Quality Externalities on Platforms: The Case of Airbnb*

Sonia Jaffe† Peter Coles‡ Steven Levitt§ Igor Popov¶

November 14, 2017

Click here for the most up-to-date version

Preliminary and Incomplete

Please do not cite or circulate without authors' permission.

Abstract

In order to screen out low-quality sellers or incentivize high quality, platforms need a good measure of seller quality. Having high-quality sellers is particularly important if buyers on a platform have limited information about the sellers; buyers' learning means that the quality of a seller in any given transaction affects whether or how frequently that buyer returns to the platform. A seller's influence on the number of times her buyers return to the platform is an externality that the seller exerts on the platform's other sellers; we propose using this externality to measure seller quality. Using data from Airbnb, a peer-to-peer accommodation platform, we calculate each listing's *guest return propensity* (GRP), defined as the average number of subsequent bookings a listing's guests complete, controlling for guest and trip characteristics. There is substantial variation in GRP across listings and its correlation with a listing's average rating is only 0.05. Using an instrumental variable analysis to account for unobservable guest characteristics, we find that our measure of GRP has a causal effect on returns: a one standard deviation increase in GRP causes guests to take an additional 0.34 trips (a 17% increase). We discuss how platforms can increase overall seller surplus by directing buyers towards higher quality sellers, either by using Pigouvian subsidies for quality or by prioritizing high-quality sellers in their search algorithms.

*The opinions and views expressed herein are those of the authors, not Airbnb. The authors thank Kevin Murphy for helpful comments and Ryan Parks and Anya Marchenko for research assistance.

†University of Chicago. Corresponding author: spj@uchicago.edu

‡Airbnb

§University of Chicago

¶Airbnb

Platforms live or die by their ability to match users and align incentives. Online marketplaces such as Lyft, eBay, and Airbnb do not directly control production of their goods or services, relying instead on market design decisions to manage their marketplaces. These online platforms, along with more traditional 'platforms' such as newspapers and farmers' markets, have inspired a robust economics literature on platform externalities, which focuses almost exclusively on quantity-based externalities (e.g., eBay buyers benefit when new sellers join the market and are potentially harmed by additional buyers). Less academic attention has been paid to externalities that arise from network *quality*.

In contrast to the academic literature, online marketplaces invest heavily in managing and incentivizing high quality for the goods and services provided through their platform. In doing so, they face a key challenge: quality is notoriously difficult to measure. Platforms can often measure purchase propensity. User-generated ratings vary in their informativeness (whether the ratings actually differ across agents) and in their effectiveness, (whether users base their decisions on the ratings). But even at their best, there is no clear mapping from these metrics to the value a seller contributes to the platform.

In this paper, we explore the 'quality externality' that platform participants exert on other users on the same side of the market and propose that platforms should measure quality based on this externality. A marketplace participant that provides a good (bad) experience makes it more (less) likely that the agent on the other side of the transaction returns to the platform. These future transactions help other participants on the same side of the market, but are not internalized by each agent.[1] A platform can estimate the effect of each seller's behavior on the propensity of their buyers to continue interacting with the platform. We call this metric the *guest return propensity* (GRP).

We begin by presenting a theoretical framework that motivates the use of GRP and guides the empirical analysis. On the buyers' side, users update their expectations about average platform quality with each purchase. We extend the basic learning model to incorporate learning about the relationship between price and quality. The key insight is that this buyer-side learning can generate a wedge between individual seller incentives and those of the platform or sellers as a group. It is likely that a buyer's subsequent purchases on the platform are from different sellers, so the initial seller does not internalize that benefit. From the platform's point of view, this setup mirrors a standard case for a Pigouvian tax, highlighting the potential efficiency gains from aligning social and private incentives across agents.

[1] The externality seems particularly important in a platform such as Groupon, where the business model is based on buyers trying *new* sellers. If bad experiences cause buyers to only purchase Groupons for sellers that they already frequent, then Groupons are less profitable for sellers.

We then present the empirical context. We describe our data from Airbnb, which coordinates millions of trips in 191 countries. We observe rich information about search and travel patterns on the platform over time. Using bookings data, we construct a GRP metric for each accommodation listed on Airbnb. We find significant variation in GRP across listings, even controlling for guest and trip characteristics. Very highly reviewed listings differ in their guests' propensity to rebook. In fact, GRP is only weakly correlated with guests' rating of that listing or with the number of guests the listing has.[2] Moreover, the GRP measure is persistent, predicting rebooking rates out of sample.

Our learning model predicts that GRP will have a stronger effect on the beliefs of new users, who have weaker priors regarding platform quality. Consistent with this, we find that GRP has a larger effect on the probability of return for inexperienced guests than for experienced ones. However, GRP's effect on total future guest spending is larger for experienced guests. This difference reflects the fact that experienced guests are more likely to be frequent travelers so, though the effect is smaller, it influences more potential future trips and therefore more future spending on the platform. Thus, from the platforms point of view, experienced guests are attractive targets for high-GRP listings, even if their beliefs are less affected by GRP.

Using granular search data, we design an instrumental variable strategy to show that the effect of a listing's GRP on guests' subsequent trips is causal rather than just reflecting unobserved guest characteristics. Frequent bookings, changes in availability, and ongoing search experiments often result in two users visiting the site on the same day entering the same search terms being shown different inventory. The size of our sample allow us to leverage these quickly changing choice sets for identification. We instrument for the GRP of the booked listing with the average GRP across the first page of search results, using only variation generated by multiple guests searching for the same market for the same travel day, with comparable trip lead time.

We find that high-GRP listings cause a significant increase in guests' future platform utilization. In our instrumental-variable specification, (exogenously) booking a listing with a one standard deviation higher GRP leads a guest to take 0.34 additional future trips. This effect is three times larger than the comparable effect of a guest booking a listing with a one standard deviation higher rating. In contrast to the effect on future travel, displaying high-GRP listings in a guest's search results does not lead to many more immediate bookings. A one standard deviation increase in the average GRP displayed only raises

[2]The lack of correlation with ratings may seem surprising, but makes sense if we think that ratings are based on things guests think are specific to a given listing, whereas whether a guest returns is not effected by things they think are listing-specific, but precisely by the things they think generalize to all listings. The low correlation with the number of guests is consistent with the idea that GRP is not observable ex-ante.

purchase probabilities by 0.6% (0.00032 percentage points). We take this as evidence that GRP reflects the quality of the actual experience and not necessarily some feature of the listing that users observe while searching.

We end by discussing the implications of our results for Airbnb and how they might generalize to other platforms. Airbnb could use our measure of listing quality to raise overall seller surplus by removing low quality hosts from the platform, or by directly trying to help low-quality hosts improve. Alternatively, it could incentivize higher listing quality and redirect guests to higher-quality listings either by implementing a Pigouvian tax on low quality or by showing guests higher quality listings in their search results.

To our knowledge, Nosko and Tadelis (2015) were the first to highlight the role of quality-driven own-side platform externalities. They argue that reputational externalities arise when buyers update their beliefs about the quality of all sellers on the platform from individual interactions. In the context of the eBay platform, they propose a quality metric that takes into account the propensity of buyers to leave feedback. They then use whether a buyer returns to eBay to validate this measure of quality. We build upon their work in two key ways. First, we provide a formal theoretical treatment of buyer learning and the implications for a platform's surplus. Second, we show that reputational externalities can be directly converted into a useful quality metric. In this way, we offer a new, complementary quality measure that is directly informed by reputational externalities.

More broadly, we contribute to the growing literature exploring two-sided market design and platform externalities (e.g., Rochet and Tirole, 2003, 2006; Evans, 2003; Armstrong, 2006; Hagiu, 2007; Rysman, 2009; Weyl, 2010; Bardey et al., 2009) This body of work studies platform design and pricing problems and platform competition in a setting where users exert cross-side membership or usage externalities. We explore the existence and implications of quality externalities in the same type of setting.

In focusing on users' learning about platform quality from interactions with individual agents, we build upon the collective reputations and reputational commons literatures. Shapiro (1982) models consumer learning and firm quality investments in equilibrium. Taking these concepts to an empirical setting, Landon and Smith (1998) estimate the effects of individual and group reputational effects in the market for Bordeaux wines. King et al. (2002) note that if individual firm actions can affect group reputation, a special case of the commons problem exists (see also, Winfree and McCluskey, 2005; Barnett, 2007). We apply these principles to the platform setting to understand how intermediaries can mitigate these commons problems.

Finally, this paper relates to the literature exploring quality and ratings metrics used by online marketplaces (e.g., Dellarocas, 2003; Cabral and Hortacsu, 2010; Mayzlin et al., 2014;

Fradkin et al., 2015; Luca, 2017). Our contribution to this literature is the introduction of a new metric that is platform-independent and grounded in the theory of quality investment and externalities.

The remainder of the paper is organized as follows. Section 1 outlines the theory, discussing both consumer learning and the externality it generates. Section 2 describes the Airbnb context and data we use in the analysis. Here, we show how we construct the GRP metric for each listing. Section 3 presents the empirical results, both the reduced-form analysis and the instrumental-variable approach using the search data. In Section 4, we discuss potential strategies the platform could use to raise the average GRP of the listings that guests visit. Section 5 concludes.

