaglar, 2022; Avery, Resnick and Zeckhauser, 1999; Bergemann and Välimäki, 1996, 1997, 2000; Che and Hörner, 2018; Kremer et al., 2014; Papanastasiou, Bimpikis and Savva, 2018). Keller, Rady and Cripps (2005), Avery et al. (1999), and Bolton and Harris (1999) model games of experimentation where players learn from each other, making information a public good.[1] Bergemann and Välimäki (1996, 1997, 2000) embed strategic experimentation into a market setting. In Bergemann and Välimäki (1996), competing sellers set low prices to incentivize a single consumer to experiment. By contrast, Bergemann and Välimäki (1997, 2000) find that, upon introducing a new product, the learning speed is inefficiently high because experimentation is expected to increase the vertical differentiation of products. Acemoglu et al. (2022) introduce a model in which consumers learn about product qualities from online reviews, comparing the effectiveness of different rating systems in facilitating social learning. We relate to this literature by estimating an empirical model that features social learning through peer-to-peer reviews.

A separate stream of theoretical literature analyzes the cold-start problem in social learning from a mechanism design perspective (Che and Hörner, 2018; Kremer et al., 2014; Vellodi, 2022). In Vellodi (2022), consumers only purchase a firm's product if its rating is sufficiently high. It takes new firms a long time to reach this threshold rating, leading to their likely market exit or lack of market entry in the first place. Paradoxically, consumers benefit if product ratings are censored. The market inefficiency in Vellodi's model arises from firms' capacity constraints, and faster learning allows consumers to benefit from less congestion at highly rated firms. Our model also features capacity constraints; unlike in Vellodi (2022), prices are endogenous and products are differentiated.

Che and Hörner (2018) highlight that exploring "ex-ante unappealing" products can be "socially valuable because some of them are ultimately worthy of consumption" (p. 872). A consumer might not try a new product because the private cost of doing so exceeds the social benefit of learning its quality. In their model, a mechanism designer observes a signal about a single product's quality at a rate proportional to the number of consumers who have tried it. The authors show that the designer can increase welfare by occasionally recommending the product to consumers, even if the signal does not confirm high quality. Similarly, Kremer et al. (2014) conclude that a principal, who recommends one of two options with unknown rewards to sequentially arriving agents, learns the reward of an option if at least one agent chooses it. They find that the principal optimally recommends the a priori worse option unless the reward of the other option is sufficiently large. In contrast to these mechanism design solutions to the cold-start problem, our paper investigates the welfare impact of inducing relatively lower prices for new products.

Apart from our paper, to our knowledge, only Pallais (2014) assesses the cold-start problem empirically. Pallais demonstrates that entry-level workers face barriers to labor market

[1] Other models of social learning feature agents who ignore the informational externalities of their actions (Banerjee, 1992; Bikhchandani, Hirshleifer and Welch, 1992; Smith, Sørensen and Tian, 2021). This literature is distinct in that it assumes agents have private information.

entry due to employers' uncertainty about their abilities. Firms do not reap all the benefits of "exploring" new workers instead of hiring seasoned ones. In Pallais's field experiment, the performance of entry-level workers in the treatment group is certified, leading to increased welfare. Similarly, in our paper, we estimate a substantial welfare increase from incentivizing consumers to try new products.

Multiple studies estimate the impact of reviews on sales (Anderson and Magruder, 2012; Chevalier and Mayzlin, 2006), revenues (Luca, 2016), and exit rates (Cabral and Hortacsu, 2010). Additionally, several studies attribute substantial consumer surplus gains to online rating systems (Fang, 2022; Lewis and Zervas, 2016; Reimers and Waldfogel, 2021; Wu, Che, Chan and Lu, 2015). We contribute to this literature by confirming the significance of ratings on Airbnb, demonstrating that unreviewed Airbnb listings are booked less frequently, are cheaper, and are more likely to be discontinued than reviewed ones.

A large body of empirical literature on Airbnb reviews examines the incentives for guests (and hosts) to write reviews and the sources of review bias (Fradkin and Holtz, 2022; Fradkin, Grewal and Holtz, 2021; Proserpio, Xu and Zervas, 2018; Zervas, Proserpio and Byers, 2021). Fradkin and Holtz (2022) highlight that, since consumers bear the cost of writing a review but do not receive all the benefits, online reviews are likely under-provided. In their field experiment on Airbnb, they incentivize reviews but do not find that this increases demand. Our paper relates to theirs by investigating guests' insufficient incentives to book new listings rather than focusing on the broader issue of review provision.

Airbnb has been the subject of extensive structural, empirical research. Most of these studies focus on the housing market (Almagro and Domınguez-Iino, 2022; Calder-Wang, 2021; Farronato and Fradkin, 2022). Huang (2022) examines pricing behavior on Airbnb and find that pricing frictions lead to substantial welfare loss, while Decker (2023) models the rating system as a mechanism designed by Airbnb to persuade consumers to join the platform. Our empirical model of the Airbnb market is the first to endogenize the accumulation of reviews on the platform.

To solve the model, we compute the oblivious equilibrium as developed in Weintraub et al. (2008) and Weintraub, Benkard and Van Roy (2010). Their model focuses on technological investment. In our application, the dynamic decision variable is the price, in the spirit of the learning-by-doing literature (Besanko, Doraszelski and Kryukov, 2014, 2019; Besanko, Doraszelski, Kryukov and Satterthwaite, 2010).

3 The Cold-Start Problem

In this section, we analyze the cold-start problem in a simple model to motivate our counterfactual analysis and provide intuition for our results.

Suppose there are an entrant and an incumbent firm supplying products E and I respectively at zero marginal cost. Let the success rate or "quality" of product $j \in \{E, I\}$ be denoted by $\omega_j \in [0, 1]$. The quality ω_I of I is publicly known and reflects the probability with which consumption of I is a "success" and yields utility of one. If it is a "failure", its utility is zero. The quality ω_E

of E is unknown and distributed according to a beta distribution with parameters $a, b > 0$.

The model has two periods, i.e., $t \in \{1, 2\}$. In each period, a risk-neutral Bayesian consumer arrives to purchase E or I. Crucially, the consumer in $t = 1$ (consumer 1) is distinct from the consumer in $t = 2$ (consumer 2). The prior belief of the consumer 1 about E's quality is characterized in Equation (1).

$$\omega_{E1} \equiv \mathbf{E}_1[\omega_E] = \frac{a}{a+b}. \tag{1}$$

If consumer 1 chooses E, she truthfully reports her experience (success/failure) with probability $v_r \in (0, 1]$ to consumer 2. Note that consumer 1's experience is a Bernoulli distribution with the success probability equal to ω_E. Conditional the consumer 1 writing a review, the posterior belief ω_{E2} of consumer 2 about the quality of E is $\frac{a+1}{a+b+1}$ in case of success and $\frac{a}{a+b+1}$ if consumer 1 experienced a failure.

The expected indirect utilities of product $j \in \{E, I\}$ are given in Equation (2). The unit price of product j is denoted by p_j and ϵ_{jt} represents the idiosyncratic taste shock.

$$u_{jt} = \omega_{jt} - p_{jt} + \epsilon_{jt} = \nu_{jt} + \epsilon_{jt}, \text{ where } \epsilon_{jt} \overset{iid}{\sim} \mathit{Gumbel}(0, \pi^2/6), \tag{2}$$

Under the distributional assumption on the taste shocks, consumer t chooses E with probability $q_{Et} = \exp(\nu_{Et}) / (\exp(\nu_{It}) + \exp(\nu_{Et}))$. The cold-start problem arises because v_{1E} does not account for the fact that the review that may be generated if consumer 1 purchases product E in expectation benefits both consumer 2 and the entrant firm. All else equal, the probability that product E is purchased is inefficiently low as a result.

Firms, on the other hand, exist for both periods and take the effect of their pricing decision in $t = 1$ on their profit in $t = 2$ into account. In $t = 1$, firm j solves the profit maximization problem in Equation (3). Notice that the firm's profit in $t = 2$ depends on p_{E1} and p_{I1} as the prices determine the likelihood that E is bought (and reviewed) in $t = 1$.

$$\max_{p_{j1}} \left(q_{j1} p_{j1} + \delta \mathbb{E}_2 \left[q_{j2} p_{j2} | p_{E1}, p_{I1} \right] \right) \tag{3}$$

Equation (3) characterizes the prices in a subgame-perfect Nash equilibrium. $\pi^i_{j2}, i \in \{0, g, b\}$, denotes firm j's profit in $t = 2$ if its product receives no review (0), a good review (g) or a bad review (b) respectively. Furthermore, $\phi_0 = -1, \phi_g = \omega^e_{E1}$ and $\phi_b = 1 - \omega^e_{E1}$.

Lemma 1. *Suppose that firms compete in Nash-Bertrand fashion. In equilibrium, the entrant firm and the incumbent firm set prices p^*_{E1} and p^*_{I1} respectively, where*

$$p^*_{E1} = \frac{1}{1 - q_{E1}(p^*_{E1}, p^*_{I1})} - v_r \delta \sum_{i \in \{0, g, b\}} \phi_i \pi^i_{E2}$$

$$\text{and } p^*_{I1} = \frac{1}{q_{E1}(p^*_{E1}, p^*_{I1})} + v_r \delta \sum_{i \in \{0, g, b\}} \phi_i \pi^i_{I2}.$$

p^*_{E1} strictly decreases while p^*_{I1} strictly increases in v_r.

The proof of the lemma can be found in Appendix A.1. Notice that both the entrant's and the incumbent's profits are convex in ω_{E1} and, in expectation, information revelation benefits both the incumbent and the entrant. Therefore, the entrant lowers its price in $t = 1$ and raises the likelihood of a review in $t = 2$, countering the cold-start problem and speeding up the revelation of its quality. To the same effect, the incumbent raises its price and further increases the likelihood that E is reviewed. Lemma 1 implies that the price difference between entrant and incumbent is larger in absolute terms if social learning plays a larger role ($v_r\delta \uparrow$).

According to Lemma 1, both firms have an incentive to alleviate the cold start problem, though if consumer choices are socially efficient remains unclear. Note that even absent any considerations for the future, firms will set different prices than is socially optimal in the first period; whoever has a higher expected quality will set a higher price, while, as there is no difference in marginal costs, it is socially optimal that prices are equal across products. We have seen that social learning increases the absolute difference between entrant and incumbent prices when E's quality is expected to be worse than I's, while the opposite is true when the entrant is expected to be better. Proposition 1 characterizes the price difference in the first period which fully alleviates the cold start problem and shows how it compares to the first-period price difference in a Nash equilibrium. Denote the difference in expected quality $\tilde{\omega}_1 = \omega_{E1} - \omega_I$, the equilibrium price difference $\tilde{p}_1^* = p_{E1}^* - p_{I1}^*$ and the socially optimal price difference $\tilde{p}_1^s = p_{E1}^s - p_{I1}^s$.[2]

Proposition 1. *Suppose a social planner sets prices in the first period.*

(i) The socially optimal price difference in $t = 1$ is negative irrespective of $\tilde{\omega}_1$, i.e.,

$$\tilde{p}_1^s < 0 \qquad \forall \qquad \tilde{\omega}_1.$$

(ii) If $\sum_i \delta v_r \phi_i(\pi_{I2}^i - \pi_{E2}^i) < 1$, there exists $\widehat{\omega_1} < 0$ such that $\tilde{p}_1^s < \tilde{p}_1^* \quad \forall \tilde{\omega}_1 > \widehat{\omega_1}$.

The proof of the proposition is relegated to Appendix A.2. According to Proposition 1(i), the social planner chooses a p_{E1} that is smaller than p_{I1} to incentivize consumer 1 to purchase E and review it, irrespective of expected quality of E and the quality of I. Proposition 1(ii) further elaborates on when the socially optimal price difference in the first period is lower than in a Nash equilibrium. To say something meaningful, we need to impose a regularity condition, $\sum_i \delta v_r \phi_i(\pi_{I2}^i - \pi_{E2}^i) < 1$. This condition makes sure that the information gain from an additional review is not excessively valuable to the incumbent compared to the entrant (or, in other words, that the incumbent profits are not more convex than the entrant profits by an order of magnitude). Then, we can show that the entrant's price relative to the incumbent's price is higher in the equilibrium than is socially optimal, as long as consumers believe that the E is not much worse than the I. Hence, the entrant charges a price that is too high and E suffers the cold-start problem. In contrast to the consumers, the entrant and the incumbent are forward-looking. However, they account only for their own future expected profit but not for the future expected consumer surplus in their pricing decision.

[2] The socially optimal price difference in the first period takes into account the Nash equilibrium played in the second period.

Proposition 1 illuminates the underlying causes of the cold-start problem. In addition, it provides the rationale for our counterfactual analysis, in which we examine the welfare implications of a tax-subsidy scheme designed to increase the price disparity between new and established products.

4 Model

In this section, we generalize the model from the previous section and apply it to the context of Airbnb. The generalized model accommodates numerous listings with varying quality and incorporates the strategic decisions of Airbnb hosts. We assume that the true quality of each listing is unknown, but reviews serve as publicly observable signals of quality. We further assume that information is symmetric. That is, Airbnb guests and hosts have the same information and, therefore, have the same beliefs about listing quality at all times. The time horizon is infinite, and hosts maximize the discounted sum of future profits through their pricing, entry, and exit decisions.

Indirect utility and demand. – There exists a set $\mathcal{I}$ of guests, indexed by i, and a set $\mathcal{J}$ of listings, indexed by j. Let N_{jt} denote the number of reviews and K_{jt} the number of good reviews of listing j accumulated up to period t. Guests have unit demand. Let u denote a guest's indirect utility of renting listing $j \in \mathcal{J}_t$ in time period $t \in \{1, .., +\infty\}$.

$$u(p_{jt}, x_{jt}) = \gamma \frac{a + K_{jt}}{a + b + N_{jt}} + \beta_{l(j)} + (1 + f)\alpha p_{jt} + \xi_{jt} + \epsilon_{jt} = \nu(p_{jt}, x_{jt}) + \xi_{jt} + \epsilon_{jt} \tag{4}$$

γ reflects the utility value of having a successful stay with certainty, $\beta_{l(j)}$ is the intercept coefficient, which differs by listing type $l(j)$, and α is the rental rate coefficient. p_{jt} is the rental rate, f is the fee Airbnb levies on consumers, the structural error ξ_{jt} captures the unobserved (to the econometrician) listing characteristics, and ϵ_{jt} is an idiosyncratic taste shock. As in Section 3, a and b govern the prior distribution over listings' success rates, i.e., $\omega_j \overset{\text{iid}}{\sim} Beta(a, b)$. We allow for four different listing types, $l \in \{1, 2, 3, 4\}$, to account for observed differences in listing characteristics, such as a listing's location and amenities. A listing with more desirable characteristics is assigned a higher type. The details about how we construct listing types are found in Section 5. Notice that a listing's type and its rating are independent of each other. For example, there may be a type 1 listing with a bad rating and a type 4 listing with a good rating. The state x_{jt} of listing j in t is $(K_{jt}, N_{jt}, l(j)) \in X$ where $X = \{(K, N, l) \in \mathbf{N}_0^2 \times \{1, 2, 3, 4\} : K \leq N, N \leq \bar{N}\}$. $\bar{N} \in \mathbf{N}_+$ is the maximal number of reviews our model allows for.

As in Section 3, the taste shocks are independently and identically drawn from a normalized Gumbel distribution. We also normalize the mean utility of taking the outside option, which can be interpreted as booking a hotel rather than an Airbnb, to zero. We abstract from guests arriving sequentially in the run-up to t to make bookings, as this would immensely complicate characterizing the demand system.[3] Rather, in each t, a discrete number of guests arrive simul-

[3] If consumers arrive sequentially, expected demand does not have a simple, closed-form solution because the current set of available listings depends on past booking decisions. While it is possible to integrate the different booking sequences numerically, it is computationally expensive.

taneously in the market to book accommodation for the duration of the period. Guest arrival follows a Poisson process with mean μ. In each t, an average number of μ guests arrive in the market to book one of the listings or take the outside option. Listings are capacity-constrained; at most, one consumer can rent listing j in t. We show in Appendix A.3 that the number of consumers who want to book listing j is Poisson distributed. Equation (5) characterizes the demand q for j in t. It equals the probability that at least one of the guests arriving in t wants to book j. If more than one guest wants to book j in t, we assume that one guest successfully books the listing while all remaining ones are forced to take the outside option.

$$q(p_{jt}, x_{jt}, P_t, s_t) = 1 - \exp\left(-\mu \frac{\exp(\nu(p_{jt}, x_{jt}))}{1 + \exp(\nu(p_{jt}, x_{jt})) + \sum_x^X (s_t(x) - \mathbb{1}_{x=x_{jt}}) \exp(\nu(P_t(x), x))} \right) \tag{5}$$

$s_t(x)$ denotes the number of listings in each state x in t. Equation (5) assumes that hosts other than j set the rental rate of their listings according to the pricing policy function $P_t(x)$. Listings that share the same state have the same rate. This will be true in the symmetric equilibrium. Notice also that $P_t(x)$ does not depend on p_{jt} or the prices of $j' \neq j$. As discussed below, this follows from the oblivious equilibrium concept we use. Hence, we can simplify the notation and drop the j subscript in what follows.

State transitions. – If a listing is booked for t, with probability $\upsilon_r \in (0, 1)$, the guest accurately reports its experience (success, failure) in a review.[4] The probability ρ_0 that a listing's state does not change, either because it is not booked or the guest fails to leave a review, is $1 - \upsilon_r q$. The listing receives a good review with probability ρ_g or a bad review with probability ρ_b, depending on the listing's quality prior.

If the review is good, both N and K increase by one in $t+1$. If the review is bad, N increases by one, but K does not. The possible transitions are illustrated in figure Figure 1.

$$\begin{tikzcd}
 & (K_{t+1}, N_{j(t+1)}, l) = (K_t, N_t + 1, l) \\
(K_t, N_t, l) \arrow[ur, dash, "\textit{bad review}"] \arrow[r, dash, "\textit{no review}"] \arrow[dr, dash, "\textit{good review}"'] & (K_{t+1}, N_{t+1}, l) = (K_t, N_t, l) \\
 & (K_{t+1}, N_{t+1}, l) = (K_t + 1, N_t + 1, l)
\end{tikzcd}$$

Figure 1: State transitions.

Equation (6) summarizes the transition probabilities. Based on Equation (6) we formulate the transition matrix $\mathbf{T}(\mathbf{P}_t, \mathbf{s}_t)$, where $\mathbf{P}_t = P_t(\mathbf{X})$ and $\mathbf{s}_t = s_t(\mathbf{X})$ in Appendix A.4.

[4] Note that we assume that reviews are truthful and the probability of leaving a review is exogenous. The incentive of consumers to leave a product review may be manifold and is the subject of analysis in many other scholarly contributions, such as Fradkin et al. (2021). Modeling this is beyond the scope of this paper.

$$\begin{aligned} &\rho_0(p_t, x_t, P_t, s_t) = 1 - v_r q(p_t, x_t, P_t, s_t) \\ \text{and } &\rho_g(p_t, x_t, P_t, s_t,) = v_r q(p_t, x_t, P_t, s_t) \frac{a + K_t}{a + b + N_t} \\ \text{and } &\rho_b(p_t, x_t, P_t, s_t) = v_r q(p_t, x_t, P_t, s_t) \left(1 - \frac{a + K_t}{a + b + N_t}\right) \end{aligned} \tag{6}$$

The host's problem. – There is a finite number of J potential hosts, but not all of them are active. Each active host operates a single listing.[5] An active host makes per-period revenue $\pi(p_t, x_t, P_t, s_t) = (\text{period length in days}) q(p_t, x_t, P_t, s_t) p_t$. We do not model a marginal cost of renting out a listing. Instead, a host incurs operating cost ϕ_{lt}, which is a fixed cost. ϕ_{lt} reflects that a host cannot use their apartment themselves for the time it is listed on Airbnb. This holds irrespective of whether the listing is booked or not. There is also a fixed cost κ_{lt} from entering the market and becoming active as the host must convert their apartment to an Airbnb rental. The operating and entry costs are random variables that vary over time. ϕ_{lt} and κ_{lt} follow exponential distributions with mean $\bar{\phi}_l$ and $\bar{\kappa}_l$ respectively. That is, $\kappa_{lt} \overset{\text{iid}}{\sim} Exponential(\bar{\kappa}_l)$ and $\phi_{lt} \overset{\text{iid}}{\sim} Exponential(\bar{\phi}_l)$. As the costs may differ for listings with different characteristics, we allow their means to vary across listing types.

Timing. – In each t the timing of events is a follows.

1. Active hosts set their prices and receive the per-period revenue.
2. Inactive hosts observe κ_{jt} and decide whether to enter at the beginning of $t + 1$ or not. They pay κ_{jt} if they enter.
3. Active hosts observe ϕ_{jt} (the cost of operating the listing in $t + 1$) and decide whether to exit at the end of t or not. They pay ϕ_{jt} if they do not exit.
4. Review outcomes are determined, and the industry takes on a new state $\mathbf{s}_{t+1}$.

Value function. – Having entered, in each t, the active host maximizes the expected discounted profit flow through her exit and pricing decisions.[6] We denote the discounted profit flow from the optimal pricing and exit behavior conditional on staying in the market in the current period by $V(x_t, P_t, s_t)$, the value function. Notice that V depends on the pricing policy function $P_t(x)$ and the number of listings $s_t(x)$ in each state. In equilibrium, all hosts behave according to a common, stationary strategy. This allows us to drop P_t and s_t as arguments of V. We write V recursively in Equation (7). Note that the host shares a common prior with consumers, which precludes any form of price signaling. In practice, this means that the host is as good as any consumer when judging the quality of its listing.

$$V(x_t) = \max_{p_t} \left\{ \pi(p_t, x_t) + \mathbb{E}_\phi \left[\max\left(0, \delta \mathbb{E}_{x_{t+1}} \left[V(x_{t+1}) | p_t, x_t \right] - \phi_{lt} \right) \right] \right\} \tag{7}$$

[5] 90% of hosts in our dataset operate a single listing. The average number of listings per host is 1.10.

[6] Notice that we do not allow hosts to invest in quality, as listing quality is assumed to be exogenous. In principle, hosts could improve their services to guests to get better reviews. We find that ratings are highly correlated across categories, such as communication, accuracy, cleanliness, check-in, and location. Location-related ratings are likely exogenous.

$\delta \in (0,1)$ denotes the discount factor. Equation (7) incorporates the host's decision to exit the market when the current operating cost exceeds the expected value of remaining in the market. A host who exits the market cannot re-enter and is replaced with an inactive host.

Exit & entry rates. – Recall that a host exits the market if ϕ_{lt} exceeds $\mathbb{E}_{x_{t+1}}[V(x_{t+1})|p_t, x_t]$ and that ϕ_{lt} is exponentially distributed. Then, the probability that an active host of a listing in state x_t exits is given by $\chi(p_t, x_t)$.

$$\chi(p_t, x_t) = \exp(-\delta \mathbb{E}_{x_{t+1}}[V(x_{t+1})|p_t, x_t]\bar{\phi}_l^{-1}) \tag{8}$$

If a host has not entered the market, it is inactive. A listing is sure to have no reviews upon entry. An inactive host, therefore, chooses to enter if $\kappa_{lt} \leq V((0,0,l))$ and remains inactive otherwise. The entry probability of an inactive host of a type l listing is λ_l.

$$\lambda_l = 1 - \exp(-\delta V((0,0,l))\bar{\kappa}_l^{-1}) \tag{9}$$

Using the fact that the operating cost is exponentially distributed, we specify hosts' exit behavior. We show in Appendix A.7 that the expected operating cost is $(1-\chi(x))\bar{\phi}_l - \mathbb{E}_{x_{t+1}}[V(x_{t+1})|p_t, x_t]\chi(x)$. We rewrite Equation (7) as follows.

$$V(x_t) = \max_{p_t} \left\{ \pi(p_t, x_t) + \delta \mathbb{E}_{x_{t+1}}[V(x_{t+1})|p_t, x_t] - (1-\chi(p_t, x_t))\bar{\phi}_l \right\}, \tag{10}$$

The *expanded* transition matrix $\mathbf{F}(\mathbf{P}_t, \mathbf{s}_t, \chi, \lambda)$ extends $\mathbf{T}$ to cover transitions from and to activity. We define $\mathbf{F}$ in Appendix A.4. $\mathbf{F}$ allows us to compute the stationary listing distribution $\mathbf{s}$ across states implied by $\mathbf{F}^T\mathbf{s} = \mathbf{s}$.

Equilibrium. – Computing a Markov perfect equilibrium, as Ericson and Pakes (1995) do in their seminal paper on industry dynamics, is computationally infeasible in our context. In the Markov perfect equilibrium, the dimension of the state space increases exponentially in the number of hosts, and so does the cost of computing the Markov perfect equilibrium. It is prohibitively large for thousands of hosts.

Instead, we compute the oblivious equilibrium introduced in Weintraub et al. (2008).[7] The oblivious equilibrium can be interpreted as an approximation of the Markov perfect equilibrium of a game with a large number of players whose strategic response to a change in a single player's behavior is negligible. In the context of Airbnb, a host is unlikely to adjust its rental rate in response to a rate change by one of its thousands of competitors. Each host considers only its own state and the stationary equilibrium listing distribution $\mathbf{s}$ when choosing its pricing strategy $P(x)$. In equilibrium, $\mathbf{s}^*$ is consistent with the hosts' pricing, exit, and entry behavior.

Let $\mathbf{s}^*$ be the stationary listing distribution that arises from each host behaving according to $\mathbf{P}^*$. $\mathbf{P}^*$ is the oblivious equilibrium strategy if and only if no host has a strict incentive to deviate from P^* to any P' in strategy set Π if all other hosts behave according to $\mathbf{P}^*$ and $\mathbf{s}^*$ remains unaffected by the host's deviation. Equation (11) formalizes this notion. Intuitively, each host

[7] There exist other equilibrium concepts for large dynamic games (Doraszelski and Judd, 2012, 2019; Fershtman and Pakes, 2012). For these, unlike for the oblivious equilibrium, the size of the state space depends on the number of players.

behaves like an atom whose deviation $\mathbf{s}^*$ does not respond to. It is important to realize that hosts have market power regardless, as listings are differentiated.

$$V(x|P^*, P^*, \mathbf{s}^*) \geq V(x|P', P', \mathbf{s}^*) \; \forall \; P' \in \Pi. \tag{11}$$

5 Data

Airbnb is the market-leading peer-to-peer platform for short-term accommodation. It enables hosts to rent their apartment or a room to guests, usually tourists. Since its founding in 2008, Airbnb has grown to feature more than four million Airbnb hosts worldwide, housing about 33 million guests annually on average.[8] Airbnb's annual revenue of roughly 6 billion US dollars in 2021 rivals the revenue of large hotel chains. With thousands of active Airbnb listings, New York City is by far Airbnb's largest market among US cities. Due to many *a priori* unknown listings, the crowd-based rating system of Airbnb is one of its key services to potential guests. Guests are encouraged to leave a review of the listing on a one-to-five stars scale within two weeks after concluding their stay, which is then published for other potential guests to see. The average star rating and the cumulative number of reviews are prominently displayed on Airbnb's website.

We use data collected by AirDNA, a data analytics company, on all Airbnb listings in Manhattan, New York, between January 2016 and December 2019.[9] Figure 2 (left) illustrates the spatial distribution of listings across Manhattan. During this time, the number of listings is relatively stable, supporting the idea that the market is in a stationary equilibrium (see Figure 2, right). We chose the timeframe to exclude the period affected by the COVID-19 pandemic. We have rental rates in USD for each listing and day during the observation period and information on whether a listing was available, reserved, or blocked for private use by the host on a given day. In addition, we observe the number of reviews and the rating on a one-to-five stars scale for each listing in a roughly bi-monthly frequency.[10] Lastly, the dataset includes various listing attributes and amenities, which we use to define the relevant market. We focus on listings that offer the entire apartment for rent (as opposed to a single room), include at least one picture of the place, permit at most two guests and no pets, and feature one bedroom and one bathroom. We drop observations if they occur before the first booking of a listing because we suspect that the earliest observation often predates the actual market entry. This way, we exclude listings that are never booked during their lifespan. We also drop listing-days with observed prices below the first and above the 99th percentile, corresponding to \$65 and \$518, respectively. We account for the cleaning fee hosts typically charge per stay by adding it to the daily rate, divided by the average reservation length of 5.5 days. If we do not observe the cleaning fee for a listing, we assume it is equal to the average cleaning fee in the data sample.[11]

[8] See https://news.airbnb.com/about-us/.

[9] See https://www.airdna.co/.

[10] The frequency of the ratings and reviews depends on the frequency of the AirDNA scraping. All other variables are available at the daily level. There are less than two weeks between 40% of review observations. Only 1% observations lie more than three months apart.

[11] This leaves us with 7,586 unique listings and 2,298,598 listing-days, excluding blocked ones.

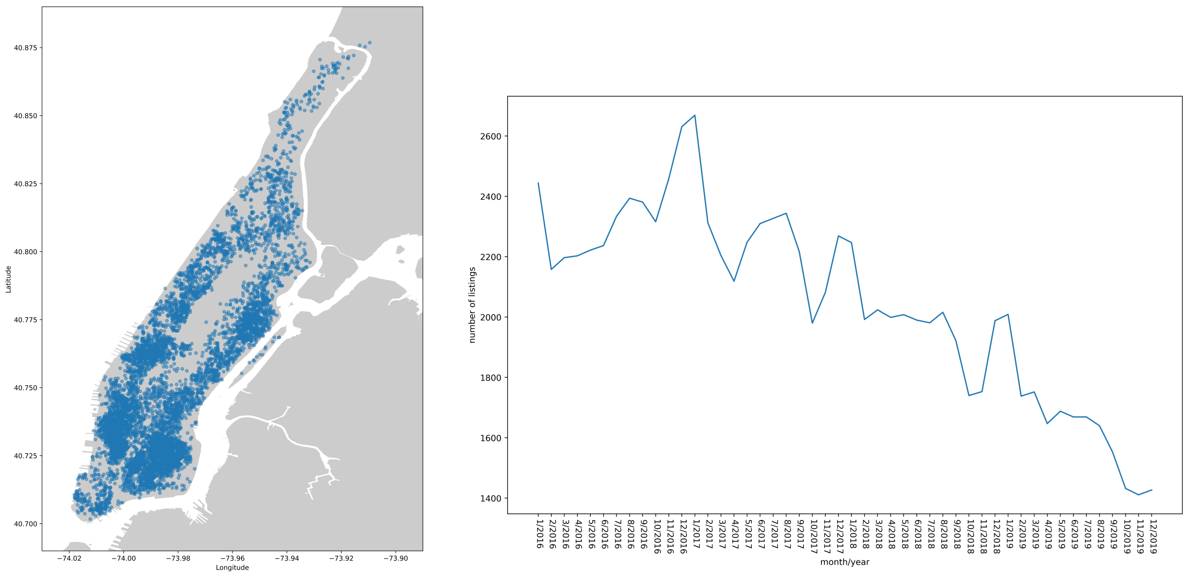

Figure 2: Number of listings over time and their location.

To estimate the model, we aggregate the data to four-week periods, or what we refer to as "months" in the following. This is because inversion of the demand system in the context of demand estimation precludes occupancy rates of 0% and 100%. After the previously described cleaning, we have 62,937 listing-month observations.

	mean	std	min	25%	50%	75%	max
Rental rate	\$193.02	\$60.13	\$70.33	\$150.75	\$184.78	\$270.95	\$562.43
Occupancy rate	60.64%	33.57%	0.00%	33.33%	69.23%	100.00%	100.00%
Number of reviews	10.30	8.19	0.00	2.00	9.00	20.00	20.00
Rating	4.51	0.72	1.00	4.40	4.67	5.00	5.00
Monthly exit rate	3.21%	0.80%	1.57%	2.68%	3.21%	4.08%	5.55%
Monthly entry rate	4.40%	1.85%	0.39%	3.17%	4.36%	6.26%	9.83%
Lifespan (in months)	17.67	16.38	1.00	3.00	12.00	39.00	52.00

Table 1: Data summary.

We construct the number of "good" reviews K as the number of five-star reviews required to achieve the observed rating if the remaining $N - K$ reviews were "bad" one-star ones. For example, for a listing with ten reviews and a rating of 3.8 stars, the implied number of good reviews is 7. We deal with missing information on the number of (good) reviews by filling in the most recent observed value. To keep a reasonably sized state space, we censor the number of total reviews at 20, i.e., $\bar{N} = 20$. Since the marginal effect of N on the posterior belief becomes small as N grows large, the change in the posterior mean tends to be minor for N exceeding 20. For example, suppose the prior quality distribution has a mean of 4.15 stars and a variance of 0.27.[12] Figure 3 illustrates how the posterior mean experiences ever smaller changes as the listing receives either a sequence of good or bad reviews. Adding a good review to a yet unreviewed listing improves its expected quality by 4.42%, whereas the expected quality of a listing with 20 good reviews would only move by 0.19%.