1 Theory

We start with a basic model of consumers' behavior and learning. We then consider the future value of information and the possibility that consumers observe a price or other signal prior to purchase. We then look at the platform objective function and how it compares to the payoffs of individual sellers.

1.1 Individual Choices

The central aspects to our model of user choice are not specific to platforms: there is a good whose quality is not known and varies across units; an individual may have multiple opportunities to consume the good and each time she does she updates her prior on the distribution of quality.

Consumers, indexed by i, first arrive in the market with a prior over the distribution of the good's quality, f_0. Each period, with probability π_i, a consumer has an opportunity to potentially consume the good; a consumer with the opportunity to consume gets a draw, $\theta_t \sim G$, which is her value for a good of zero quality (a normalization). In addition, consumers get utility $u(q)$ if they purchase the good and its quality is $q \gtrless 0$, for some concave function $u(\cdot)$.

Myopic Individuals

Start with the case of consumers who are myopic and base their purchase decision only on the current period's utility; they will purchase the good if

$$\theta_t + E[u(q_t)] > 0.$$

This implies a threshold that depends on the beliefs that period, where the consumer purchases if and only if $\theta > \underline{\theta}(\cdot)$, where $\underline{\theta}$ is a function of individual's beliefs about the distribution of quality. If a consumer purchases the good, she observes a quality draw q_t and updates her prior accordingly.

For simplicity, assume that the prior is normal.[3] In the most basic model, consumers may know (or think they know) the variance of quality in the market, but be uncertain about the mean. In this context, their prior can be described by three parameters: the variance in quality across sellers, $1/\gamma$, the individual's belief about the average quality, μ_0, and the confidence they have in that belief, τ_0.

The first time a consumer has an opportunity to purchase the good, the cutoff will be

$$\underline{\theta} = -E\left[u(q_1)|q_1 \sim N\left(\mu_0, \frac{1}{\tau_0} + \frac{1}{\gamma}\right)\right].$$

If she chooses to purchase, she will get a draw q_1 and update her prior to

$$\mu_1 = \frac{\tau\mu_0 + \gamma q_1}{\tau_0 + \gamma}$$
$$\tau_1 = \tau_0 + \gamma.$$

The less variance there is seller quality, the more weight the individual puts on a quality draw.

Lemma 1. *The more confidence consumers have in their beliefs (higher τ) or the more variation there is across sellers in quality (lower γ), the less an individual's beliefs will be influenced by a given quality draw.*

Lemma 1 suggests that quality externalities may be smaller on a platform like Etsy[4] – where the artistic nature of the products may lead consumers to think that sellers differ a lot – then a platform like Uber where all the "sellers" are offering a fairly homogeneous product – "a ride" – so consumers may have less reason to expect big differences across sellers. Similarly, if consumers have stronger priors because a platform has been around longer or is more widely known, a single purchase will affect their beliefs less.

In addition to not knowing the mean level of quality across sellers, individuals may be uncertain about its variance. In this case, their prior consists of four components: a belief about the mean and variance of the quality distribution – μ_0, σ_0 – and the precision of those beliefs – τ_0 and d_0, respectively. Each time a consumer observes a quality draw, q_k, she

[3]Without the assumption that $u(\cdot)$ is concave, the assumption of a normal prior would just be a normalization; since quality has no inherent unit, it could be re-scaled to be normally distributed. The assumptions of concavity and normality together are substantive, but they allow us to get interesting comparative statics with closed-form results.

[4]Etsy is a platform for selling primarily handmade or vintage items and craft supplies.

updates her beliefs to a posterior

$$\mu_k = \frac{\mu_{k-1}\tau_{k-1} + q_k}{\tau_{k-1} + 1}$$
$$\tau_k = \tau_{k-1} + 1$$
$$\sigma_k = \frac{d_{k-1}\sigma_{k-1} + \frac{\tau_{k-1}}{\tau_{k-1}+1}(q_k - \mu_{k-1})^2}{d_{k-1} + 1}$$
$$d_k = d_k + 1$$

where k indexes only those periods in which the individual purchased the good. An individual's purchase threshold depends only on the mean and variance, $\underline{\theta}(f_k) = \underline{\theta}(\mu_k, \sigma_k)$. Later posteriors are less effected by the quality draw.

Lemma 2. *When a consumer has more past quality draws, that is $k > k'$, then*

1. *The mean of the consumer's posterior is less effected by an additional draw: That is $\frac{\partial \mu_k}{\partial q} < \frac{\partial \mu_{k'}}{\partial q}$.*
2. *If the prior mean and variance are the same, and the consumer's threshold cutoff $\underline{\theta}$ is decreasing in the quality draw,[5] then the cutoff will be less sensitive to the quality draw when the consumer has more past quality draws. That is if $\mu_k = \mu_{k'}$ and $\sigma_k = \sigma_{k'}$ then*

$$0 > \frac{\partial \underline{\theta}(\mu_k, \sigma_k)}{\partial q} > \frac{\partial \underline{\theta}(\mu_{k'}, \sigma_{k'})}{\partial q}.$$

When $q < \mu$, increasing the quality both increases the mean of the posterior and decreases its variance; these two changes both push $\underline{\theta}$ down, but less so when a consumer is more experienced. If $q > \mu$, then increasing the quality increases the variance of the posterior, which has the opposite effect on $\underline{\theta}$ as increasing the mean. We cannot sign the derivative $\frac{\partial \underline{\theta}}{\partial q}$ and so the sign of the derivative is uncertain and sensitive (smaller absolute value of the derivative) to a quality draw when she has more previous quality draws.

Though a given individual's beliefs will be more effected by observed quality when she has fewer past draws, individuals who have many past purchases are not randomly selected from the population; they are more likely to be individuals who get lots of opportunities to purchase – those with a high π_i. The effect on an individual's probability of purchase – as distinct from beliefs – is

$$\pi_i g(\underline{\theta}) \frac{\partial \underline{\theta}}{\partial q}.$$

[5] While this is the case in general, at very high levels of quality it is the increase in posterior variance from raising quality can push the threshold up more than the increase in the posterior mean pushes it down.

If we only observe one period's decision, then the expectation of an individual's type, $E[\pi_i]$ is higher if they purchase than if they do not.[6] In general, we may expect more experienced travelers to be positively selected on π. Therefore, the effect of quality on their behavior could be larger than for inexperienced consumers, even if the effect on their beliefs is smaller.

Forward-looking consumers

Forward looking consumers have the same process for learning about quality, but a different decision rule. With a prior f_k, the utility of purchase relative to not purchase is

$$W(\theta|f_k) = \theta + E[u(q)] + \delta \left(E_q \left[\tilde{W}(f_{k+1}(f_k, q)) \right] - \tilde{W}(f_k) \right),$$

where $\tilde{W}(f_k) = \frac{\pi_i}{1-(1-\pi_i)\delta} E_\theta[\max\{0, W(\theta|f_k)\}]$ and $f_{k+1}(f_k, q)$ is the prior updated after observing q. This is similar to a single-armed bandit problem, though the (known) value of the outside option changes from period-to-period.[7] Non-myopic consumers will have a lower threshold $\underline{\theta}$ for purchasing because they recognize that purchasing the good this period provides information value for future periods. Overtime, as the precision of f_k increases, that information value will decrease and forward looking consumers will behave more like myopic consumers.

Prices

In many markets goods vary not only in quality, but also in price. Allowing goods to vary by price means individuals may learn about the distribution of price and, more importantly, its correlation with quality.[8] In a contrived setting where a price-quality pair (q, p) is randomly drawn, updating can follow a simple rule. If the beliefs at the beginning of period t are means $\mu_t = (\mu_t^q, \mu_t^p)$, variance-covariance $\sigma_t^2 = \begin{pmatrix} \sigma_{q,t}^2 & \sigma_{qp,t}^2 \\ \sigma_{qp,t}^2 & \sigma_{p,t}^2 \end{pmatrix}$, and confidence

[6]If one observes a sequence of purchase choices, it is actually possible that the individual with more purchases is likelier to have a lower π_i. If the true mean of the distribution is above the prior, then once an individual purchases once, we expect her to purchase more, because on average the posterior beliefs will indicate better quality than the prior. If an individual purchases in period 1 and then does not purchase for many periods, then we think the individual must have a low π_i (and the first period was fluke) because otherwise she would have returned in a subsequent period; whereas if an individual does not purchase in the first period, her lack of subsequent purchasing can be explained by the low prior, so she may still have a fairly high π_i.

[7]This implies that unlike the simple single-armed bandit, a user may chose not to purchase in one period, but then chose to purchase in subsequent periods.

[8]It may also mean that some purchases may be more valuable if they transact at a higher price, which we discuss in Section 1.2.

τ_t, and a consumer draws $(q,p)_{t+1}$, the consumer's posterior beliefs will be

$$\mu_{t+1} = \frac{\tau_t\mu_t + (q,y)_{t+1}}{\tau_t+1}, \qquad \tau_{t+1} = \tau_t + 1,$$
$$\sigma^2_{t+1} = \frac{\tau_t}{\tau_t+1}\left(\sigma^2_t + \frac{1}{\tau_t+1}\left((q,y)_{t+1}-\mu_t\right)\left((q,y)_{t+1}-\mu_t\right)^T\right).$$

As one would expect, the believed correlation increases whenever $(q-\mu^q_{t+1})(p-\mu^p_{t+1})>0$.

If, more realistically, a consumer chooses from among a set of goods, after observing their prices, then the consumer is not learning about the distribution of price, but only about the distribution of quality and its correlation with price. In this case, there is no conjugate prior for the beliefs about the correlation, so rather than look at the updating process, we focus on the effect of a consumer's belief about the correlation between prices and quality on her decision of whether to purchase and what price-level of the good to choose.

For notational convenience, let $v(\mu,\sigma) = E\left[u(q)|q\sim N(\mu,\sigma)\right]$ be the expected value given the beliefs about the mean and variance of quality. Because $u(\cdot)$ is concave, $v(\cdot)$ is increasing in the mean and decreasing in the variance and $\frac{\partial^2 v}{\partial\mu^2}<0$, $\frac{\partial^2 v}{\partial\mu\partial\sigma}>0$. For a given correlation ρ and price p, the distribution of quality is $q\sim N\left(\mu^q_t + (p_j-\mu^p_t)\rho, (1-\rho^2)\sigma^2_q\right)$. Increasing the correlation has two effects for the consumer. First, it either increases or decreases the expected quality, depending on whether p is above or below the mean price. Second, it decreases the variance of the distribution of quality conditional on price, which is always good.

If the consumer's overall utility is $v(\mu,\sigma)-p$ (quasilinear in price), then her optimal price satisfies

$$\frac{\partial v}{\partial\mu}\rho = 1.$$

When the correlation changes, we get

$$\frac{\partial p^*}{\partial\rho} = \frac{\frac{\partial v}{\partial\mu}}{-\frac{\partial^2 v}{\partial\mu^2}\rho^2} + \frac{(p-\mu^p)}{\rho} - 2\sigma\frac{\frac{\partial^2 v}{\partial\mu\partial\sigma}}{-\frac{\partial^2 v}{\partial\mu^2}} \tag{1}$$

The first effect is a direct effect of higher correlation pushing for a higher p because an increase in p brings a larger corresponding increase in (expected) q. However, if the optimal p was already above the mean, that the increase in expected quality (for a fixed p) decreases the marginal return to expected quality, so if the correlation is positive, the second term is negative. Lastly, higher correlation implies a lower variance in q conditional on p; this decrease in variance means choosing a lower price involves less left-tail risk, which also pushes for a lower optimal price. In our empirical analysis we look at whether the effect of quality differs by the price of the listing.

A similar analysis applies for other characteristics that the consumer may observe prior to purchase, and can therefore base their decision on. Consumers could also update their priors on the correlation between quality and ratings or any other ex-ante observable characteristic. The quasi-linearity we used for prices, would be less reasonable, but there would still be competing effects. If rating and quality are more highly correlated, it means a consumer gets more quality for a given increase in rating, which pushes for a higher rating; it also means that a given high rating predicts higher quality, so if there are decreasing returns to quality, a consumer may choose a lower quality.

1.2 Externality

Consumers learning about quality means that higher quality this period results in more purchases in later periods. What is special to the platform context (since learning about quality also applies to a single seller) is that an individual seller will only receive a fraction of the returning consumer's later purchases, and does not care about the additional purchases that other sellers receive. Therefore, there is a *quality externality*: the private benefit to a seller of having higher quality is lower than the social benefit.[9]

If there is no variation in price so both the sellers and the platform are just trying to maximize purchases, then the platform's payoff from consumer i is

$$W = \pi_i \sum_{t=1}^{\infty} \delta^t G\left(\underline{\theta}\left(q_1 \ldots q_{k(t)} | \mu_0, \sigma_0, \tau_0, d_0\right)\right),$$

where δ is the discount factor and $k(t)$ is the number of purchases a consumer has made at the start of time t. Each seller j gets some share, d_j of sales and a corresponding share d_j of the payoff

$$\omega_j = d_j \pi_i \sum_{t=1}^{\infty} \delta^t G\left(\underline{\theta}\left(q_1 \ldots q_{k(t)} | \mu_0, \sigma_0, \tau_0, d_0\right)\right) = d_j W.$$

A seller's quality affects the consumer's cutoff in period t directly, but can also have an indirect effect via the other purchase decision the consumer makes in the intervening periods. If a good seller in period 1 causes the consumer to purchase (with higher probability) in period 2, then in period 3, the consumer's prior is based on the quality observed in period 1 and 2 instead of just the quality observed in period 1. Nevertheless, the seller's incentive to

[9]This differs somewhat from the "reputation commons" problem because even if buyers can differentiate among sellers, they still update their beliefs about the distribution of quality among other sellers. Bad quality will make them less likely to purchase even if they know a specific seller was not the one who had low quality in the previous period. As long as the sellers are not available every period (or at some of the locations the consumer goes to purchase), then good quality will also benefit other sellers.

improve quality is also d_j of the social value.

Proposition 1. *The value to the seller of improving quality in period 1 is*

$$\frac{\partial \omega_j}{\partial q_j} = d_j \pi_i \sum_{t=1}^{\infty} \delta^t g^i\left(\underline{\theta}_t\right) \frac{d\underline{\theta}_t}{dq_1} d_j$$

where $g = G'$ and $\underline{\theta}_t = \underline{\theta}_t\left(q_1 \ldots q_{k(t)} | \mu_0, \sigma_0, \tau_0, d_0\right)$. The value to the seller is the social value times the seller's share

$$\frac{\partial \omega_j}{\partial q_j} = d_j \frac{\partial W}{\partial q_j}.$$

The share d_j enters the seller's value twice – once for the probability that the consumer buys from that seller in period 1 and once for the probability that the buyer returns to that seller. The first instance of d_j also enters the platform's marginal value of quality, since the seller's quality is only relevant if the buyer purchases from that seller; the latter instance of d_j does not enter the platform's value because they care whether the guest returns, even if it is to a different seller. The larger the seller's share, the less misaligned her incentives are, but the larger the effect on welfare of a decrease in her quality.

If either δ is small or quality effects the cutoff primarily in the periods directly subsequent to purchase, then the value to the platform of a seller's quality (relative to zero quality) for a consumer who purchases in period 1 is just

$$W(q_j) - W(0) = \delta\left(G(\underline{\theta}_2(q_j|\cdot)) - G(\underline{\theta}(0|\cdot))\right).$$

At the other extreme, if $\delta \to 1$, the value to the platform of a seller's quality is

$$W(q_j) - W(0) = \sum_t \left(G(\underline{\theta}_t(q_j|\cdot)) - G(\underline{\theta}_t(0|\cdot))\right).$$

We use this measure of quality in our empirical analysis. We measure a seller's quality as the number of purchases a consumer makes after purchasing from them, controlling for the number of purchases predicted for that consumer.[10]

Prices

If goods also vary in price then the platform may care about revenue instead of the quantity of transactions. A seller's quality may affect not just whether a consumer purchases, but which product-price she chooses. However, that can all be summarized in the seller's

[10]For computational convenience, we do not discount future trips by how far in the future they are. All the trips are within a few years and three quarters of those who return do so within 6 months, so while $\delta = 1$ is an approximation, a high δ is reasonable.

effect on total spending. In the empirical analysis we look at the effect of a listing's quality on subsequent spending by a guest on the platform:

$$W(q_j) - W(0) = E\left[\sum_t p_{j(t)} | q_1 = q_j\right] - E\left[\sum_t p_{j(t)} | q_1 = 0\right].$$

where if the consumer does not purchase in period t then $j(t) = 0$ and $p_0 = 0$.

2 Data

To get a sense of the potential magnitude and variance in quality externalities, we use data from Airbnb, a global peer-to-peer accommodations platform. Airbnb matches travelers looking for accommodations to hosts who offer their home, apartment, extra bedroom, or other accommodation to guests. Hosts can list their space, set their pricing and availability, and accept bookings. Guests search for where they wish to stay and rate the home and experience after traveling.

The Airbnb platform is large, providing ample data; there are over four million active Airbnb listings in 191 countries. Figure 3 shows the geographical distribution of Airbnb listings. Airbnb data also allow us to follow agents in the network, and their platform behavior, over time. This information is essential for understanding return propensities and the effect that quality may have on decision making. The other advantage of this context is that accommodation quality is heterogeneous and not well-observed ex-ante. The market has the opportunity to incorporate quality information into its marketplace design, making this a relevant setting for platform policy with respect to quality.

While limiting to a subset of locations might give us a more homogeneous set of trips, we want to be able condition on a guest's travel history at Airbnb, so we use trips in all locations. Similarly, though we have many trips, the number of listings with a large number of trips is smaller, so in order to have more precision in estimating listing effects, we use all guests.

Throughout our analysis, we use 'trip' to refer to a check-in by a guest at a listing; guests might book multiple such 'trips' for a single episode of travel. The summary statistics are in Table 1. Markets (which can include multiple cities) have had an average of 33.5K trips and 7.5K listings that have hosted a trip. The average rating of listings as of January 1, 2016 was 4.55. For users who have taken a trip, the average number of trips taken is 2.4; about 55% are female and the average age is 36. For trips since 2011, the average cost is $96 per night. On average trips are booked just over a month in advance.