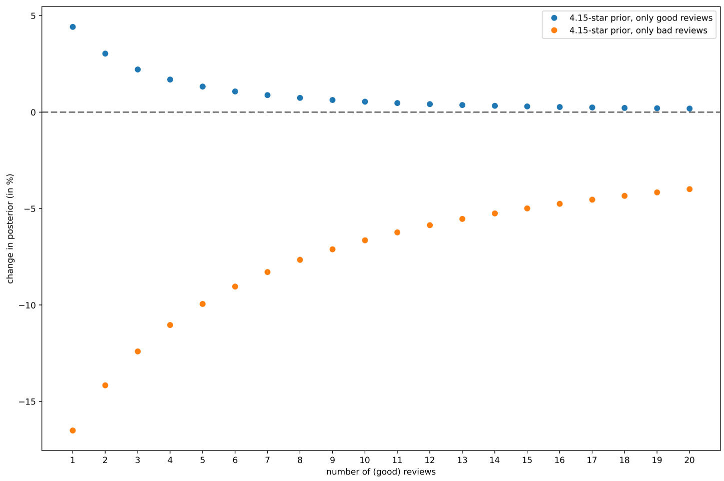

Figure 3: Example posterior changes.

Table 1 summarizes key variables in the data on the listing-month level. The average rental rate is $193.02 in our sample. Rental rates exhibit little variation. The majority of listings are priced in the $50 range around the average rate. On average, Airbnb listings are occupied 60.64% of the time and have been reviewed about nine times (after censoring the number of reviews). On average 3.21% of hosts leave the market each month and 4.40% have just entered.[13] The average lifespan within our sample is 1.3 years, but most listings exist for less than a year.[14]

[12] These values correspond to our estimates (see Section 6).

[13] We use the whole sample with dates before January 2016 and after December 2019 to determine entry and exit. Rates are relative to the active hosts in the market.

[14] Here, the lifespan are the number of months for which we observe a listing in our sample, hence it has a natural maximum of 4 years.

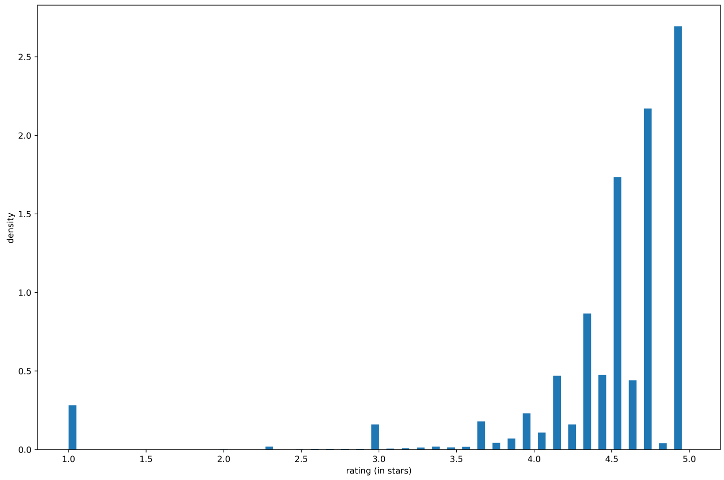

Figure 4: Rating distribution in the data.

At 4.5 stars, the average rating is high. In comparison, at 3.8 stars, the average rating on TripAdvisor is substantially lower (Zervas et al., 2021). It is a priori unclear if this is mainly because highly rated listings remain in the market longer or because the listings' quality is generally high. Figure 4 shows the rating distribution in our sample. The distribution is left-skewed because most listings are rated four stars or higher. Listings with a lower rating tend to have fewer reviews. Almost all listings with a 1-star, three-star, or 3.66-star rating have less than four reviews.

Conditional on the rental rate and the reviews, we want listing types to capture as much heterogeneity in the utility that consumers derive from different listings as possible, both unobserved and observed. We choose to be parsimonious and distinguish only four types, thereby limiting the size of the state space and keeping the model tractable. To construct the four listing types, we regress the occupancy rate on the rental rate, K-N fixed effects, month-year fixed effects, and listing fixed effects. We divide the estimated listing fixed effect coefficients into type 1, type 2, type 3, and type 4, depending on the quartile they fall into. Figure 5 shows how the type of a listing relates to the fixed effect we estimate for that listing and the listing's rental rate.

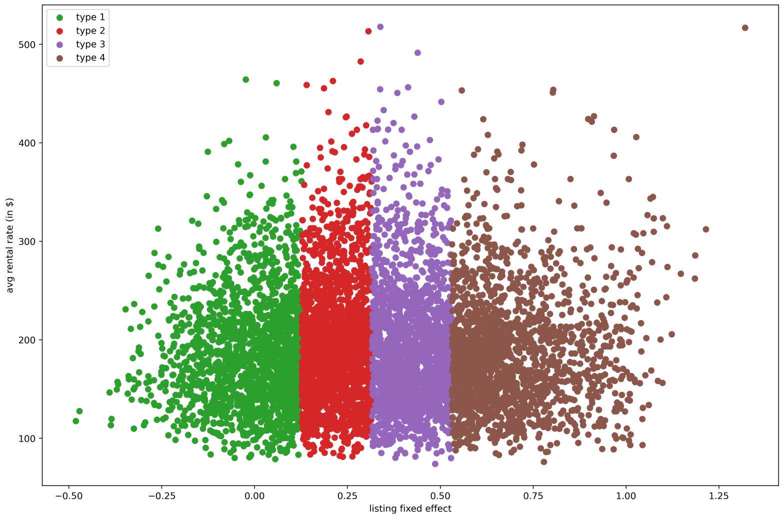

Figure 5: Listing type definition.

We report summary statistics for the different types in Table 2. On average, higher types feature a higher occupancy rate. While a type 4 listing is, on average, booked 81% of the time, the average occupancy rate of a type 1 listing is only 35%. The correlation between the type and the number of reviews is very weak (Pearson correlation coefficient: -0.013). The rating tends to increase in the type (0.09). As a results, also the rental rate is positively correlated with the listing type (0.08). Higher-type listings also tend to be older. The correlation coefficient of a listing type and the lifespan is 0.11.

	Price	**Occupancy**	**Reviews**	**Rating**	**Lifespan**
type 1	$186.30	34.60%	9.89	4.39 stars	14.6 months
type 2	$192.83	53.47%	10.48	4.48 stars	20.2 months
type 3	$189.45	68.58%	11.04	4.56 stars	21.2 months
type 4	$204.40	80.92%	9.32	4.58 stars	14.7 months

Table 2: Summary: listing types.

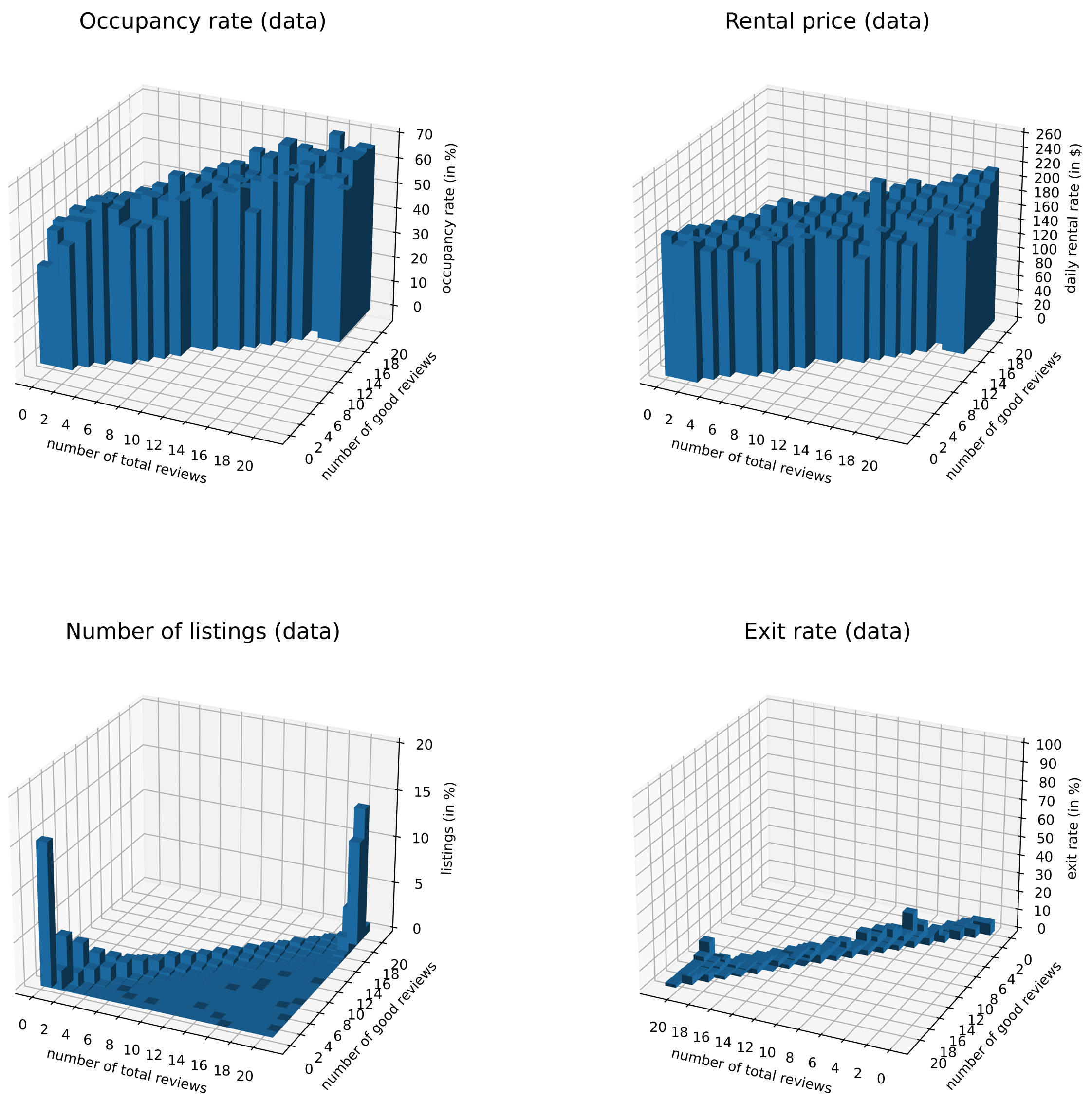

Figure 6: Occupancy rate (top left), rental rate (top right), number of listings (bottom left), and exit rate (bottom right) by number of (good) reviews in the data.

We plot the average occupancy rates, rental rates, number of listings, and exit rates depending on N and K in Figure 6. Occupancy rates tend to be larger for listings with many reviews, assuming these reviews are good (Figure 6, top left). Listings without reviews, for example, are around 40% less likely to be booked than listings with 20 good reviews. Similarly for rental rates, though the relationship is less stark. Listings with 20 good rather than zero review have a 12% higher rate on average.[15]

Hosts of listings with many reviews and a good rating make more revenue and remain in the market, whereas they leave the market otherwise. Hence, our dataset contains very few or no

[15] Moreover, prices of entrants are relatively high, higher than the price of a listing with 20 reviews and a 4.6-star rating.

observations for many states associated with small K, while most observations are concentrated on high-K states (see Figure 6, bottom left). The right-skewed listing distribution of firms over (good) reviews is driven by a selection effect reminiscent of Jovanovic (1982). High-quality listings survive, whereas low-quality ones fail and exit. The listing distribution features a "pitch fork" shape, as most listings either have few or no reviews or – due to right-tail compression – have the maximum amount. Notably, there are roughly 5.5 times as many reviewed listings as unreviewed ones. About 39% of reviewed listings have the maximal review count. We also observe the selection effect in hosts' exit behavior (see Figure 6, bottom right). Compared to a reviewed listing, an unreviewed listing is around three times as likely to exit in a given month.

To support the graphical intuitions of Figure 6, we regress various dependent variables on the rental rate, the number of reviews, the number of good reviews, their interaction, listing types as well as year-month. In this way, our regression results account for differences in time-invariant listing characteristics and seasonality. They are reported in Table 3. Considering an average listing with ten reviews and a 4.5 star rating as baseline, a listing with one more review has a price that is higher by \$0.98 if the review is good and a price that is lower by \$10.55 if the review is bad. Comparing the same listings, the occupancy rate increases by about 1.33 percentage points and decreases by 1.08 percentage points after receiving a good and a bad review, respectively. Notice that the occupancy rate barely responds to an increase in the rental rate by \$1 – it is a mere 0.1 percentage points lower in response. The exit rate is generally lower for a listing with an additional review, whether the review is good or bad seems to play a minor role. This could be the result of the strong selection effect in the dataset.

We do not estimate all the parameters of our model, but we calibrate some of them using data moments. Recall that a period is a four-week interval. We estimate the four-week review rate v_r, i.e., the number of additional reviews a listing has four weeks after being booked, as the change in the number of reviews over the number of bookings of listings with less than 20 reviews in our data sample. In this way, we determine that v_r is 0.992. Given that the average reservation length is 5.5 days, the probability that a guest leaves a review after the stay is around 20%. It is important to note that v_r does not depend on the number of (good) reviews, the rental rate, or whether the guest experiences a success or a failure. Each booking is equally likely to result in a review. At any given time, we observe, on average, 1,210 active listings, of which 735 are booked. We set the mean number of guests arriving in the market to 10,000. Assuming that one in three consumers looking for accommodation considers Airbnbs, this implies that Airbnb's mean market share is 2.1%, consistent with the 2% median market share across 50 US cities in Farronato and Fradkin (2022). We set the number of potential listings J to 10,000. For the discount factor δ, we choose 0.995, implying an annualized interest rate of 6.7%. Finally, Airbnb charges – both in the model and in reality – a commission fee of 14.2%.[16] The calibrated parameters are summarized in Table 4.

[16] See `https://www.airbnb.ca/resources/hosting-homes/a/how-much-does-airbnb-charge-hosts-288`.

	Dependent variable:		
	rental rate	occ rate	exit rate
	(1)	(2)	(3)
rental rate		-0.001^{***}	-0.0001
		(0.000)	(0.000)
no of reviews	-10.340^{***}	-0.002^{***}	-0.003***
	(0.105)	(0.004)	(0.001)
no of good reviews	11.795^{***}	0.035^{***}	-0.001
	(0.492)	(0.002)	(0.001)
no of reviews $\times$	-0.024	-0.001^{***}	0.0001***
no of good reviews	(0.018)	(0.000)	(0.000)
Type FE	Yes	Yes	Yes
Observations	62,937	62,937	62,937

Note: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parenthesis clustered at type level.

Table 3: Rental rate, occupancy rate, exit rate and reviews in the data.

Parameter		Value
Discount factor	δ	0.995
Revenue fee	f	0.142
Arrival rate	μ	10,000
Review rate	v_r	0.992
Maximum number of reviews	$\bar{N}$	20
Maximum number of listings	J	10,000

Table 4: Calibrated parameters.

6 Demand Estimation

We estimate the parameters governing demand by the generalized method of moments. Some listings still have zero bookings in a given month (or, more rarely, are fully booked out), in which case we have to drop those observations for the demand estimation. This leaves us with 49,214 observations and 7171 unique listings. We invert demand and back out ξ_{jt} to compute the moment conditions. Equation (12) characterizes the regression equation we estimate.

$$\ln\left(\frac{-\ln(1-B_{jt})}{\mu_t}\right) - \ln\left(\frac{1+\sum_j^{\mathcal{J}_t}\ln(1-B_{jt})}{\mu_t}\right) = \nu(p_{jt}, x_{jt}) + \xi_{jt}. \tag{12}$$

We account for seasonality by allowing the arrival rate to vary by month. We compute μ_t as μ times the average percent deviation of the total number of Airbnb bookings from the mean in a given month. For example, in the first four weeks of the year (i.e., in January), the arrival rate is 22.8% lower than average. Following Ferrari and Cribari-Neto (2004) and Dickstein (2018), we do not estimate a and b directly, but estimate ψ and ι instead. They are defined in Equation (13). ψ determines the prior mean, whereas ι is closely related to the variance of the prior distribution.

$$\frac{1}{1+\exp(-\psi)} = \frac{a}{a+b} \quad \text{and} \quad \exp(\iota) = a+b. \tag{13}$$

This alternative formulation facilitates the estimation but also restricts the set of possible solutions. Naturally, the type coefficients and the coefficient of the rental rate are identified by variation in the listing types and the rental rate, respectively. The identification of γ, ψ, and ι comes from within- and between-listing variation in the number of (good) reviews. ι pins down to what extent an additional review moves the posterior mean away from the prior mean. If ι is small, the prior belief is precise, and guests make only marginal adjustments to their beliefs after observing the rating. ψ depends on variation in the rating to be identified. The posterior mean responds more strongly to reviews if the rating differs greatly from the prior mean. In particular, the posterior mean increases by more as the rating improves if the prior mean is low. In this way, the relationship between the rating and the posterior mean allows us to estimate the prior mean. γ captures the impact of the posterior mean on guests' booking decisions. If the reviews have little effect on the occupancy rate of the listings, regardless of their current rating and review count, γ must be low.