We are particularly interested in whether guests return to Airbnb after taking a trip. Figure 2 shows the fraction of guests that return by trip number, for trips that occurred at

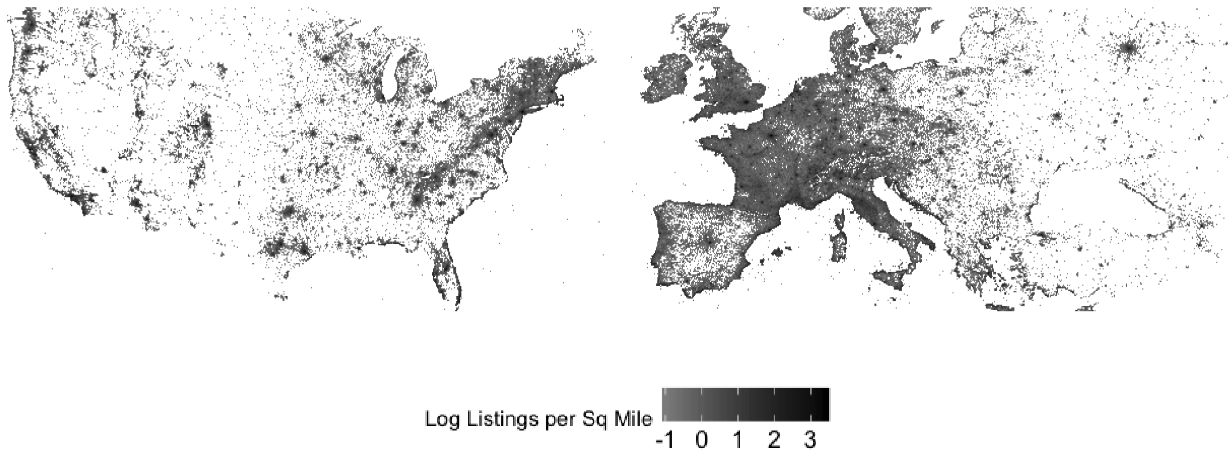


Figure 1: Log density of listings per square mile

Note: These maps show the density of Airbnb listings in the United States and Europe. Listings per square mile is calculated for each cell of .1 deg longitude by .1 deg of latitude, using the approximation of 53 miles per degree latitude.

Table 1: Summary of users, listings, trips, and markets

Statistic	Mean	St. Dev.	Min	Median	Max
Markets					
Listings	1,900	7,137	1	303.5	165,151
Trips	33,578	137,987	1	3,769	2,998,067
Listings					
Rating	4.55	0.58	1.00	4.71	5.00
Trips	18	35	1	5	1,524
Guests					
Age	35.82	12.30	18	32	99
Female (0/1)	0.55	0.50	0	1	1
Trips	2.36	2.93	1	1	475
Trips					
Num of Guests	2.53	1.70	1	2	104
Days in advance booked	33.88	43.01	0	18	1,097
Nightly Price	96	94	0	73	32,216

Note: This table summarizes data about the listings, guests, trips and markets. All variables refer only to listings and guests with at least 1 trip. The rating is the average rating as of January 1, 2016.

least 400 days before the end of the data. About a third of first time guests return within 400 days; that fraction climbs to over 95% for guests with at least 25 trips.

For part of our analysis, we use data on searches made on Airbnb's website. There are many more searches than bookings or trips. We drop any searches the platform thinks were

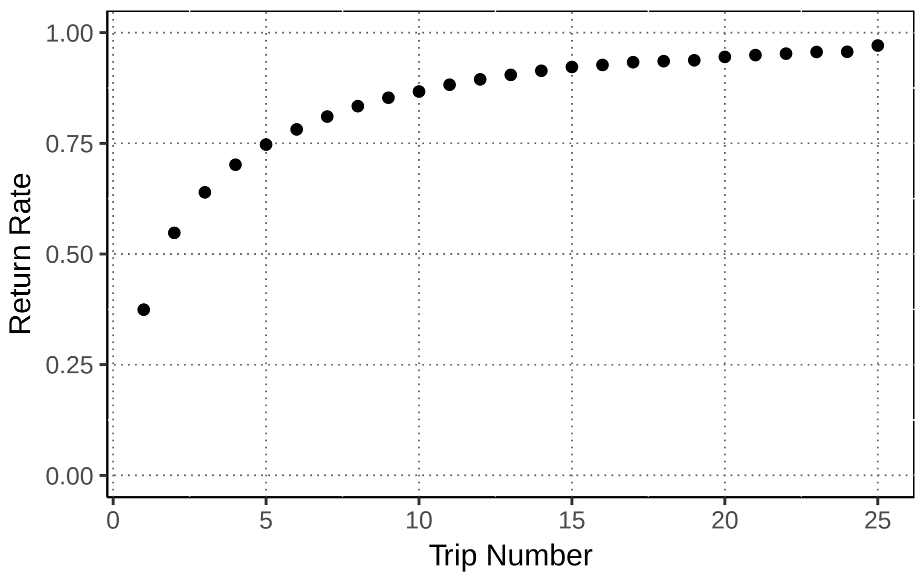


Figure 2: Fraction of guests that return to Airbnb markets

Note: The graph shows the fraction of guests that take an additional trip within 400 days after their first, second, etc trip. The dot for 25 also includes all trips after the 25^{th}. It is based only on trips that happened at least 400 days prior to the end of the data.

likely performed by a bot and restrict the data to the first search a potential guest made for a given market and check-in day. Of these 'first searches' we drop any that had a minimum or maximum price. We also drop searches for the same user and market with a check-in date within two days, since these are in some sense not the user's first search. We are left with approximately 30 million searches per month in 2016. To keep the analysis manageable, we use only 1 month of searches when looking at all searches and use 6 months when looking only at searches that resulted in a booking.

Table 2 summarizes the search variables. The first section of Table 2a is characteristics of the search or searcher, the second describes the search results and the third describes the subsequent outcomes we link to the searches. About 78% of searchers are not people the site recognizes as having previously booked a trip.[11] On average they search for trips 77 days before the searched-for check-in date. Unsurprisingly, the results on average have a substantially higher price than the trips described in Table 1 (people prefer lower-priced listings); the ratings are also slightly higher. The construction of the guest return propensity (GRP) variable is explained below. We only have quality and rating measures for a subset of listings, so we weight the measured average quality (rating) by the share of listings that have a quality (rating) measure, using the mean quality (rating) for the remaining listings.

[11]The site either does not recognize the user or the user is linked (by logging in or by the site recognizing their device) to an account with zero previous trips.

If, within 6 days of that initial search, a user booked a trip for the market searched for, with a check-in date within 1 day of the date searched for, we say that search 'resulted in' that booking. If there are multiple such bookings, we take the first one. As we see in the last section of Table 2, about 3% of searches result in a booking. Much of the analysis is, of necessity, limited to the searches where we can identify the user id of the searcher; users are identified if they log in or if the website recognizes their device. For those users, we match their searches to the set of trips they booked and took subsequently. Searchers we can link take an average of 1.8 trips following the search, spending \$646 dollars. For searches by identified users, 5.7% result in bookings.[12]

Table 2b summarizes some variables that are only available for the booked searches. The average GRP and rating booked are slightly higher than the average shown. The users whose searches we link to a booking spend an average of \$360 on the linked trip and take an average of 2.0 trips, spending \$583, after the linked trip.

2.1 Quality

In addition to the raw data, we construct a quality measure for each listing.[13] As discussed in Section 1.2, we measure host quality as the *guest return propensity*, number of trips a guest takes after staying with that host (controlling for guest characteristics). For each trip taken prior to 2016 by guest i at listing l in period t, we calculate the number of trips the guest booked and took subsequently to get the outcome variable, y_{ilt}. We then regress

$$y_{ilt} = \alpha_0 + \alpha_l + \alpha_1 X_i + \alpha_2 Z_{it} + \varepsilon_{ilt},$$

where α_l is the market the listing is in, X_i is a vector of guest characteristics – whether the guest is verified, if this was their 1^{st}, 2^{nd}, 3^{rd}, or 4^{th}+ trip, gender and age bin[14] – and Z_{it} is a vector of characteristics – the date, number of guests, and (binned) number of nights – of the trip taken by guest i in period t. For each listing we then calculate a *listing effect*

$$f_l = \frac{1}{n_l} \sum_{it} \varepsilon_{ilt} = \frac{1}{n_l} \sum_{it} \left(y_{ilt+} - (\hat{\alpha}_0 + \hat{\alpha}_1 X_i + \hat{\alpha}_2 Z_{it}) \right),$$

where m_l is the number of trips listing l has.

The guest return propensities will be imprecisely measured for listings with a small number of trips hosted. For some of our analysis, we use only listings with at least 20

[12] This rate is by necessity larger than the overall average since any searches where we cannot identify the user did not result in a booking.

[13] We could use a similar approach to try to measure guest quality, but since there are many fewer listings than guests, the number of guests with sufficiently many listings to get an accurate measure is fairly small.

[14] Gender and age are only observed for a subset of the guests. We use 'unknown' as a category for both age and gender, but the results are similar if we use only guests for whom we observe these demographics.

Table 2: Summary search variables

(a) Searches (Jan 2016)

Statistic	Mean	St. Dev.	Min	Median	Max	N
New Traveler	0.781	0.414	0	1	1	28,584,517
Days Advance Searched	76.900	78.900	−1	47	730	28,584,517
Num Nights	7.940	21.700	0	3	731	28,584,517
Num Guests	2.610	2.630	0	2	24	28,584,517
Entire Only	0.194	0.396	0	0	1	28,584,517
Search Results						
Num Search Results	2,416	4,423	1	690	204,673	28,584,517
Avg price of results	247	367	0	120	2,639	28,487,971
Avg Jan 1 Rating	4.630	0.198	1.000	4.660	5.000	28,325,786
Avg GRP (quality)	0.047	0.538	−3.150	0.074	3.150	25,032,831
Outcomes						
Booking (0/1)	0.031	0.174	0	0	1	28,584,517
Trips After	1.790	3.400	0	1	339	15,452,138
Spent After	646	1,719	0	76	642,443	15,452,138

(b) Booked Searches (Jan-June 2016)

Statistic	Mean	St. Dev.	Min	Median	Max	N
GRP (Quality) booked	0.184	0.684	−3.610	0.286	3.628	4,190,158
Rating booked	4.641	0.347	1.000	4.714	5.000	5,304,413
Total price booked	360	564	0	203	60,300	6,711,504
Trips After Check-in	1.987	4.101	0	1	338	6,366,733
Spent After Check-in	583	2,367	0	104	1,525,804	6,366,733

Note: This table summarizes the search data. The top panel is all searches, the bottom panel is only those searches which we are able to match to a booking. “New Traveler” includes searchers who are not linked (by logging in or by the site recognizing their device) to an account or are linked to an account with zero previous trips. “Days in advance” is the number of days between the search date and the searched for check-in date. “Trips taken after” and “Dollars spent after” only include trips that are booked after the search; these variables are only available for searchers who are linked to an account so we can calculate their later behavior. The bottom panel summarizes characteristics that are only available for booked searches. The first three lines give the quality, rating, and price of the trip linked to the search. The next lines give the number of trips taken and number of dollars spent by those guests subsequent to the trip linked to the search.

guest-trips on which to base the estimation. In addition, we adjust the listing effects using a shrinkage estimator to account for the noise in the estimate (see Morris, 1983). To our knowledge, prior work with shrinkage estimators has not accounted for the fact that, in many contexts, one might expect the number of observations might be correlated with the effect being estimated.[15] The number of guests a listing has is likely not independent of its quality, so it does not make sense to to shrink all effects towards the same mean; we modify the usual application of shrinkage estimators accordingly. We group listings by the number of trips they have and the time (quarter) that they were first created.[16] If a group, g, has m listings, and the average listing effect in the group is $E_g[f]$, then the weight for listing l with n_l trips is

$$\delta_l = 1 - \frac{m-3}{m}\frac{1}{n_l}\frac{E_{i\in l}[(\varepsilon_{ilt} - f_l)^2]}{E_{l\in g}[(E_g[f] - f_l)^2]}.$$

and the *adjusted listing effect* is

$$\tilde{f}_l = \delta_l f_l + (1-\delta_l)E_g[f].,$$

Lastly, we winsorize the adjusted listing effects at 3 standard deviations from the mean to get our final measure of GRP.

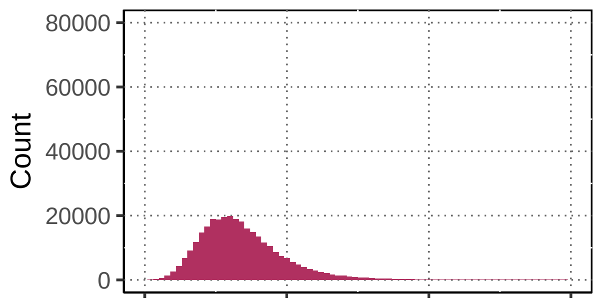
Figure 3a shows the distribution of the raw averages of the number of times a listing's guests return. It is limited to the 308k listings that had at least 20 trips prior to 2016. The average across listings is 3.5 return trips.[17] Figure 3b and 3c show the listing effects after adding guest controls or guest and trip controls. Figure 3d shows the final the residualized, shrunk, and winsorized listing effects, as described above. The standard deviation of the raw effects is 1.87; after all controls it is 1.45; and for the final effects it is 0.55. Guest and market controls do not decrease the variance much, the drop comes mostly from adjusting for the noise from small samples. For the regression analysis below, we normalize the listing effects to have a standard deviation of one (and a mean of zero).

Table 3 shows the correlations between our calculated listing effect and other metrics. The adjusted listing effect is very highly correlated with the unwinsorized version and moderately correlated with the raw averages. However, its correlation with other characteristics of the listings is very low. The correlations with rating, price, number of guests, number of ratings,

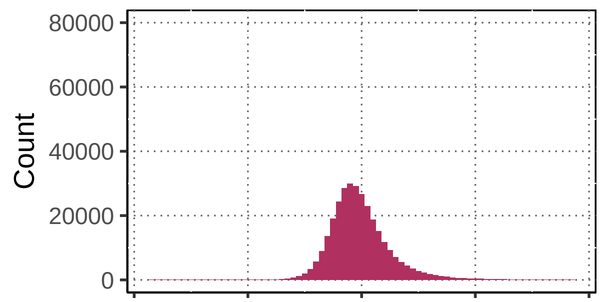
[15]This seems potentially relevant for teacher quality – a common application of shrinkage estimators in economics.

[16]If two listings have the same number of guests, but one has been around twice as long, the newer one is clearly the more attractive listing. Rather than trying to quantify this trade off, we group listings by age interacted with the number of guests.

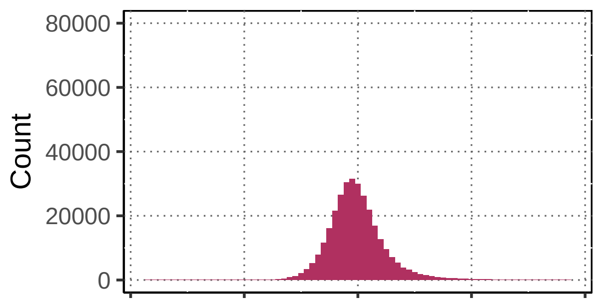
[17]The average across guests is also 3.5 subsequent trips. Unlike the entire sample summarized in Table 1, these trips all happened prior to 2016, so the guests have had more time to accumulate additional trips.



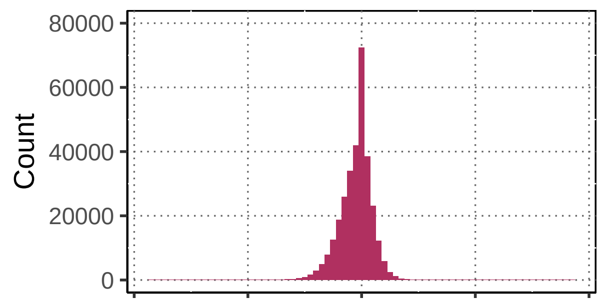
(a) Raw averages



(b) Controlling for guest characteristics



(c) Controlling for guest and trip characteristics



(d) Residualized and shrunk effects

Figure 3: Guests' return trips by listing

Note: These histograms show the distribution across listings of the number of times their guests return, for the 308k listings with at least 20 guests prior to 2016. Figure (a) is the raw averages. Figure (b) is the average listing residual, controlling for guest characteristics; Figure (c) adds controls for the market and number of guests on the trip. Figures (a), (b), and (c) censor 481, 837, and 747 observations, respectively. Figure (d) shows the effects after shrinking towards a group mean to account for small samples, and winsorizing them at 3 standard deviations from the mean.

and the share of guests who left ratings are all below 0.06. We do not know why the correlation with ratings is so low. It may be that ratings are based on things guests think are specific to a given listing, whereas whether a guest returns is not effected by things they think are listing-specific, but precisely by the things they think generalize to all listings.[18]

Hosts on Airbnb can manage multiple listings. We compare the within host variance in listing quality to the overall variance. Both overall and for listings with at least 20 guests, the ratio of within host variance in quality to overall variance is about 85%, so we think that quality is primarily a characteristic of the listing, rather than the host.

[18]It is also possible that for ratings guests have in mind some absolute standard rather than whether they thought the experience was good enough to try again. We thought the latter might be better captured by the "value rating" that guests leave, but its correlation with quality is no higher.

Table 3: Correlation of quality measures across listings

	Guest Return Propensity	Unwinsorized	Raw Average
Guest Return Propensity	1	0.976	0.439
Unwinsorized	0.976	1	0.464
Raw Average	0.439	0.464	1

Correlation of Guest Return Propensity with:				
Rating	Price	Number of Guests	Number of ratings	#Ratings / #Guests
0.0478	0.0341	0.0216	0.0346	0.0509

Note: The top part of the table gives the correlation between the guest return propensity (our main measure of quality), the unwinsorized version and the version not controlling for guest or trip characteristics, for the 308k listings with at least 20 guests prior to 2016. The bottom panel gives the correlation between the guest return propensity (GRP) and the rating, the price, the number of guests, the number of reviews left, and the fraction of guests who left a review. The price is the average price for October 2015 as of 09/01/15; the correlation is based on the 241k listings for which both price and GRP is available. The other measures are as of the end of 2015 and the correlations are based on the 292k listings for which those measures and GRP are available.