Column (1) in Table 5 shows the estimation results assuming independence of the structural error regarding the rental rate and the number of (good) reviews. Clearly, this assumption may be violated in reality. First, the four listing types may not fully capture the time-invariant listing characteristics. In this case, the structural error is correlated with the occupancy rate and the number of reviews, which result from past bookings. Second, rental rates may be correlated with time-varying unobserved demand shocks. Third, hosts may exert effort to improve their listing's desirability and their incentive to do so may depend on market conditions. Using suitable instrumental variables, we account for the potential endogeneity of the rental rate and the number of (good) reviews.

We instrument p with the average reservation length of a listing. The reservation length serves as a proxy for any cost of welcoming new guests (we assume in the model that there is none). Properties occupied for a longer period by the same guests are less costly to maintain, which allows the host to charge a lower rate. If the average reservation length is unrelated to a listing's characteristics, it satisfies the exclusion restriction. We instrument the number of (good) reviews with the de-meaned five-, six–, seven–, and eight-month lagged occupancy rates, their squares, and their interaction with the rating. Intuitively, an unexpectedly higher occupancy rate in the past has no bearing on the present occupancy rate other than through the reviews that resulted from it. The results of estimating the demand coefficients using instruments are reported in column (2) in Table 5.

As expected, the standard errors of coefficients that are identified with the instruments are relatively larger. The coefficient of the posterior mean narrowly fails to be significant on the 10% level. ψ and ι imply a prior mean and variance of 4.15 stars and 0.27, respectively. Figure 7 relates the estimated prior distribution to the rating distribution in the data. Recall that the rating is an imprecise signal of the true quality and features a larger variance. Also, the rating is subject to the selection effect; highly rated listings are observed in the data, whereas poorly rated listings are not.

		(1)	(2)
prior	ψ	-0.1810	1.3171**
		(0.1463)	(0.7421)
	ι	1.9633***	1.6212
		(0.1233)	(1.1995)
rental rate	α	-0.0020***	-0.0086***
		(0.0001)	(0.0015)
types	β_1	-10.6106***	-10.5354***
		(0.1109)	(1.7158)
	β_2	-10.1166***	-9.8218***
		(0.1110)	(1.7193)
	β_3	-9.6580***	-9.4401***
		(0.1116)	(1.7221)
	β_4	-9.2443***	-8.9099***
		(0.1114)	(1.7415)
expected quality	γ	1.6000**	2.8607
		(0.1358)	(1.8711)
Observations		49,214	25,824

Note: *p<0.1; **p<0.05; ***p<0.01. Robust standard errors in parenthesis.

Table 5: Demand estimates.

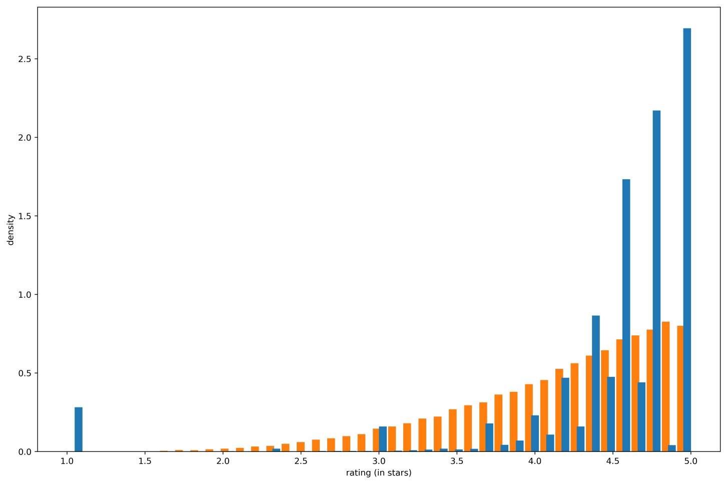

Figure 7: Estimated prior distribution and rating distribution in the data.

The magnitude of the rental rate coefficient is four times larger in column (2) compared to column (1), suggesting that the rental rate is indeed endogenous in column (1). The estimated rental rate coefficient corresponds to an average listing-month price elasticity of -1.05. To compare, Farronato and Fradkin (2022) report an average price elasticity of the demand for Airbnb and hotel accommodations of -4.27. In Huang (2022), the average price elasticity of demand for Airbnb accommodation in San Francisco is -2.51. While our estimated elasticity is somewhat lower in absolute terms compared to the literature, we believe it is reasonable. It is possible that our estimate of the price coefficient reflects search frictions; if guests are not aware of all available listings, they would substitute to other options at a comparatively lower rate in response to a price increase, leading to relatively low price elasticity.

	mean	std	min	25%	50%	75%	max
Own-price elasticity	-1.0426	0.7294	-5.4299	-1.4287	-0.9022	-0.4941	-0.0005
Good review semi-elasticity	2.5640	2.7082	0.0041	0.6497	1.3687	3.6880	16.6112

Table 6: Summary of elasticities.

We also compute the semi-elasticity of demand if a listing receives a good review. On average, the occupancy rate increases by 2.56% in response to a good review. Figure 8 shows that the demand for listings with fewer reviews is more elastic. Guests who see an additional review will adjust their expectations of a listing's quality only slightly if the listing already has many reviews and the rating is precise. Conversely, observing a review has a much greater impact on guests' beliefs if the listing has few or no reviews and there was previously little information available to assess its quality. Accordingly, the average good review semi-elasticity is 5.80 for unreviewed listings but a mere 0.68 for listings with 19 reviews. Table 6 provides summary statistics for the two elasticities we estimate.

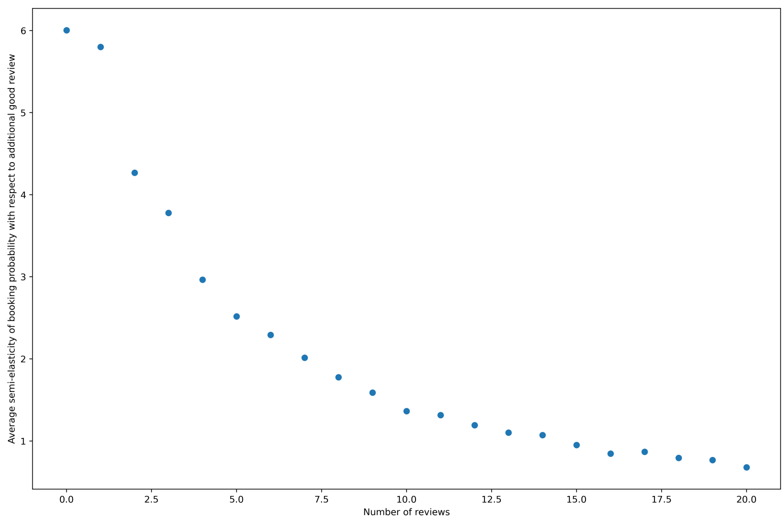

Figure 8: Mean good review semi-elasticity conditional on the number of reviews.

The prior distribution parameter estimates, together with the estimates of the expected quality and the rental rate, imply that guests' willingness to pay is *ceteris paribus* $11.56 higher per day for a listing with one review and a 5-star rating compared to an unreviewed listing. By contrast, an additional good review of a listing with 19 reviews and a four-star rating is worth only $3.34 per day to guests.

Recall that Airbnb has a low single-digit market share. The type coefficients are estimated to be negative and large in magnitude, reflecting that most guests take the outside option (i.e., book a hotel) rather than book an Airbnb. Remember that a listing is attributed a higher type if it has a higher occupancy rate after accounting for the rental rate and the number of (good) reviews. Hence, the type coefficients increase in the type.

7 Supply Estimation

We estimate the cost parameters of the model, denoted by c, conditional on the demand parameter estimates $\theta = \{\psi, \iota, \alpha, \beta_{,1}, \beta_2, \beta_3, \beta_4, \gamma\}$, by Maximum Likelihood. The log-likelihood of the number of listings across states is depicted in Equation (14). $s^*(x|c,\theta)$ denotes the number of listings in state x in the oblivious equilibrium of our model. y_{jt} is a variable indicating the state of listing j in month t. $s^*(x|c,\theta)$ is determined by the demand parameters (as they determine how frequently listings change states) and the cost parameters (because the listings' entry and exit rates depend on them). Conditional on the demand parameters, the operating costs, and the entry costs are identified by variation in the number of listings across states. We assume that listings are equally likely to have any of the four types. Hence, the average number of listings of a certain type outside the market is 2,500 minus the total number of listings of that type in the market. Denote the set of states associated with type l by X_l

$$\begin{aligned} L(c|y_{jt},\theta) = \sum_t \Bigg\{ & \sum_j \left[\sum_{x\in X} \mathbb{1}(y_{jt}=x) \ln(s^*(x|c,\theta)) \right] \\ & + \sum_{l\in\{1,2,3,4\}} \left(\frac{J}{4} - \sum_{x\in X_l} \sum_j \mathbb{1}(y_{jt}=x) \right) \ln\left(\frac{J}{4} - \sum_{x\in X_l} s^*(x|c,\theta) \right) \Bigg\} \end{aligned} \tag{14}$$

The maximum likelihood estimation requires us to solve the model and determine $\mathbf{s}^*$ repeatedly for different cost parameter candidates. We solve the model in four steps. We start by formulating the initial guess of the pricing policy function $P_0(x)$, the number of listings $s_0(x)$ per state, and the host's value function $V_0(x)$. Initially, half of the hosts are active and half of them are inactive. All hosts charge a rate of $200 and their value function is the present discounted value of their revenue. The number of listings is identical across states.

Step 1 – Based on the guess, we determine a host's best response $P_1(x)$ if all remaining hosts adhere to $P_0(x)$. We solve for $P_1(x)$ using Newton's method. Specifically, we iterate over Equation (15), where k is the iteration step and $P_1^0 = P_0(x)$, until the change in the rental rate, $P_1^{k+1} - P_1^k$ for any state x is less than $0.1.

$$P_1^{k+1}(x) = P_1^k(x) - \frac{v'(x)}{v''(x)} \tag{15}$$

Functions v' and v'' are the respective first- and second-order derivatives of the host's value function with respect to the rental rate.[17]

Step 2 – Assuming that all hosts set their rates according to $P_1(x)$, we compute the occupancy rate $q_1(x)$ and transition matrix $T_1(x)$. We use them to compute the value function $V_1(x)$ from Equation (10).

Step 3 – We use $V_1(x)$ to update the exit rate in state x to $\chi_1(x)$ and the entry rate for each type l to λ_l. Together with the occupancy rates $\mathbf{q}_1$, $\boldsymbol{\chi}_1$ and $\boldsymbol{\lambda}_1$ allow us to compute the expanded transition matrix $\mathbf{F}_1$. We use $\mathbf{F}_1$ to solve for the new, stationary listing distribution $\mathbf{s}_1$, where $\mathbf{s}_1$ is given by:

$$\mathbf{s}_1 = \mathbf{s}_1 \mathbf{F}_1 \tag{16}$$

Step 4 – If the absolute difference between $V_1(x)$ and $V_0(x)$ or P_1 and V_0 or $s_1(x)$ and $s_0(x)$ for any x exceeds 0.000001, we update the guess to $P_1(x), s_1(x)$, and $V_1(x)$ for all x and repeat steps 1 to 3. Otherwise, $(\mathbf{P}_1, \mathbf{s}_1, \mathbf{V}_1)$ constitutes the model solution $(\mathbf{P}^*, \mathbf{s}^*, \mathbf{V}^*)$.

Doraszelski and Satterthwaite (2010) establish the existence of a symmetric equilibrium in pure strategies for a closely related model. Note that the model may have multiple equilibria. (Doraszelski and Satterthwaite, 2010; Weintraub et al., 2008).

The estimation results are shown in Table 7. All estimated parameters are highly significant. Our estimates imply that hosts on average incur a \$3,065 cost upon entry.[18] This compares to the average entrant's lifetime profit of \$5,940 in present discounted value terms. Listings of any two types are by construction equally likely observed in the market although higher type listings are more profitable. Hence, the entry costs increase in listing type. Our estimates suggest that the next higher type requires 30 to 75% higher entry costs. For instance, the entry costs of a type-3 listing are on average \$898 or 33% higher than a type-2 listing.

We estimate that on average hosts pay \$2,584 in operating costs. To compare, the average revenue per month is \$3,422. This implies an average profit margin of 24.5%. As with entry costs, higher-type listings incur higher operating costs. Intuitively, higher-type listings are more desirable to guests but also the host. The host's opportunity cost of renting out the apartment, rather than using it herself, is therefore higher. A type-4 listing, for example, costs on average \$3,656 or 173% more to maintain than a type-1 listing. Nonetheless, hosts earn more profit from higher-type listings. The present discounted values of type-1, type-2, type-3, and type-4 listings are \$5,213, \$8,510, \$11,481, and \$13,235 respectively.

8 Model Fit

We use our estimates to simulate four years worth of data. We compare the simulated data to the actual data to assess the fit of our model. Table 8 shows key simulated data moments. As is to be expected, most variables exhibit less variation in the simulated data compared to the real data. The average rental rate is roughly 5% lower than in the actual data. The average

[17] The exact expressions can be found in the Appendix A.5.

[18] Note that the average incurred entry cost is by many orders of magnitude lower than the average entry cost draw as firms will only enter and incur the cost if the draw is low enough.

		(1)
mean entry cost	$\bar{\kappa}_1$	181,368***
		(3,946.58)
	$\bar{\kappa}_2$	264,819***
		(5,136.71)
	$\bar{\kappa}_3$	426,997***
		(8,345.07)
	$\bar{\kappa}_4$	796,930***
		(11,485.67)
mean operating cost	$\bar{\phi}_1$	2,323***
		(1.40)
	$\bar{\phi}_2$	3,587***
		(2.02)
	$\bar{\phi}_3$	4,345***
		(2.79)
	$\bar{\phi}_4$	5,572***
		(3.19)
Observations		62,937

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors in parenthesis.

Table 7: Supply estimates.

occupancy rate in the simulated data is almost identical to the one in the real data. Nonetheless, the simulated listings have on average fewer reviews (7) than those in the actual data (10). This is likely because hosts tend to exit at a relatively higher rate in the simulated data. The average monthly exit rates are 12.88% and 3.21% in the simulated and actual data respectively. The average rating in the simulated data is 4.63 stars, which is comparable to the 4.51-star average rating in the actual data.

	mean	std	min	25%	50%	75%	max
Rental rate	$182.58	$31.53	$119.31	$160.41	$182.07	$209.10	$248.22
Occupancy rate	60.01%	48.99%	0.00%	0.00%	0.00%	100.00%	100.00%
Number of reviews	7.32	7.60	0.00	1.00	4.00	14.00	20.00
Rating	4.63	0.86	1.0	4.71	5.00	5.00	5.00
(Monthly) exit rate	12.88%	0.33%	0.00%	0.00%	0.00%	0.00%	100.00%

Table 8: Simulated data summary.

Figure 9 compares the rating distribution in the simulated data ("rating (model)") to the rating distribution in the actual data ("rating (data)"). The frequency of one-star, three-star, $3\frac{2}{3}$-star ratings (i.e., ratings comprised of two good reviews and one bad review), and four-star listings (i.e., ratings comprised of three good reviews and one bad review) are remarkably similar in the simulated and the actual data. The model predicts more five-star ratings than we find in the actual data. In the actual data, listings at the high end of the distribution typically have ratings between four and five stars. We believe that the difference can be explained by the fact that the number of reviews is censored at 20. Hence, the model allows for little variation in the ratings close to 5 stars. Ratings in the simulated data are based on at most 20 reviews but possibly on many more in the actual data. Note that Figure 9 illustrates the selection effect that underlies the simulated and the actual data. Recall that the mean quality of *entrant* listings is estimated to be 4.15 stars. This is lower than the average rating of *all active* listings, both in the simulated and actual data. the 25th percentile is also fairly similar in the simulated and actual data. Less than 25% of listings have a quality of around 4.5 stars or less.