3 Results

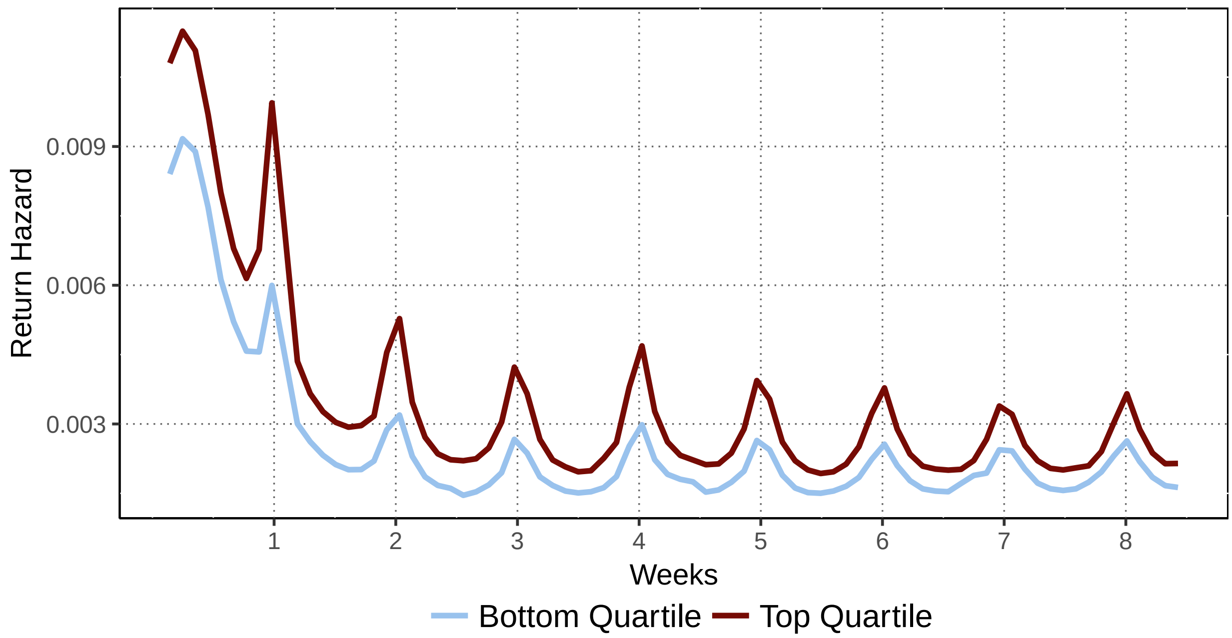
The guest return propensities (GRP) are calculated using the post-trip behavior of each listing's trips from 2011 through 2015. We look at the effect of GRP on subsequent guest behavior using trips starting in 2016.

3.1 All trips

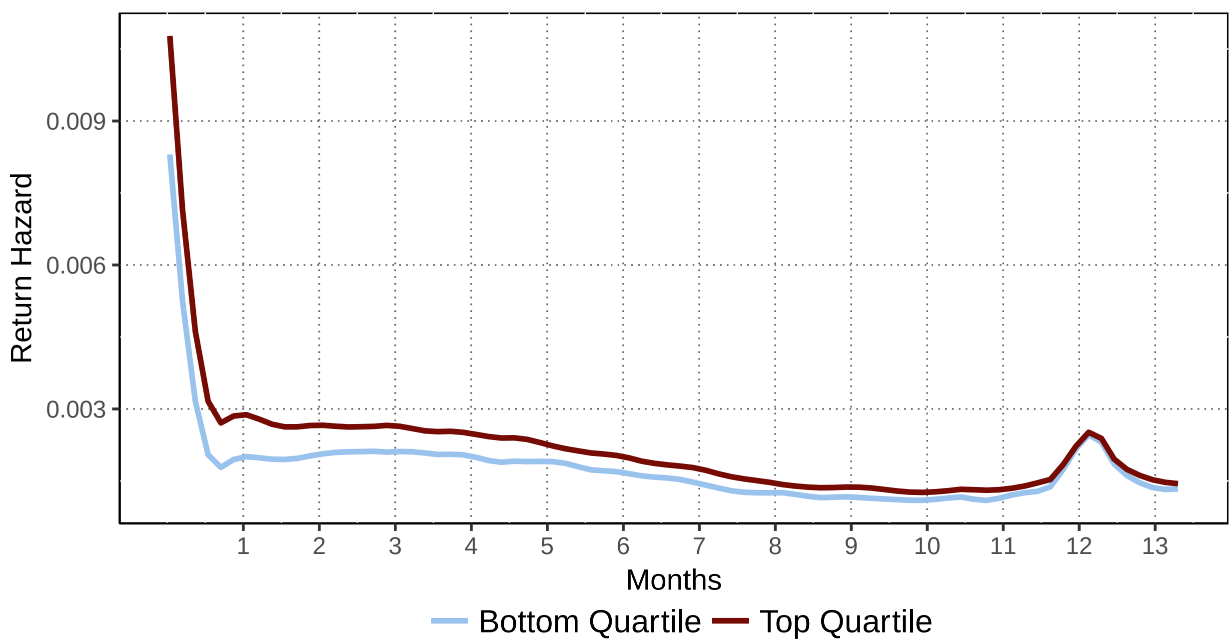
We focus on trips where the listing's GRP is based on at least 20 trips. For each such trip we calculate how many days until the guest either returns or 'leaves the sample' (the data ends).[19] Figure 4 shows hazard rates for a guest returning to Airbnb for listings in the top and bottom quartiles of listing GRP; the hazard rates are the average probability that a guest returns a given amount of time after staying at a listing, conditional on not having returned between the stay at that listing and that point in time. At all time horizons the probability of return is higher for guests that stayed at a listing in the top GRP quartile; the difference falls by about half over the first 6 months.

Table 4 shows the regression results analogous to Figure 4. The first column is the raw effect and the later columns control for ratings and guest and trip characteristics. These are

[19]For this part of the analysis, we exclude all trips where the subsequent trip was booked prior to the trip in question; though the guest could have cancelled that subsequent trip, we think that the effect of a listing's quality on whether the already-booked trip occurred would be negligible. For trips after the already booked trip, it would not be clear whether to use the initial listing or the one from the intervening trip.



(a) First 60 days



(b) First 400 days

Figure 4: Return hazard by quartile of listings' guest return propensity

Note: For listings in the top and bottom quartile of listings' guest return propensity, these graphs show the hazard rates of guests returning to Airbnb; the y-axis shows the (smoothed) probability that a guest returns x weeks or months after staying with one of those listings, conditional on the guest not having returned between the stay at that listing and time x. A listing's GRP is measured by its guests prior to 2016, the returns are shown for guests starting in 2016. Because we cannot control for market in the graphs, we use a measure of GRP that does not control for market. The cyclicality in the top graph is because guests tend to start trips on the same day of the week as their previous trip started on.

the coefficients from Cox hazard regression;[20] since guest return propensity is normalized to have a standard deviation across listings of 1,[21] the interpretation is that switching to a listing with a 1 standard deviation higher GRP corresponds to a guest being exp(.098) $\approx$ 1.1 times as likely to return (at any given point in time, conditional on not having returned previously). The remaining columns repeat the analysis separately for returns after a guest's first, second, third and fourth or more trips. The increase in return probability from a higher GRP host is about 50% larger for first time guests than returning guests. This is consistent with the idea that new guests know less about Airbnb so they update their priors more based on the quality of the listing. The coefficient on rating bounces around, but even when it is significant, it is qualitatively quite small.[22]

Table 4: Effect of listing guest return propensity on hazard of return

	All Trips		1st	2nd	3rd	4+
GRP Booked	0.237***	0.098***	0.123***	0.101***	0.096***	0.082***
	(0.001)	(0.001)	(0.003)	(0.004)	(0.004)	(0.002)
Rating at booking		−0.002	0.003	0.015***	0.006	−0.008**
		(0.002)	(0.005)	(0.006)	(0.006)	(0.003)
Controls	0	1	1	1	1	1
Observations	9,003,856	5,773,355	1,668,324	1,074,219	740,012	2,290,800
R^2	0.004	0.134	0.114	0.066	0.047	0.067

*p<0.1; **p<0.05; ***p<0.01

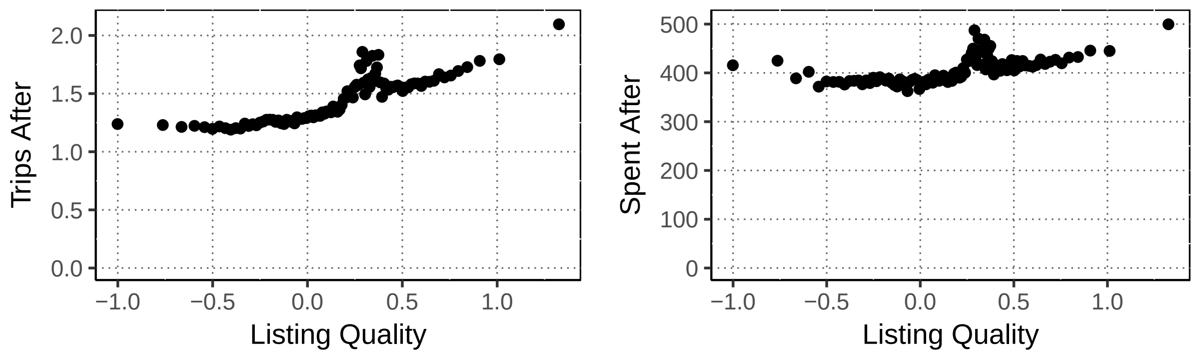
Note: This tables show the coefficients from Cox Hazard regressions, analyzing the effect of listing GRP on the probability that guest returns to Airbnb over time. Because we cannot include the thousands of market fixed effects in the hazard regression, we use a measure of GRP that does not control for market. Column (2) adds the listing's rating at the time of booking and controls for when the initial trip happened, the guest's age bin, gender, experience, and verified-status as well as the trip's number of nights, number of guests, nightly price and whether it was booked through instant-book. Columns (3)-(7) run the same analysis as Column (2) separately for different subsamples of guests – those for whom the trip was their 1st, 2nd, 3rd or 4th or more trip on Airbnb.