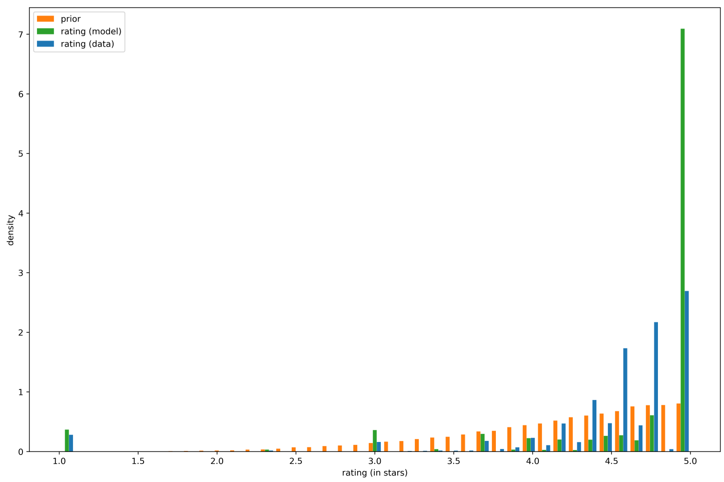

Figure 9: Estimated prior distribution, rating distribution in the data, and model rating distribution.

Recall that we estimate the cost parameters of the model by maximizing the likelihood of the equilibrium state distribution. According to our model, in equilibrium, there are on average 1,152 active listings, compared to 1,210 in the data. Figure 10 (bottom left) shows that the number of listings generated by the model for each number of (good) reviews closely matches the observed data. As in the actual data, around 20% of active listings are unreviewed. The model predicts that there are few to no poorly rated listings, especially for states associated with a relatively large number of reviews. Highly rated listings, on the other hand, tend to remain in the market and gather reviews. The model gives rise to a "pitched-fork" shaped distribution which is consistent with the data. As discussed, the model produces rental rates and occupancy rates that are comparable to the actual data (Figure 10, top right and top left, respectively).

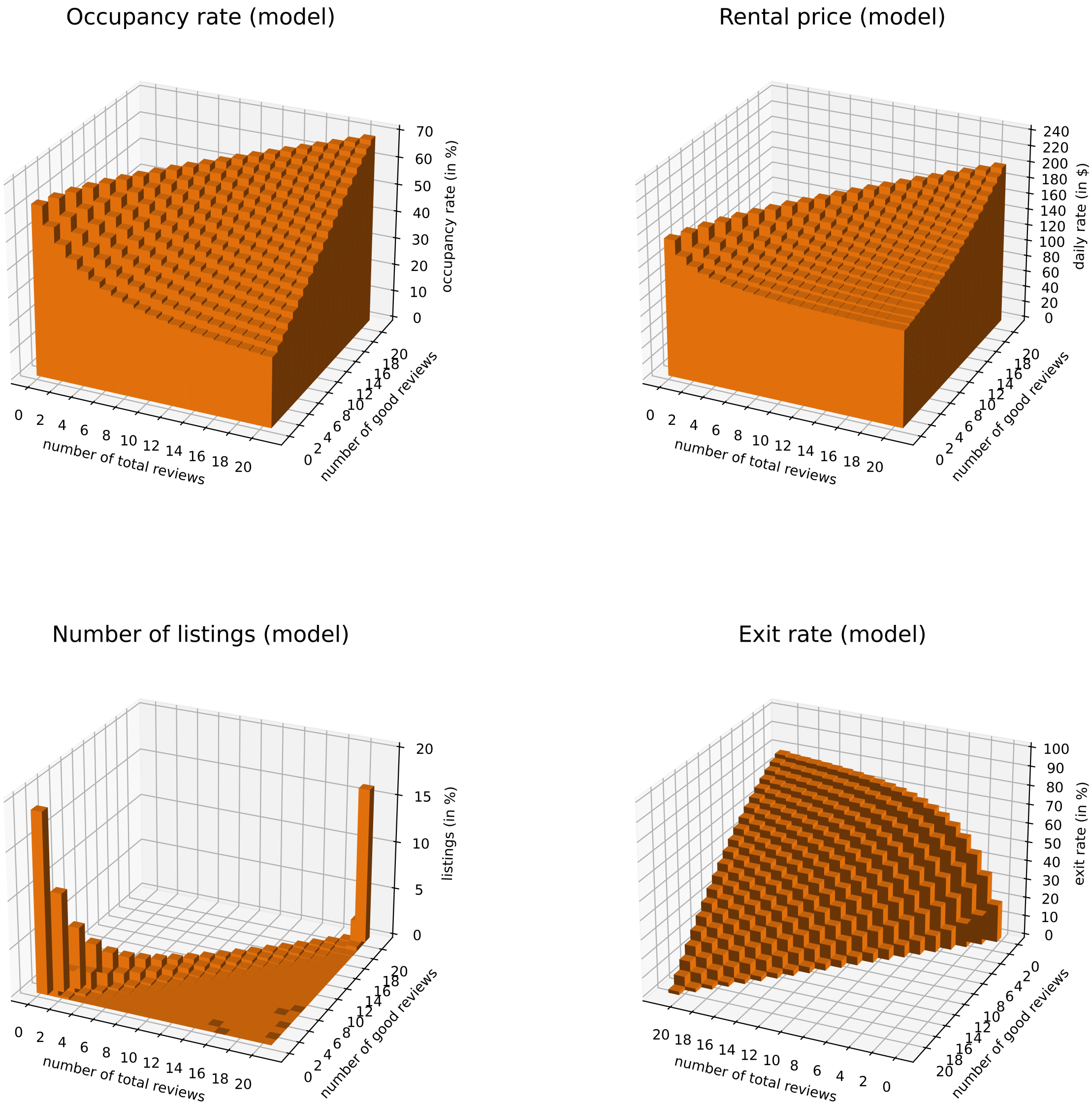

Figure 10: Estimated occupancy rate (top left), rental rate (top right), number of listings (bottom left), and exit rate (bottom right) by number of (good) reviews.

It is critical for our counterfactual analysis that we get the behavior of rental rates, occupancy rates, and exit rates in the number of (good) reviews right. Due to the selection effect, we do not have data for many states. For the states we do have data for, the occupancy rate seems to respond more strongly to the accumulation of reviews compared to the model. Changes in rental rates, on the other hand, appear to be similar in the data and the model. To validate these observations, we repeated the regressions from Table 3 using the simulated data. The estimation results are shown in Table 9. Again, we translate these estimates into the impact of an additional positive review on each outcome variable for a listing with a 4.5-star rating and ten reviews, which are values close to their averages in the actual data. For ease of comparison, we report the prediction using the estimates from Table 3 based on the actual data in parentheses. For the simulated data, we find that *ceteris paribus* a good review raises the rental rate of by \$1.55 (\$0.98), whereas a bad review decreases the rental rate by \$11.79 (\$10.55). Guests generally expect high listing quality and significantly lower their expectations, when being confronted with a negative review. The host must drastically reduce the rate to maintain the listing's appeal.

	Dependent variable:		
	rental rate	occ rate	exit rate
	(1)	(2)	(3)
rental rate		-0.006^{**}	-0.001
		(0.003)	(0.002)
no of reviews	-10.872^{***}	-0.071^{***}	0.071^{***}
	(0.705)	(0.021)	(0.026)
no of good reviews	14.497^{***}	0.095^{***}	-0.095^{***}
	(1.079)	(0.030)	(0.036)
no of reviews × no of good reviews	-0.105^{***}	-0.001^{***}	0.001^{***}
	(0.012)	(0.000)	(0.000)
Type FE	Yes	Yes	Yes
Observations	59,904	59,904	59,904

Note: *p<0.1; **p<0.05; ***p<0.01. Standard errors in parenthesis clustered at type level.

Table 9: Reduced form regression results with simulated data.

Moreover, Table 3 suggests that a good review increases the occupancy rate of the average host's listing by 0.3 (1.3) percentage points and a bad review decreases it by 8.1 (1.08) percentage points, all else equal. Unlike the hosts in our model, actual hosts might not accurately reflect the rating of their listing in their pricing strategies, as doing so requires solving a complex, dynamic problem. Figure 10 (bottom right) presents the exit rate in the model equilibrium. As expected, state associated with few good reviews feature high exit rates. Hosts of listings with a good rating, on the other hand, are relatively unlikely to leave the market. Directly comparing the model's exit rates with the data is challenging, as our dataset includes exit events for only a few states and for many of these states exit events rarely occur. Our reduced-form estimates suggest that the rating matters more for hosts' exit decisions in the model than in the data. All else equal, the average host is 0.4 (0.2) percentage points less likely to exit the market on any given day after receiving a good review; the monthly exit rate increases by 7.8 (0.2) percentage points if the review is bad.

Despite its limitations, our model effectively captures many key aspects of the data. Notably, the model produces a state distribution similar to the empirical one. Additionally, the direction and magnitude of the responses of hosts and guests to reviews in the model are broadly consistent with the actual data. Our counterfactual policies aim to increase the rate at which the quality of new listings is revealed by boosting their occupancy rates. According to our model, there is significant potential to enhance the performance of new listings. Currently, a new listing's occupancy rate is typically 8% below the average, and its rental rate is $18 lower than the average. Taken together, new listings generate 17% less revenue than the average listing. Consequently, our model predicts that new listings are around 60% more likely to exit the market in a given month compared to the average listing. In the actual data, new listings are booked 31% less frequently and generate 31% less revenue than the average listing. Therefore, we believe that our counterfactual analysis will yield relatively conservative estimates of the welfare effects of the policies we examine.

9 Counterfactual analysis

Having estimated the model's parameters, we use it to explore how rental rates would need to change to ameliorate the cold start problem. Our welfare results suggest the extent of the underlying inefficiency in welfare terms. Our results also demonstrate the efficacy of a two-part tariff regime in implementing socially desirable prices.

We consider per-booking subsidies on unreviewed listings ("entrants") and reviewed listings ("incumbents") respectively. We allow the subsidies to take negative values, in which case we refer to them as taxes. Equation (17) shows the mean utility of a listing in state x, where the subsidy is denoted by τ.

$$v(P(x), x) = \gamma \frac{a + K(x)}{a + b + N(x)} + \beta(x) + \alpha((1 + f)P(x) - \tau(x)) \tag{17}$$

We search for the subsidy or tax values that maximize long-run equilibrium welfare. Based on our analysis in Section 3, it is unclear whether a cold-start problem exists, but if it does, we expect the optimal subsidy-tax scheme to widen the rental rate gap between entrants and incumbents. A relatively higher subsidy or lower tax diverts demand from reviewed listings to those with no reviews. A guest booking an unreviewed listing benefits future guests by potentially uncovering a higher-than-expected quality. However, this comes at the cost of not choosing a frequently reviewed, high-quality listing. The better rental rate compensates the guest who makes the booking for this opportunity cost. By contrast, guests are penalized for booking listings with well-known quality instead of choosing a lesser-known listing and potentially reviewing it to inform future guests. In this regard, the policy is Pigouvian. It aims to internalize the positive externality of booking an unreviewed listing instead of a reviewed listing.

To make sure the tax-subsidy scheme is revenue neutral and effects are not driven by pouring government money into the market, we implement a lump-sum tax on hosts with subsidized listings, equal to the total subsidies they received in a given month. Similarly, hosts are fully reimbursed for the tax levied on their listings through a lump-sum subsidy. This approach mitigates the direct impact of the policy on host profits: hosts effectively finance the subsidy and receive the tax revenue of their own listings. It is important to recognize that combining per-unit and lump-sum taxes and subsidies in this way is equivalent to directly regulating hosts' rental rates.

It is important to note, that not only the price difference between incumbent and entrant, but also the overall price level may be sub-optimal from a social welfare perspective. However, welfare gains or losses from differences in the price level are unrelated to the cold-start problem, for reasons we discuss in more details in the Section 9.3. Therefore, in our main counterfactual in Section 9.2, we constrain the tax-subsidy scheme we search over by imposing the restriction that the average tax/subsidy is zero, and therefore prices are largely unaffected. We explore the welfare gain from social learning when allowing the price level to be affected by the tax-subsidy regime in a second counterfactual in Section 9.3.

9.1 Welfare

Generally, we determine the tax-subsidy scheme that maximizes social welfare by solving the model with varying amounts of subsidies and taxes. We calculate the welfare change as the change in the sum of producer profits, the total entry cost, and the compensating variation in response to the policy. Since the policy is revenue-neutral, we do not need to separately account for the subsidy cost and tax revenue.

The monthly producer profit, i.e. the host revenue less the operating costs, is characterized in Equation (23), where all variables take their new oblivious equilibrium values. The first term of Equation (23) captures the hosts' revenues, Airbnb fees, tax revenue, and subsidy cost. The second term represents the operating cost. For a derivation of the expected operating cost, see Appendix A.7.

$$\sum_{x}^{X} s(x)\left(28q(x)((1+f)P(x)-\tau(x)) - ((1-\chi(x))\bar{\phi}(x) - \chi(x)\delta T(x)V(x))\right) \tag{18}$$

Equation (19) describes the total entry cost in equilibrium. The entry cost is the number of potential entrants times their type-specific, expected cost of entry. the number of potential type-l entrants equals the total number of listings of that type less the number of type-l listings that are already active. We derive of the expected entry cost in Appendix A.8.

$$\sum_{l}\left(\frac{J}{4} - \sum_{x}^{X_l} s(x)\right)(\lambda_l \bar{\kappa}_l - (1-\lambda_l)\delta V((0,0,l))) \tag{19}$$

Recall that each listing-week can only be booked once. Ignoring the capacity constraint of listings risks overstating the compensating variation of the subsidy. To address this, we assume that if multiple consumers want to book the same listing, only one makes the booking, while the others must choose the outside option, such as booking a hotel room. As shown by Williams (1977) and Small and Rosen (1981), if the random taste shocks are independently and identically Gumbel distributed and the utility is linear in income, the expected utility of a single consumer is the natural logarithm of the sum of the mean utilities, $\ln(1 + \sum_x^X s(x)\nu(P(x), x))$, plus a constant of integration. If capacities were not constrained, consumer surplus would be the total market size 28μ times the expected utility. However, since capacities are constrained, not every guest can book their preferred listing. From Equation (5), we know that under the capacity constraint, $q(x)$ consumers book a particular listing in state x in expectation. Conversely, the expected number of guests who want to book that listing is $-\ln(1 - q(x))$, which is greater than $q(x)$. When calculating consumer surplus, we correct for this difference, as shown in Equation (20).

$$28\left(\mu - \sum_x^X s(x)(-\ln(1 - q(x)) - q(x))\right) \ln\left(1 + \sum_x^X s(x)\nu(P(x), x)\right) + \text{constant} \tag{20}$$

We calculate the compensating variation as the change in consumer surplus divided by the price coefficient α (McFadden, 2012; Small and Rosen, 1981).

9.2 The cost of the cold-start problem without aggregate tax distortions

Our focus is the inefficiency due to social learning only. Therefore, our main counterfactual restricts the feasible set of taxes and subsidies over which we optimize. Specifically, we impose that taxes and subsidies cancel out in the aggregate, i.e., the average tax level in the whole Airbnb market is zero, such that taxes do not mechanically expand or contract the Airbnb market. First, this ensures that the optimal tax/subsidy scheme we find does not mechanically expand or contract the entire Airbnb market, which will have welfare effects on its own. Second, the inefficiency due to social learning, or the extent of the cold-start problem, may depend on the market size of Airbnb. In other words, by restricting the feasible set of taxes and subsidies, we only capture the eliminated inefficiency due to the excessively slow pace of social learning at the current market size.

We find that implementing a \$14.86 tax per booking on incumbent listings and a \$89.32 subsidy on bookings of entrant listings maximizes welfare while leading to no tax burden in the aggregate. This policy increases welfare by \$177,241 per month, with the consumer surplus overcompensating a decrease in producer surplus. As shown in Table 10, guests pay rental rates that are only 0.8% higher than without the tax-subsidy scheme on average. Crucially, prices paid by consumers for a new listing are lower by almost 30% while incumbent listings charge almost 5% more. This translates into a substantially higher occupancy rate for entrants. Still, the number of entrant listings decreases for two reasons. First, despite the higher occupancy rate, we impose a privately sub-optimal price on the entrant, which reduces its revenue as the subsidy is deducted as a lump-sum tax again. Second, the high occupancy rate implies that entrants accumulate reviews faster and therefore transition faster to become an incumbent.

	total		*entrant*		*incumbent*	
		in %		in %		in %
Δ per-period welfare	\$177,241	–	–	–	–	–
Δ per-period consumer surplus	\$207,910	–	–	–	–	–
Δ per-period profit	-\$30,700	–	–	–	–	–
Average tax	\$0.00	0.00%	-\$89.32	-54.58%	\$14.86	8.12%
Δ average net rental rate	\$1.42	0.78%	\$40.25	24.41%	-\$6.23	-4.59%
Δ average gross rental rate	\$1.42	0.78%	-\$49.07	-29.76%	\$8.64	4.62%
Δ # listings	27	2.35%	-49	-22.73%	76	8.19%
Δ average occupancy rate	0.90 pp	1.39%	15.15 pp	25.34%	-1.82 pp	-2.75%

Note: Percentage changes are indicated relative to prices, number of listings, or occupancy without taxes.

Table 10: Effects of the constrained optimal tax-subsidy scheme.

One should also note that taxing all incumbent hosts raises their prices in a coordinated fashion and increases their revenue (despite the new prices being privately sub-optimal). Consequently, incumbent hosts remain in the market longer, which, together with the faster transition from the entrant state, results in approximately 8% more incumbents and a 1.4% increase in the average number of listings. As shown in Figure 11 (bottom left), the number of listings with 20 reviews increases by about 30% compared to the scenario without taxes. This increase is driven by a significant reduction in exit rates, which drop by 7% on average and up to 20% (bottom right).

Still, hosts are slightly worse off overall and experience a decrease in their aggregate profits by \$30,700. This decrease is more than compensated by the increase in consumer surplus of \$207,910 as consumers value the increase in (good) booking options. The tax-subsidy scheme incentivizes guests to book and review new entrants, thereby increasing the availability of information about product quality for all guests and facilitating the identification of more high-quality listings.

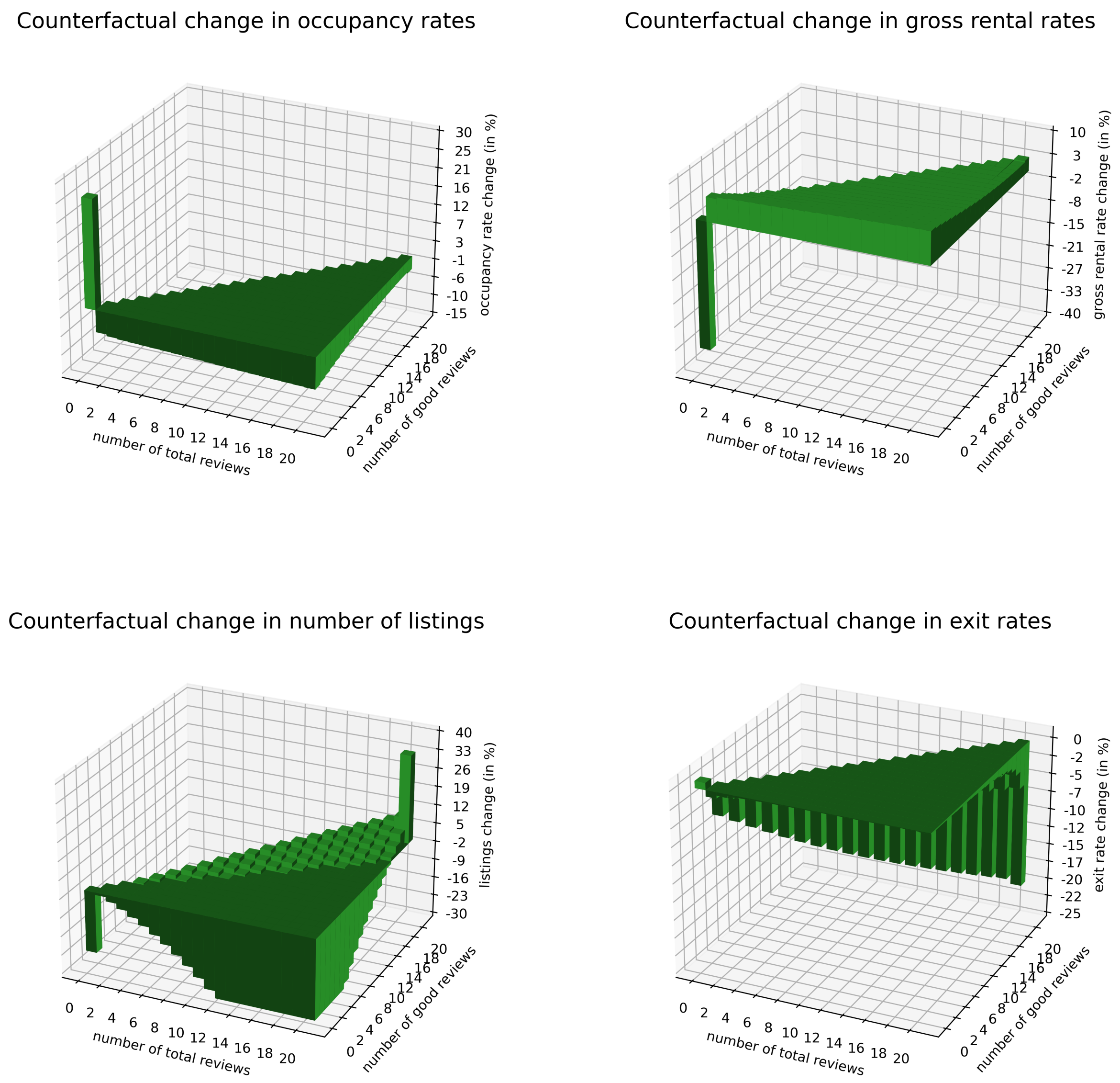

Figure 11: Counterfactual changes in occupancy rate (top left), rental rate (top right), number of listings (bottom left), and exit rate (bottom right) by number of (good) reviews.

This is illustrated in Figure 11 (top left). The occupancy rate of entrant listings increases by 25.3%, while incumbent listings see an average decrease of 2.7%. This disparity is due to an increase in the price difference as a result of the tax-subsidy scheme. Guests pay an average of \$116 for an entrant listing and \$196 for an incumbent listing — a \$80 difference (top right). At \$22, the rental rate difference between entrant and incumbent listings is much lower without the tax. Among incumbents, the change in the gross rental rate varies across states. Although all incumbent listings are taxed equally, incumbent hosts adjust their rates to different extents. In other words, the pass-through rate is state-specific. Listings with relatively high mean utility experience a relatively smaller increase in gross rental rates in response to the tax, as these hosts reduce their prices more significantly. Hence, the gross rental rates of highly rated listings tend to be closer to their pre-tax levels.

At the annualized interest rate of 6.7% used in our model, the social welfare effect is valued at approximately \$35.45 million in present discounted terms. This is significant. For comparison, the present discounted value (PDV) of all Airbnb host revenues is \$788 million, and the PDV of all Airbnb host profits is \$193 million. The welfare effect is roughly 30 times larger than the average lifetime profit of an entrant, which is \$5940.

9.3 The cost of social learning at the optimal aggregate tax level

If we do not impose that the counterfactual policy must satisfy an average tax/subsidy of zero, the optimal tax-subsidy scheme may affect market efficiency for reasons beyond the cold start problem. If there is a strictly positive or negative average tax burden, the policy will not only affect the demand for entrants compared to incumbents, but also change the aggregate demand for Airbnb listings compared to hotels overall.

First, a change in the overall tax level will affect Airbnb's overall market-share. An average tax leads to a higher price level and distorts guests' choices towards the outside option. Conversely, an average subsidy will decrease the overall price level. Since marginal costs are zero, this will allow guests to pay rates closer to the efficient prices. In other words, a lower price level generally decreases the deadweight loss and therefore increases total welfare in a static setting. Second, in a dynamic market as the one we model, differences in profits affect entry and exit. Since we consider a two-part tariff which ensures revenue neutrality, listings would always be worse if they were the only one affected by taxes/subsidies. However, rival listings are affected as well, and therefore listings may actually benefit from some tax/subsidy schemes. A higher aggregate price level due to positive average taxes may therefore increase profits and incentivizes more hosts to enter Airbnb. Mankiw and Whinston (1986) points out that if market entry is free, the number of firms can be too small because firms do not account for the social benefit of introducing new varieties to the market. Indeed, there are inefficiently few active listings. Note that the change in aggregate host profitability does not pertain to social learning.

Hence, the welfare effect will conflate the effect of a change in the tax level and the cold-start problem. As mentioned in the previous section, there may also be interactions of different effects, so the extent of the cold-start problem may vary depending on the average tax rate. Therefore, in addition to the total welfare effect described in this section, we also identify the extent of the cold-start problem at the new average tax level in Section 9.3.1.

Without imposing a zero average tax, we find that the welfare-maximizing tax-subsidy scheme is a \$41.86 tax per booking on incumbent listings and a \$1.15 subsidy on bookings of entrant listings. The policy increases welfare by \$453,191 per month, with approximately 85% of this increase attributed to a larger consumer surplus and the remainder to a larger producer surplus. Both consumers and producers (i.e., Airbnb and hosts) benefit from this tax-subsidy scheme on incumbents and entrants. Similar to arguments in the previous section, incumbent listings are better off because of a coordinated price increase, an effect that is now stronger than before as the price increase is larger. This leads to an approximately 20% increase in the average number of listings, while the number of entrants again decreases.

	total		*entrant*		*incumbent*	
		in %		in %		in %
Δ per-period welfare	$453,191	–	–	–	–	–
Δ per-period consumer surplus	$386,078	–	–	–	–	–
Δ per-period profits	$67,113	–	–	–	–	–
Average tax	$35.20	19.26%	-$1.15	-0.71%	$41.86	22.44%
Δ average net rental rate	-$12.44	-6.81%	-$1.30	-0.79%	-$15.37	-8.22%
Δ average gross rental rate	$22.76	12.45%	-$2.45	-1.48%	$26.50	14.17%
Δ # listings	239	20.74%	-2	-1.02%	241	25.81%
Δ average occupancy rate	-3.78 pp	-5.80%	0.43 pp	0.71%	-4.82 pp	-7.25%

Note: Percentage changes are indicated relative to prices, number of listings, or occupancy without taxes.

Table 11: Effects of the optimal tax-subsidy scheme.

Despite much higher prices, the increase in booking options again benefits guests who value variety and are better off by $389,078. Also, hosts are better off and experience a slight increase in aggregate profits by $67,113.

It is crucial to recognize that, even though the policy conflates welfare effects due to the change in price level, it also incorporates a change in the speed of social learning and it therefore has implications for the cold-start problem. Specifically, entrants are taxed to a lesser extent compared to incumbents, making them relatively cheaper. As in the previous counterfactual, this increases the pace of learning about entrant quality for all guests.

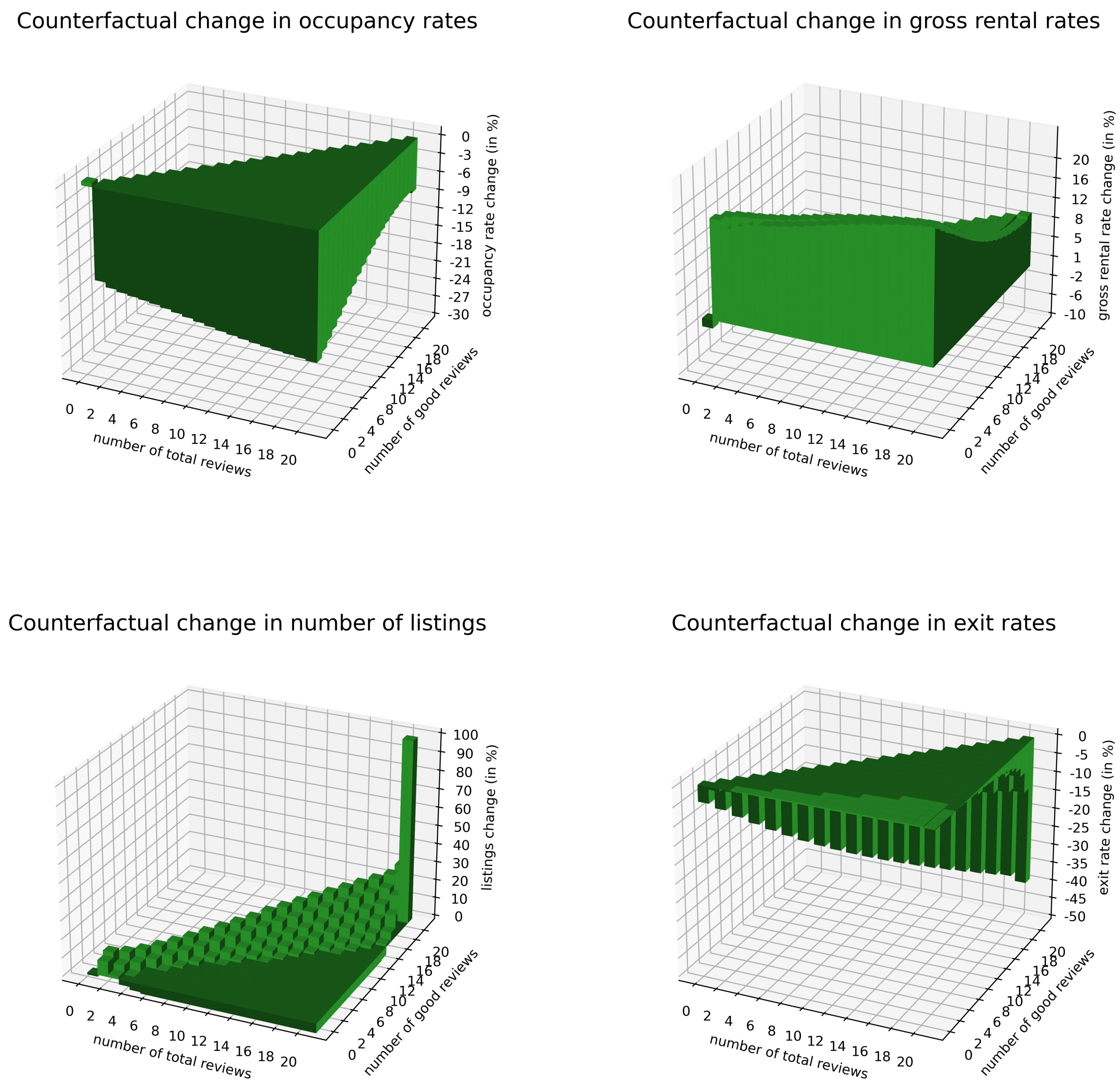

Figure 12: Counterfactual changes in occupancy rate (top left), rental rate (top right), number of listings (bottom left), and exit rate (bottom right) by number of (good) reviews.

This is illustrated in Figure 12 (top left). The occupancy rate of entrant listings slightly increases by 0.7%, while incumbent listings see an average decrease of 7.3%. This disparity is due to an increase in the price difference as a result of the tax-subsidy scheme. Guests pay an average of $162 for an entrant listing and $213 for an incumbent listing — a $51 difference (top right). Without the tax, the rental rate difference between entrant and incumbent listings is much lower, at $22.

9.3.1 Decomposition

We are primarily interested in the welfare change resulting from a change in the speed of social learning due to increasing the relative attractiveness of entrants, rather than the overall effect of the taxes. To do this, we determine and subtract the contribution of the change in the overall tax level to welfare under the optimal scheme of a \$1.15 subsidy for entrants and a \$41.86 tax for incumbents. Specifically, we isolate the effect of the change in tax level by computing a counterfactual scenario where we impose uniform taxes amounting to the average tax of \$35.2. The welfare difference from this counterfactual to the baseline welfare will be solely driven by imposing a positive average tax on all market participant. By subtracting the welfare effect of a change in tax levels, we are left with the welfare component that is due to the difference in taxes around the new level for entrant and incumbent. We attribute this component to the larger speed of social learning, as it is the tax difference which makes unreviewed entrant listings relatively cheaper and more attractive.