The platform cares about how many times guest return and how much money they spend

[20] Since we cannot control for the thousands of market fixed effects in the Cox regression, for this analysis, we use a measure of guest return propensity that does not adjust for the listing's market. If we use the fully-adjusted measure the effects are about half as large; however, we think our approach – controlling for market in neither the guest return propensity calculations nor the regressions – is closer to controlling for market in both analyses than this alternative of adjusting for market the guest return propensity calculations, but not the regressions.

[21] The standard deviation across listings is 1, the standard deviation across trips in this sample is 0.45

[22] Since the standard deviation of across listings of the rating is 0.58, the coefficient of 0.015 implies that 2^{nd} time guests are 1.009 times as likely to return if they stay at a listing with a one standard deviation higher rating.



(a) Trips after

(b) Spent after

Figure 5: Binscatter of trips taken and dollars spent by ventile of listing GRP

Note: These graphs group listings into GRP percentiles. Each dot shows the average GRP of a group of listings (x-axis) and the average number of trips taken (panel (a)) or dollars spent (panel (b)) by a guest after staying at those listings. GRP is calculated based on trips prior to 2016. The y-axis measures are based on trips starting in 2016. Because we cannot control for market in the graphs, we use a measure of GRP that does not control for market.

on the platform, not just whether they return. For each trip we calculate how many trips they booked (and took) after that trip and how many dollars were spent on those subsequent trips. Figure 5 shows a binscatter of the trips taken and dollars spent after a trip by listing GRP. These measures are also increasing in listing GRP, though the spending measure is high variance.

Tables 5 and 6 give the corresponding regression results. On average, a 1 standard deviation higher GRP host predicts guests taking 0.19 additional trips and spending $48 additional dollars on Airbnb. Here the effects tend to be larger for guests with more experience – because these guests travel more and spend more, even if though their beliefs about Airbnb are less affected, small changes in their propensity to use it when traveling lead to larger changes in the number of trips and dollars spent on the platform.[23] The standard deviation of rating across listings[24] is 0.58 so the effect of a one standard deviation increase in the rating is about a tenth of the effect of a standard deviation increase in GRP. The effects for spending are more comparable.[25]

[23]Host quality could also affect the extent to which guests tell their friends about Airbnb and thereby "recruit" new potential guests. It seems likely that guests that are more likely to return are also more likely to encourage friends to try Airbnb, so in general we think that if this is an important margin, we likely understate the effects of host quality. However, the extent to which guests 'recruit' could also vary with guest experience – if experienced guests have already told all their friends, the effect of quality on new guests they recruit will be small. For a growing platform like Airbnb, a differential effect on new recruits could be large enough to outweigh the differences in travel frequencies. With information on referrals, the platform could adjust our metric to include effects on new guests.

[24]Like GRP, the standard deviation across trips is smaller than across all listings, it is 0.25.

[25]Because the variance in spending is so high, we repeat the analysis with winsorized spending. The

Table 5: Effect of listing GRP on trips after

	All Trips		1st	2nd	3rd	4+
GRP Booked	0.011***	0.192***	0.070***	0.075***	0.084***	0.323***
	(0.002)	(0.003)	(0.003)	(0.005)	(0.006)	(0.007)
Rating at booking		0.018***	0.026***	0.010	0.002	0.015
		(0.006)	(0.006)	(0.009)	(0.012)	(0.013)
All Controls	0	1	1	1	1	1
Observations	11,499,195	7,710,665	2,016,255	1,384,152	989,703	3,320,555
R^2	0.00000	0.136	0.116	0.120	0.131	0.128

*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the effect of the GRP and rating of a listing on the number of trips a guest takes subsequently. The first column looks only at the effect of GRP. The other columns add the listing rating at the time the trip was booked and guest, trip and market controls. Columns (3)-(7) run the analysis separately for guests for whom the trip was their 1st, 2nd, 3rd or 4th or more trip on Airbnb.

Table 6: Effect of listing GRP on spent after

	All Trips		1st	2nd	3rd	4+
GRP Booked	387.351***	47.980***	24.306***	12.976***	24.536***	73.378***
	(0.997)	(1.591)	(1.324)	(3.144)	(4.338)	(3.082)
Rating at booking		26.475***	35.897***	26.580***	36.755***	20.921***
		(3.152)	(2.580)	(6.112)	(8.469)	(6.245)
All Controls	0	1	1	1	1	1
Observations	11,499,195	7,710,665	2,016,255	1,384,152	989,703	3,320,555
R^2	−0.046	0.059	0.072	0.030	0.031	0.068

*p<0.1; **p<0.05; ***p<0.01

Note: This tables show the effect of the GRP and rating of a listing on the amount of money a guest spend on Airbnb trips subsequent to staying there. The first column looks only at the effect of GRP. The other columns add the listing rating at the time the trip was booked as well as guest, trip and market controls. Columns (3)-(6) run the analysis separately for guests for whom the trip was their 1st, 2nd, 3rd or 4th or more trip on Airbnb.

Prices

The theory suggests that new information may also change guests' beliefs about the relationship between price and quality and therefore their optimal price point. We look at whether high GRP has a different effect on guest behavior when it is a cheaper verse a more expensive listing. First, we regress the nightly price of a trip on the number of guests, the number of nights, and the market interacted with check-in date. We use the residual from this regression as a measure of how expensive the listing was.[26]

Table 7 shows the effect of GRP by (residualized) price quartile, looking both at the number of subsequent trips taken and the price of the next trip booked. Once we control for guest and trip characteristics, there is only a slight difference in the effect of GRP by price quartile, with GRP at lower priced listings having a slightly larger effect. We do not see a significant difference in the effect of GRP by price quartile on the price of the subsequent booking. This suggests that either guests do not update a lot their beliefs about the correlation between price and quality or the different effects of correlation,[27] shown in Equation (1), cancel out.

If consumers do not properly account for the price level of a market or the added cost of certain types of trips, then the relevant measure would be the actual price, not the price residual. These results are shown in Appendix Table A2. The effect of GRP on subsequent trips differs slightly more across quartiles (.20 for the lowest price quartile and 0.08 for the highest price). The effect on prices is reversed: at a listing in the top price quartile, a higher GRP leads guests to book a *lower*-priced subsequent trip, but the effect is tiny.[28] A standard deviation increase in GRP corresponds to a place that is \$.78 cheaper per night (relative to a standard deviation of \$87).

---

results, shown in Appendix Table A1, are similar, though the effect for GRP is slightly smaller and that for listing is somewhat larger. The largest effect of GRP is still on those with the most experience.

[26] We again exclude all trips where the subsequent trip was booked prior to the initial trip in question. See Footnote 19.

[27] Correlation increases the expected quality return to choosing a higher priced listing (pushing towards choosing higher-priced listings) but also increases the expected quality at a given price, which pushes towards lower priced listings if there are decreasing returns to quality.

[28] As we discuss in Section 1.1, the choice to book a cheaper listing is consistent with the idea that there are decreasing returns to quality and those listings are is some sense nicer than the guest wanted. The guest's response is effectively, "if this is how good an expensive place is, than a moderately-priced place is probably just fine." But the effect is so small, we do not put much stock in it.

Table 7: Effect of GRP by price quartile: price of next booking and subsequent trips

	Trips	Trips	Price	Price
Price Residual	-0.0004^{***}	-0.0001^{**}	0.234^{***}	0.217^{***}
	(0.00003)	(0.00003)	(0.001)	(0.001)
Quality x Price Q1	-0.020^{***}	0.147^{***}	0.353^{***}	0.094
	(0.005)	(0.006)	(0.125)	(0.140)
Quality x Price Q2	-0.012^{***}	0.125^{***}	0.714^{***}	0.173
	(0.005)	(0.006)	(0.121)	(0.136)
Quality x Price Q3	0.038^{***}	0.138^{***}	0.796^{***}	0.217
	(0.005)	(0.006)	(0.121)	(0.137)
Quality x Price Q4	0.064^{***}	0.116^{***}	0.994^{***}	0.470^{***}
	(0.005)	(0.006)	(0.126)	(0.140)
All Controls	0	1	0	1
Observations	11,460,917	7,701,187	5,072,031	4,207,258
R^2	0.0005	0.247	0.071	0.084

*p<0.1; **p<0.05; ***p<0.01

Note: This tables show the effect of the GRP of a listing on the number of subsequent trips a guest takes and the price residual of the next listing booked by a guest. The effect is shown separately by the quartile of the price residual of the initial booking. The first and third columns looks only at the effect of GRP. The second and fourth columns add guest, trip and market controls as well as the rating of the initial listing at the time the trip was booked.