	Δ W	Δ CS	Δ PS
Total effect	\$453,191	\$386,078	\$67,113
- Tax level effect	\$421,842	\$350,264	\$71,578
= Social learning effect	\$31,349	\$35,815	-\$4,466

Table 12: Counterfactual decomposition.

	total		*entrant*		*incumbent*	
	in \$	in %	in \$	in %	in \$	in %
Δ Average tax	\$0.00	0.00%	-\$36.35	-19.90%	\$6.66	3.24%
Δ net rental rate	\$0.51	0.30%	\$13.55	9.03%	-\$2.29	-1.32%
Δ gross rental rate	\$0.51	0.25%	-\$22.80	-12.31%	\$4.38	2.09%
Δ # listings	-2	-0.15%	-20	-8.62%	18	1.58%
Δ occupancy rate	0.46 pp	0.76%	6.49 pp	12.08%	-0.79 pp	-1.27%

Table 13: Comparing the optimal tax-subsidy scheme to a uniform tax of \$35.20.

Table 13 shows that, the optimal tax-subsidy scheme leads to lower prices for entrants and higher prices for incumbents compared to a uniform tax, increasing the price gap by around \$27. This translates into an increase in occupancy for entrants, while the occupancy rate for incumbents slightly decreases. Overall, the number of listings actually decreases, driven by the decrease in the number of entrants. Despite the decrease in listings, increasing the price gap increases total welfare by \$31,349. Hence, at the new average tax level of \$35.2, increasing the pace of social learning seems to lead to a smaller welfare increase compared the main counterfactual where we impose a zero average tax.

9.4 Discussion

By implementing a tax policy, we have demonstrated that significant welfare gains can be achieved by addressing the cold start problem in the Airbnb market. However, it is important to note that we do not advocate for the implementation of this policy. Market conditions have evolved since the period of our study from 2016 to 2019. As of September 5th, 2023, the City of New York has mandated that short-term rentals must register with the Mayor's Office of Special Enforcement.[20] This measure aims to curb short-term home-sharing and support the long-term rental market.

Calder-Wang (2021) finds that although Airbnb's presence generally contributes positively to welfare, it has the adverse consequence of increasing rents in New York City, thereby making most households worse off. Our model does not account for housing market effects and they are not covered by our analysis.

The primary objective of our study is to illustrate how the inefficiently slow speed of learning affects the market and to quantify the extent of this problem. The tax policy we analyze serves as an auxiliary tool to highlight these effects. The general nature of the problem suggests that virtually any digital market with a review system may suffer from significant welfare losses due to the too-slow accumulation of reviews.

10 Conclusion

In this paper, we estimate a model of the Airbnb market that incorporates social learning through peer-to-peer reviews. Our model addresses the cold-start problem, where neither Airbnb guests nor hosts fully account for the social value of new listings being booked and reviewed. We find that welfare would significantly increase if the rental rate difference between new and established listings were larger by $104, or 64% of the average rate for a new listing. This adjustment would shift demand from established to new listings. Consequently, in the new market equilibrium, guests leave reviews that provide more information about listing quality. As guests become better informed about listing quality, consumer surplus increases.

We interpret the welfare change an estimate of the welfare cost attributed to the cold-start problem at current average tax levels of zero. Our results indicate that the cold-start problem significantly impacts Airbnb, with a welfare loss of $177,241 per month in Manhattan alone. We believe that the significance of the cold-start problem we identified for Airbnb extends to many other markets, both online and offline, that feature social learning. It would be interesting to see a study like ours applied to a different market.

Our model is agnostic about why guests leave reviews. Future research could model the decision-making process behind guest reviews which would allow testing the long-term effects of policies aimed at changing guests' reviewing behavior rather than their purchase behavior.

[20] See https://www.nyc.gov/site/specialenforcement/registration-law/registration.page.

For instance, in the experimental study by Fradkin and Holtz (2022), guests were offered a coupon to review listings that had no prior reviews. On Airbnb and many other digital markets, consumers view search results compiled by the platform based on their search queries, which is not modeled in this paper. Incorporating this into our model could open up opportunities for counterfactual scenarios that manipulate search results to address the cold-start problem. We leave this for future research.

Our paper demonstrates the inefficiency of social learning that may plague many digital markets. We examine one particular way platform providers or policymakers may address this inefficiency. Determining its relevance for other industries is left for future research. Doing this for traditional markets, where social learning usually happens through word-of-mouth, would be particularly interesting.

Appendix

A.1 Proof of Lemma 1

In $t = 1$, firm j maximizes

$$p_{j1}q_{j1} + v_r\delta q_{E1}\omega^e_{E1}\pi^g_{j2} + v_r\delta q_{E1}(1-\omega^e_{E1})\pi^b_{j2} + (1 - v_r\delta q_{E1})\pi^0_{j2}.$$

The necessary and sufficient conditions for a maximum are as follows.

$$q_{j1} - p^*_{j1}q_{j1}(1-q_{j1}) + v_r\delta q_{E1}(1-q_{E1})\frac{\partial(-p_{E1}+p_{I1})}{\partial p_{j1}} \sum_{i\in\{0,g,b\}} \phi_i\pi^i_{j2} = 0 \quad \text{(FOC)}$$

$$q_{j1}(1-q_{j1})(p^*_{j1}(1-q_{j1})-1) + v_r\delta q_{E1}(1-q_{E1})^2\left(\frac{\partial(-p_{E1}+p_{I1})}{\partial p_{j1}}\right)^2 \sum_{i\in\{0,g,b\}} \phi_i\pi^i_{j2} < 0 \quad \text{(SOC)}$$

Rearranging the FOC yields p^*_{j1}.

$$p^*_{j1} = \frac{1}{1-q^*_{j1}} + \frac{\partial(-p_{E1}+p_{I1})}{\partial p_{j1}} \sum_{i\in\{0,g,b\}} v_r\delta\phi_i\pi^i_{j2} \tag{21}$$

Substituting p^*_{j1} into the SOC reveals that the sufficient condition for a maximum is satisfied.

In $t = 2$, j maximizes $p^i_{j2}q^i_{j2}$. It is straightforward to verify that $p^*_{j2} = 1/(1-q_{j2})$. Hence, $\pi^i_{E2} = q^i_{E2}/(1-q^i_{E2})$ which is increasing and convex in ω^e_{E2}.

$$\frac{\partial}{\partial\omega^e_{E2}}\left(\frac{q_{E2}}{1-q_{E2}}\right) = \frac{q_{E2}(1-q_{E2})^2 + q^2_{E2}(1-q_{E2})}{(1-q_{E2})^2} = \frac{q_{E2}}{1-q_{E2}} > 0$$

$$\frac{\partial^2}{\partial(\omega^e_{E2})^2}\left(\frac{q_{E2}}{1-q_{E2}}\right) = \frac{q_{E2}}{1-q_{E2}} > 0$$

By Jensen's inequality $\sum_{i\in\{0,g,b\}} \phi_i\pi^i_{E2}$ is larger than zero.

$$\omega^e_{E1}\pi_{E2}\left(\frac{a+1}{a+b+1}\right) + (1-\omega^e_{E1})\pi_{E2}\left(\frac{a}{a+b+1}\right) - \pi_{E2}\left(\omega^e_{E1}\right)$$
$$> \pi_{E2}\left(\omega^e_{E1}\frac{a+1}{a+b+1} + (1-\omega^e_{E1})\frac{a}{a+b+1}\right) - \pi_{E2}(\omega^e_{E1}) = 0$$

Notice that $\partial p^*_{E1}/\partial v_r < 0$. As $q_{It} = 1 - q_{Et}$, $\pi^i_{I2} = (\pi^i_{E2})^{-1}$ and π_{I2} is decreasing and convex in ω^e_{E2}. Again, by Jensen's inequality $\sum_{i\in\{0,g,b\}} \phi_i\pi^i_{I2}$ is larger than zero and $\partial p^*_{I1}/\partial v_r > 0$.

A.2 Proof of Proposition 1

Part (i) The social planner solves the following maximization problem.

$$\max_{\tilde{p}_1} \left(cs_1 + \pi_{E1} + \pi_{I1} + \mathbb{E}_2[cs_2 + \pi_{E2} + \pi_{I2} | p_{E1}, p_{I1}]\right)$$

We write $\pi_{E1} + \pi_{I1}$ as $p_{I1} + q_{E1}\tilde{p}_1$. For brevity, denote $\ln\left(1 + \exp\left(\tilde{\omega}_2 - \tilde{p}_2\right)\right)$ by u_2, where $\tilde{\omega}_2$ and $\tilde{p}_2$ are the difference in second-stage quality expectations and Nash-equilibrium prices, respectively.

Before proceeding with the FOC, we establish two helpful facts:

1. u_2 is increasing and convex in w_{E2}:

$$\frac{\partial u_2}{\partial w_{2E}} = \frac{1}{\left(1 + \exp\left(\tilde{\omega}_2 - \tilde{p}_2\right)\right)} \exp\left(\tilde{\omega}_2 - \tilde{p}_2\right) = q_{E2} > 0$$
$$\frac{\partial^2 u_2}{\partial w_{2E}^2} = q_{E2}(1 - q_{E2}) > 0$$

2. $\pi_{E2} + \pi_{I2}$ is convex in w_{E2}:

$$\pi_{E2} + \pi_{I2} = \exp\left(\tilde{\omega}_2 - \tilde{p}_2\right) + \frac{1}{\exp\left(\tilde{\omega}_2 - \tilde{p}_2\right)}$$
$$\frac{\partial(\pi_{E2} + \pi_{I2})}{\partial w_{2E}} = \exp\left(\tilde{\omega}_2 - \tilde{p}_2\right) - \frac{1}{\exp\left(\tilde{\omega}_2 - \tilde{p}_2\right)}$$
$$\frac{\partial^2(\pi_{E2} + \pi_{I2})}{\partial w_{2E}^2} = \exp\left(\tilde{\omega}_2 - \tilde{p}_2\right) + \frac{1}{\exp\left(\tilde{\omega}_2 - \tilde{p}_2\right)} > 0$$

The necessary and sufficient conditions for a maximum are as follows.

$$-q_{E1}(1 - q_{E1})\tilde{p}_1^s + q_{E1} - q_{E1} - q_{E1}(1 - q_{E1}) \sum_{i \in \{0,g,b\}} v_r \delta \phi_i \left(u_2^i + \pi_{E2}^i + \pi_{I2}^i\right) = 0 \quad \text{(FOC)}$$
$$-q_{E1}(1 - q_{E1}) + q_{E1}\left(1 - q_{E1}\right)\left(2q_{E1} - 1\right)\tilde{p}_1^s - q\left(1 - q_{E1}\right)\left(2q_{E1} - 1\right) \sum_{i \in \{0,g,b\}} v_r \delta \phi_i \left(u_2^i + \pi_{E2}^i + \pi_{I2}^i\right) < 0 \quad \text{(SOC)}$$

Rearranging the FOC yields $\tilde{p}_1^s$.

$$\tilde{p}_1^s = -\sum_{i \in \{0,g,b\}} v_r \delta \phi_i \left(u_2^i + \pi_{E2}^i + \pi_{I2}^i\right) \tag{22}$$

From facts 1. and 2. and Jensen's inequality, it follows that $\tilde{p}_1^s < 0$. It easy to see that the SOC is satisfied at the socially optimal price difference, i

Part (ii) From Equation (21) and Equation (22), we know:

$$\tilde{p}_1^s < \tilde{p}_1^* \iff -\sum_i v_r \delta \phi_i u_2^i < \frac{1}{1 - q_{E1}^*(\tilde{\omega}_1)} - \frac{1}{q_{E1}^*(\tilde{\omega}_1)}$$

$$= \exp\left(\tilde{\omega}_1 - \tilde{p}_1^*(\tilde{\omega}_1)\right) - \frac{1}{\exp\left(\tilde{\omega}_1 - \tilde{p}_1^*(\tilde{\omega}_1)\right)} \tag{23}$$

We will show this part of the proposition in two steps: First, we show that the above inequality is satisfied at $\tilde{\omega}_1 = 0$. Second we show that, as $\tilde{\omega}_1$ increases, the increase in $\tilde{p}_1^s$ is smaller than the increase in $\tilde{p}_1^*$ for all $\tilde{\omega}_1$, as long as $\sum_i \delta v_r \phi_i(\pi^i_{I2} - \pi^i_{E2}) < 1$.

Step 1: At $\tilde{\omega}_1 = 0, \tilde{p}_1^s < \tilde{p}_1^*$: $\tilde{p}_1^*$ is given implicitly by the difference in Equation (21) for entrant and incumbent:

$$\tilde{p_1^*}(\tilde{\omega}_1) = \exp\left(\tilde{\omega}_1 - \tilde{p}_1^*(\tilde{\omega}_1)\right) - \frac{1}{\exp\left(\tilde{\omega}_1 - \tilde{p}_1^*(\tilde{\omega}_1)\right)} - \sum_i \delta v_r \phi_i(\pi^i_{E2} + \pi^i_{I2}) \tag{24}$$

Notice that for any ω_{E1}, Equation (23) implies that, if $\tilde{\omega}_1 = 0$, $\tilde{p_1^*} < 0$. Therefore,

$$\exp\left(-\tilde{p}_1^*(0)\right) - \frac{1}{\exp\left(-\tilde{p}_1^*(0)\right)} > 0 > -\sum_i v_r \delta \phi_i u_2^i.$$

Step 2: $\frac{d\tilde{p}_1^s}{d\tilde{\omega}_1} < \frac{d\tilde{p}_1^*}{d\tilde{\omega}_1}$ iff $\sum_i \delta v_r \phi_i(\pi^i_{I2} - \pi^i_{E2}) < 1$: Since $\frac{\partial \tilde{\omega}_1}{\partial \omega_{I1}} = -1$, we consider a marginal decrease in ω_{I1} here:

$$\frac{d(-\sum_i v_r \delta \phi_i u_2^i)}{d\omega_{I1}} = \sum_i v_r \delta \phi_i q^{i*}_{E2}\left(1 + \frac{d\tilde{p}_2^*}{d\omega_{I2}}\right) > 0 \text{ since } -1 < \frac{d\tilde{p}_2^*}{d\omega_{I2}} < 0.$$

Hence, the LHS of inequality (23) is decreasing as ω_{I1} decreases or $\tilde{\omega}_1$ increases.

$$\frac{d\left(\exp\left(\tilde{\omega}_1 - \tilde{p}_1^*(\tilde{\omega}_1)\right) - \frac{1}{\exp\left(\tilde{\omega}_1 - \tilde{p}_1^*(\tilde{\omega}_1)\right)}\right)}{d\omega_{I1}} = -\left(\exp\left(\tilde{\omega}_1 - \tilde{p}_1^*(\tilde{\omega}_1)\right) + \frac{1}{\exp\left(\tilde{\omega}_1 - \tilde{p}_1^*(\tilde{\omega}_1)\right)}\right)\left(1 + \frac{d\tilde{p}_1}{d\tilde{\omega}_{1I}}\right)$$

As long as $-1 < \frac{d\tilde{p}_1}{d\tilde{\omega}_{1I}}$, the RHS of inequality (23) increases as ω_{I1} decreases or $\tilde{\omega}_1$ increases.

Using the Implicit Function Theorem on Equation (24), we can derive $\frac{d\tilde{p}_1}{d\tilde{\omega}_{1I}}$:

$$\frac{d\tilde{p}_1}{d\tilde{\omega}_{1I}} = -\frac{\exp\left(\tilde{\omega}_1 - \tilde{p}_1(\tilde{\omega}_1)\right) + \frac{1}{\exp(\tilde{\omega}_1 - \tilde{p}_1(\tilde{\omega}_1))} + \sum_i \delta v_r \phi_i(\pi^i_{I2} - \pi^i_{E2})}{\exp\left(\tilde{\omega}_2 - \tilde{p}_1(\tilde{\omega}_1)\right) + \frac{1}{\exp(\tilde{\omega}_1 - \tilde{p}_1(\tilde{\omega}_1))} + 1}$$

Hence, $\frac{d\tilde{p}_1}{d\tilde{\omega}_{1I}} > -1$ iff $\sum_i \delta v_r \phi_i(\pi^i_{I2} - \pi^i_{E2}) < 1$.

A.3 Poisson process of bookings

Let the probability that a single consumer wants to book listing j at time t by σ_{jt}. We show that the expected number of arriving consumers who wish to book the listing is $\mu\sigma_{jt}$. Recall from Equation (4) that the ϵs are Gumbel distributed such that σ_{jt} takes the usual logit form.

$$\sigma_{jt} = \frac{\exp(\nu(p_{jt}, x_{jt}))}{1 + \exp(\nu(p_{jt}, x_{jt})) + \sum_x^X (s_t(x) - \mathbb{1}_{x=x_{jt}}) \exp(\nu(P_t(x), x))} \tag{25}$$

The probability that m of M consumers wish to book the listing is binomial distributed with probability σ_{jt}.

$$\binom{M}{m}\sigma_{jt}^{m}(1-\sigma_{jt})^{M-m} \tag{26}$$

Knowing that M is Poisson distributed with mean μ, we integrate over M.

$$\sum_{M=m}^{\infty}\binom{M}{m}\sigma_{jt}^{m}(1-\sigma_{jt})^{M-m}\frac{\mu^{M}\exp(-\mu)}{M!} \tag{27}$$

$$=\frac{\sigma_{jt}^{m}\exp(-\mu)}{m!}\sum_{M=m}^{\infty}\frac{(1-\sigma_{jt})^{M-m}\mu^{M}}{(M-m)!} \tag{28}$$

$$=\frac{(\sigma_{jt}\mu)^{m}\exp(-\mu)}{m!}\sum_{M-m=0}^{\infty}\frac{((1-\sigma_{jt})\mu)^{M-m}}{(M-m)!} \tag{29}$$

$$=\frac{(\sigma_{jt}\mu)^{m}\exp(-\mu)}{m!}\exp((1-\sigma_{jt})\mu) \tag{30}$$

The last step follows from the definition of the exponential function, $\exp(x) = \sum_{x}^{\infty}\frac{x^{n}}{n!}$. $\frac{(\sigma(x)\mu)^{m}\exp(-\sigma(x)\mu)}{m!}$ is the probability density function of a Poisson distribution with mean $\sigma(x)\mu$.

A.4 Transition matrix

Conditional on $\mathbf{P}_t, \mathbf{s}_t$, and l, transition matrix $\mathbf{T}$ is given by the following matrix.

$(K_t, N_t) \backslash (K_{t+1}, N_{t+1})$	$(0,0)$	$(0,1)$	$(1,1)$	$(0,2)$	$(1,2)$	$(2,2)$	...	$(\bar{N}, \bar{N})$
$(0,0)$	$\rho^0(0,0)$	$\rho^b(0,0)$	$\rho^g(0,0)$	0	0	0	...	0
$(0,1)$	0	$\rho^0(0,1)$	0	$\rho^b(0,1)$	$\rho^g(0,1)$	0	...	0
$(1,1)$	0	0	$\rho^0(1,1)$	0	$\rho^b(1,1)$	$\rho^g(1,1)$	...	0
$(0,2)$	0	0	0	$\rho^0(0,2)$	0	0	...	0
$(1,2)$	0	0	0	0	$\rho^0(1,2)$	0	...	0
$(2,2)$	0	0	0	0	0	$\rho^0(2,2)$	...	0
...	...	...	...	...	...	...	...	
$(\bar{N}, \bar{N})$	0	0	0	0	0	0	...	1

Conditional on $\mathbf{P}_t, \mathbf{s}_t$, and l expanded transition matrix $\mathbf{F}$ is given by the following matrix.

$(K_t, N_t) \backslash (K_{t+1}, N_{t+1})$	$(0,0)$	$(0,1)$	$(1,1)$	$(0,2)$	$(1,2)$	$(2,2)$	...	$(\bar{N}, \bar{N})$	*inactive*
$(0,0)$	$(1-\chi(0,0))\rho^0(0,0)$	$(1-\chi(0,0))\rho^b(0,0)$	$(1-\chi(0)\rho^g(0,0)$	0	0	0	...	0	$\chi(0,0)$
$(0,1)$	0	$(1-\chi(0,1))\rho^0(0,1)$	0	$(1-\chi(0,1))\rho^b(0,1)$	$(1-\chi(0,1))\rho^g(0,1)$	0	...	0	$\chi(0,1)$
$(1,1)$	0	0	$(1-\chi(1,1))\rho^0(1,1)$	0	$(1-\chi(1,1))\rho^b(1,1)$	$(1-\chi(1,1))\rho^g(1,1)$	...	0	$\chi(1,1)$
$(0,2)$	0	0	0	$(1-\chi(0,2))\rho^0(0,2)$	0	0	...	0	$\chi(0,2)$
$(1,2)$	0	0	0	0	$(1-\chi(1,2))\rho^0(1,2)$	0	...	0	$\chi(1,2)$
$(2,2)$	0	0	0	0	0	$(1-\chi(0,2))\rho^0(0,2)$	...	0	$\chi(0,2)$
...	...	...	...	...	...	...	...		...
$(\bar{N}, \bar{N})$	0	0	0	0	0	0	...	$1-\chi(\bar{N}, \bar{N})$	$\chi(\bar{N}, \bar{N})$
inactive	λ	0	0	0	0	0	...	0	$1-\lambda$

A.5 First- and second-order derivatives of the value function

$$
\begin{aligned}
v'(x) =& 28(q_0(P_0(x),x)+q_0'(p_t,x)P_0(x))+(1-\chi_0(P_0(x),x))\delta T_0'(P_0(x),x)V_0(x)\\
v''(x) =& 28(2q_0'(p_t,x)+q_0''(P_0(x),x)P_0(x))+(1-\chi_0(P_0(x),x))\delta T_0''(P_0(x),x)V_0(x)-\chi_0'(P_0(x),x)\delta T_0'(P_0(x),x)
\end{aligned}
\tag{31}
$$

$\chi_0(x)=\exp(-T_0(x)V_0(x)\bar{\phi}_l^{-1})$ is the exit rate in state x and $T_0(x)$ are the transition probabilities.

A.6 Elasticities

Let the probability that a single consumer wants to book a listing j by σ_{jt}.

$$
\sigma(x)=\frac{\exp(\nu(P^*(x),x))}{1+\sum_x^X s^*(x)\exp(\nu(P^*(x),x))} \tag{32}
$$

The own-price elasticity of listing j at time t is given by the following expression.

$$
-100\frac{\exp(-\mu\sigma(x))}{1-\exp(-\mu\sigma(x))}\mu\sigma(x)(1-\sigma(x))\alpha P^*(x) \tag{33}
$$

The semi-elasticity of demand with respect to a good review of listing j at time t is given by the following expression.

$$
\frac{-\exp(-\mu\sigma(K+1,N+1,l))+\exp(-\mu\sigma(K,N,l))}{1-\exp(-\mu\sigma(K,N,l))} \tag{34}
$$

A.7 Expected operating cost

The expected operating cost is given by the following expression.

$$
\begin{aligned}
\mathbb{E}[\phi_l|\phi_l\le\delta T(x)V(x)](1-\chi(x)) =& \bar{\phi}_l-\mathbb{E}[\phi_l|\phi_l>\delta T(x)V(x)]\chi(x)\\
=& \bar{\phi}_l-(\bar{\phi}_l+\delta T(x)V(x))\chi(x)\\
=& (1-\chi(x))\bar{\phi}_l-\delta T(x)V(x)\chi(x)
\end{aligned}
$$

A.8 Expected entry cost

The expected cost of entry is given by the following expression.

$$
\begin{aligned}
\mathbf{E}_\kappa[\kappa|\kappa\le\delta\mathbb{E}_l[V((0,0,l))]] =& \frac{\bar{\kappa}-Pr(\kappa>\delta\mathbb{E}_l[V((0,0,l))]\mathbb{E}_\kappa[\kappa|\kappa>\delta\mathbb{E}_l[V((0,0,l))]]}{Pr(\kappa\le\delta\mathbb{E}_l[V((0,0,l))])}\\
=& \bar{\kappa}-\frac{\exp\left(-\delta\mathbb{E}_l[V((0,0,l))]\bar{\kappa}^{-1}\right)\delta\mathbb{E}_l[V((0,0,l))]}{1-\exp\left(-\delta\mathbb{E}_l[V((0,0,l))]\bar{\kappa}^{-1}\right)}\\
=& \bar{\kappa}\left(1+\frac{v_a-\lambda}{\lambda}\ln\left(1-\frac{\lambda}{v_a}\right)\right)
\end{aligned}
\tag{35}
$$

$$q'p + q = 0 \iff p = -\frac{q}{(1+f)q'}.$$

References

Acemoglu, Daron, Ali Makhdoumi, Azarakhsh Malekian, and Asuman Ozdaglar, "Learning from reviews: The selection effect and the speed of learning," *Econometrica*, 2022, *90* (6), 2857–2899. 5

Almagro, Milena and Tomás Domınguez-Iino, "Location sorting and endogenous amenities: Evidence from Amsterdam," *Working Paper*, 2022. 6

Anderson, Michael and Jeremy Magruder, "Learning from the crowd: Regression discontinuity estimates of the effects of an online review database," *The Economic Journal*, 2012, *122* (563), 957–989. 6

Avery, Christopher, Paul Resnick, and Richard Zeckhauser, "The market for evaluations," *American Economic Review*, 1999, *89* (3), 564–584. 5

Banerjee, Abhijit V, "A simple model of herd behavior," *The Quarterly Journal of Economics*, 1992, *107* (3), 797–817. 5

Bergemann, Dirk and Juuso Välimäki, "Learning and strategic pricing," *Econometrica*, 1996, *64* (5), 1125–1149. 5

_ and _ , "Market diffusion with two-sided learning," *The RAND Journal of Economics*, 1997, pp. 773–795. 5

_ and _ , "Experimentation in markets," *The Review of Economic Studies*, 2000, *67* (2), 213–234. 5

Besanko, David, Ulrich Doraszelski, and Yaroslav Kryukov, "The economics of predation: What drives pricing when there is learning-by-doing?," *American Economic Review*, 2014, *104* (3), 868–97. 6

_ , _ , and _ , "How efficient is dynamic competition? the case of price as investment," *American Economic Review*, 2019, *109* (9), 3339–64. 6

_ , _ , _ , and Mark Satterthwaite, "Learning-by-doing, organizational forgetting, and industry dynamics," *Econometrica*, 2010, *78* (2), 453–508. 6

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026. 5

Bolton, Patrick and Christopher Harris, "Strategic experimentation," *Econometrica*, 1999, *67* (2), 349–374. 5

Cabral, Luis and Ali Hortacsu, "The dynamics of seller reputation: Evidence from eBay," *The Journal of Industrial Economics*, 2010, *58* (1), 54–78. 6

_ and Michael H Riordan, "The learning curve, market dominance, and predatory pricing," *Econometrica*, 1994, *62* (5), 1115–1140. 2

_ and _ , "The learning curve, predation, antitrust, and welfare," *The Journal of Industrial Economics*, 1997, *45* (2), 155–169. 2

Calder-Wang, Sophie, "The distributional impact of the sharing economy on the housing market," *Working Paper*, 2021. 6, 40

Che, Yeon-Koo and Johannes Hörner, "Recommender Systems as Mechanisms for Social Learning," *The Quarterly Journal of Economics*, 2018, *133* (2), 871–925. 2, 4, 5

Chevalier, Judith A and Dina Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, 2006, *43* (3), 345–354. 6

Decker, Christian, "What's in an Airbnb Five-Star Rating? An Empirical Model of Bayesian Persuasion," *Working Paper*, 2023. 6

Dickstein, Michael J, "Efficient provision of experience goods: Evidence from antidepressant choice," *Working Paper*, 2018. 21

Doraszelski, Ulrich and Kenneth L Judd, "Avoiding the curse of dimensionality in dynamic stochastic games," *Quantitative Economics*, 2012, *3* (1), 53–93. 12

_ and _ , "Dynamic stochastic games with random moves," *Quantitative Marketing and Economics*, 2019, *17* (1), 59–79. 12

_ and Mark Satterthwaite, "Computable Markov-perfect industry dynamics," *The RAND Journal of Economics*, 2010, *41* (2), 215–243. 26

Ericson, Richard and Ariel Pakes, "Markov-perfect industry dynamics: A framework for empirical work," *The Review of Economic Studies*, 1995, *62* (1), 53–82. 12

Fang, Limin, "The effects of online review platforms on restaurant revenue, consumer learning, and welfare," *Management Science*, 2022, *68* (11), 8116–8143. 6

Farronato, Chiara and Andrey Fradkin, "The Welfare Effects of Peer Entry: The Case of Airbnb and the Accommodation Industry," *American Economic Review*, 2022, *112* (6), 1782–1817. 6, 19, 24

Ferrari, Silvia and Francisco Cribari-Neto, "Beta regression for modelling rates and proportions," *Journal of Applied Statistics*, 2004, *31* (7), 799–815. 21

Fershtman, Chaim and Ariel Pakes, "Dynamic games with asymmetric information: A framework for empirical work," *The Quarterly Journal of Economics*, 2012, *127* (4), 1611–1661. 12

Fradkin, Andrey and David Holtz, "Do incentives to review help the market? Evidence from a field experiment on Airbnb," *American Economic Review: Insights*, 2022. 6, 41

_ , Elena Grewal, and David Holtz, "Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on Airbnb," *Marketing Science*, 2021, *40* (6), 1013–1029. 6, 10

Huang, Yufeng, "Pricing Frictions and Platform Remedies: The Case of Airbnb," *Available at SSRN 3767103*, 2022. 6, 24

Jovanovic, Boyan, "Selection and the Evolution of Industry," *Econometrica*, 1982, *50* (3), 649–670. 19

Keller, Godfrey, Sven Rady, and Martin Cripps, "Strategic experimentation with exponential bandits," *Econometrica*, 2005, *73* (1), 39–68. 5

Kremer, Ilan, Yishay Mansour, and Motty Perry, "Implementing the "wisdom of the crowd"," *Journal of Political Economy*, 2014, *122* (5), 988–1012. 2, 4, 5

Lewis, Gregory and Georgios Zervas, "The welfare impact of consumer reviews: A case study of the hotel industry," *Working Paper*, 2016. 6

Luca, Michael, "Reviews, reputation, and revenue: The case of Yelp. com," *Working Paper*, 2016, (12-016). 6

Mankiw, N Gregory and Michael D Whinston, "Free entry and social inefficiency," *The RAND Journal of Economics*, 1986, pp. 48–58. 36

McFadden, Daniel, "Computing willingness–to–pay in random utility models," in "Trade, theory and econometrics" 2012, pp. 275–296. 33

Pallais, Amanda, "Inefficient hiring in entry-level labor markets," *American Economic Review*, 2014, *104* (11), 3565–3599. 5

Papanastasiou, Yiangos, Kostas Bimpikis, and Nicos Savva, "Crowdsourcing exploration," *Management Science*, 2018, *64* (4), 1727–1746. 5

Proserpio, Davide, Wendy Xu, and Georgios Zervas, "You get what you give: theory and evidence of reciprocity in the sharing economy," *Quantitative Marketing and Economics*, 2018, *16*, 371–407. 6

Reimers, Imke and Joel Waldfogel, "Digitization and pre-purchase information: the causal and welfare impacts of reviews and crowd ratings," *American Economic Review*, 2021, *111* (6), 1944–71. 1, 6

Small, Kenneth A and Harvey S Rosen, "Applied welfare economics with discrete choice models," *Econometrica*, 1981, pp. 105–130. 32, 33

Smith, Lones, Peter Norman Sørensen, and Jianrong Tian, "Informational herding, optimal experimentation, and contrarianism," *The Review of Economic Studies*, 2021, *88* (5), 2527–2554. 5

Vellodi, Nikhil, "Ratings design and barriers to entry," *Working Paper*, 2022. 4, 5

Weintraub, Gabriel Y, C Lanier Benkard, and Benjamin Van Roy, "Markov perfect industry dynamics with many firms," *Econometrica*, 2008, *76* (6), 1375–1411. 3, 6, 12, 26

_ , _ , and _ , "Computational methods for oblivious equilibrium," *Operations Research*, 2010, *58* (4-part-2), 1247–1265. 6

Williams, Huw CWL, "On the formation of travel demand models and economic evaluation measures of user benefit," *Environment and Planning A*, 1977, *9* (3), 285–344. 32

Wu, Chunhua, Hai Che, Tat Y Chan, and Xianghua Lu, "The economic value of online reviews," *Marketing Science*, 2015, *34* (5), 739–754. 6

Zervas, Georgios, Davide Proserpio, and John W Byers, "A first look at online reputation on Airbnb, where every stay is above average," *Marketing Letters*, 2021, *32*, 1–16. 6, 16