3.2 Searches

The evidence above suggests that these listing quality differences are persistent and meaningful, but it is still possible that the difference is actually on the guest side – certain guests are (unobservably) more likely to return and take more trips and also tend to chose certain listings. To account for the endogeneity of guest choices, we use an instrumental variables analysis based on guests' search results. Frequent bookings, changes in availability, and ongoing search experiments mean that guests searching for the same market and check-in date on the same day are shown different search results. Their choice for where to stay is affected by these results. At a macro scale, the results may be correlated with unobserved guest characteristics – if some guests plan farther in advance and there may be higher-GRP listings available then or certain types of guests may look to travel in certain markets that have higher or lower average GRP. However, we believe that after controlling for the market interacted with the check-in date searched for and the amount of time in advance of the search, the time of day interacted with the timezone in which the search happened, and the search filters, any residual variation in the GRP of the listings shown is as good as random.

Because the sample size is smaller, and the IV analysis is less affected by measurement error, we use all listings for this analysis, not just those whose GRP is measured based on at least 20 guests. Table 8 shows the reduced-form effects of the average GRP, rating, and price on trips taken subsequently. The first two columns of Table 8 show the reduced-form effect of the average GRP and rating of the listings shown on the number of trips the searcher takes after *searching*. The first column is for all searches we can identify; the second is for those who book a trip related to the search – the sample relevant for the IV analysis. The first column uses searches from January, 2016. When we limit to booked searches the sample size shrinks dramatically, so we use searches January – June 2016.[29] Columns (3) and (4) use the same sample of booked searches as Column (2) but look at the effect on the number of trips the searcher takes after *check-in*. For the IV, where we're interested in the effect of the listing characteristics, it makes more sense to use this measure, which does not include trips taken prior to the trip at that listing.[30] Column (3) uses the number of past trips the guest had taken at time of search as a control, Column (4) uses the number of past trips the guest had taken at the time of check-in – again the variable more relevant for the IV analysis. All columns include the full set of market, check-in, and search timing controls.

Recall that GRP is normalized to have a standard deviation of 1 across listings. However, the standard deviation in the average GRP of the search results is only 0.16. So Column (4) of Table 8 implies that raising the GRP of all the listings by one standard deviation of listing GRP results in 0.14 more trips, but raising the average GRP by one standard deviation of average GRP shown only results in 0.02 additional trips. The effects of rating are even smaller. The standard deviation across listings in the rating as of January 1 is 0.58 and the standard deviation in average rating shown is 0.12. So raising the rating of all the listings by one standard deviation of listing rating would result in 0.05 more trips, but raising the average rating shown by one standard deviation of average rating shown only results in 0.006 additional trips.

One reason why the search results have a fairly small effect on subsequent behavior is that their effect on the type of listing booked is modest. The first two columns of Table 9 show the effect of the average GRP and rating shown on the GRP and rating booked.

[29]Since our quality measure is as of the end of 2015, the fraction of search results with a quality measure available decreases over time, as the fraction of listings created since the beginning of January, 2016 increases. To the extent that quality changes over time, the quality measure also becomes less accurate. For these reasons, we use only the first half of 2016.

[30]Trips would be include in the first measure and not the second if they were booked after the search and took place before the associate booking.

Table 8: Effect of average GRP and rating of search results on number of subsequent trips

	Trips After Search		Trips After check-in	
	January Searches	Booked Searches		
Avg GRP shown	0.054***	0.052***	0.127***	0.142***
	(0.004)	(0.009)	(0.015)	(0.015)
Avg Jan 1 Rating	−0.020***	0.090***	0.016	0.051**
	(0.007)	(0.014)	(0.022)	(0.022)
Avg price of results	−0.00005***	−0.0001***	−0.0001***	−0.0001***
	(0.00000)	(0.00001)	(0.00002)	(0.00002)
1 Past Trips	0.814***	0.674***	0.589***	
	(0.002)	(0.003)	(0.005)	
2 Past Trips	1.465***	1.162***	1.036***	
	(0.002)	(0.004)	(0.007)	
3+ Past Trips	3.073***	2.678***	3.055***	
	(0.002)	(0.003)	(0.005)	
1 Past Trips				0.568***
				(0.005)
2 Past Trips				0.996***
				(0.007)
3+ Past Trips				2.893***
				(0.005)
Observations	14,483,541	6,546,771	6,211,480	6,211,480
R^2	0.304	0.522	0.493	0.491
Adjusted R^2	0.215	0.212	0.152	0.149

*p<0.1; **p<0.05; ***p<0.01

Note: The first column shows the effect of the average GRP and rating of search results on the number of trips taken afterwards for all guests we can identify for searches in January 2016. The remaining columns are limited to searches that resulted in a booking; to maintain a large enough sample we use searches in January - June 2016. Column (2) repeats the analysis of Column (1) of trips taken *after the search* for the IV-subsample. Columns (3) and (4) show the effect of the average GRP and rating of search results on the number of trips taken *after the check-in date*. For the first three columns, "Past Trips" refers to the number of trips the user took prior to the search; for the last column it is the number of trips prior to the trip's check-in date; for both zero is the reference group. All regressions control for market x check-in-date x how far in advance the search happened, the search date, the hour of the search interacted with the timezone the searcher was searching from and any search filters the searcher used.

The estimated effect of average GRP shown on average GRP booked is 0.45; for ratings the analogous effect is 0.37. The cross effects are negligible. Our measures of the characteristics of the listings shown use only the average of the results on the first page and only the first search the user makes for a given market and check-in range (to avoid guests' desire to search more influencing the search results); other results they see may differ, so the effect of the listings shown on the listing booked is less than 1.

The last column of Table 9 puts the preceding results together in an IV regression. We instrument for the characteristics of the booked listing with the characteristics of the search results and look at the effect on trips taken afterwards. We find that a one standard deviation increase in the GRP of the listing booked leads to 0.34 additional trips by that guest. On average, in this sample, a guest takes 2 subsequent trips, so that is an increase of 17%.[31] The coefficient on the rating is smaller and recall that there is also less variance across listings in their rating so a one standard deviation increase in the rating of a listing leads to 0.10 additional trips by the guest.[32]

The identifying assumption for our IV regressions is that given the controls – (1) the market searched for, interacted with the check-in date searched for, interacted with bins for how far in advance the search was[33] (2) the date the search happened (3) the timezone the searcher was in interacted with the hour of day the search happened (4) the search filters the searcher used – the remaining variation in the GRP and rating of listings shown is uncorrelated with guest characteristics. The remaining variation results from listings being added, listings being booked, and experiments in Airbnb's search algorithm.[34]

In addition to affecting what place a searcher books if the search results in a booking, the search results can also affect whether the search results in a booking. Table 10 looks at the effect of search results on this latter, extensive margin effect. The effects are fairly small: a 1 standard deviation increase in the average GRP shown increases the booking rate by $.002 * .16 \approx .00032$ percentage points. Since the baseline booking rate for known guests is about 5.7%, this an increase of 0.6%. The fact that these effects are non-zero means that

[31] The 0.34 coefficient is not directly comparable to the standard deviation of 0.55 trips that GRP had before we normalized it to 1, because it is based on an earlier sample. The mean number of return trips for trips prior to 2016 was 3.5.

[32] If instead of controlling for the average price shown, we instrument for the price booked with the price shown, the coefficients on GRP and rating are largely unaffected and coefficient on price is still tiny.

[33] We do separate bins for each of 0 through 30, then bin by week for searches less than a year in advance. For searches more than a year in advance we have two bins: less than 400 and greater than 400. Since we have thousands of markets and thousands of check-in dates, this binning makes the interaction tractable. The exact search date is also controlled for, but not interacted with the market and check-in date.

[34] We do not directly use this experimental variation because the first stage F-statistic is about 1. Since the search algorithm is not considering this measure of GRP, and it is not strongly correlated with prices or ratings, the experiments do not have large effects on the average GRP shown. We are working with Airbnb to implement an experiment that directly varies the GRP of listings shown.

Table 9: First stage and IV results:
Effect of search results on booking characteristics, and instrumented effect of booking characteristics on subsequent trips

	First stage		
	GRP Booked	Rating Booked	Trips After (IV)
GRP booked			0.340***
			(0.043)
Rating booked			0.190**
			(0.092)
Avg GRP shown	0.449***	0.002	
	(0.003)	(0.002)	
Avg Jan 1 Rating	0.006	0.329***	
	(0.005)	(0.002)	
Avg price of results	−0.00004***	−0.00001***	−0.0001**
	(0.00000)	(0.00000)	(0.00003)
1 Past Trips	0.003**	0.009***	0.570***
	(0.001)	(0.001)	(0.007)
2 Past Trips	0.001	0.014***	0.991***
	(0.002)	(0.001)	(0.009)
3+ Past Trips	0.006***	0.026***	2.871***
	(0.001)	(0.001)	(0.007)
F-Stat	NA	NA	32899
Observations	4,094,066	4,115,700	3,878,265
R^2	0.519	0.513	0.555
Adjusted R^2	0.090	0.079	0.143

*p<0.1; **p<0.05; ***p<0.01

Note: The first three columns show the effect of search results on the characteristics of the booking, the first stage for the IV regression. The second column is limited to trips at listings for whom the GRP measure is based on at least 20 trips. The last column shows the IV results where we use the average GRP and rating of search results to instrument for the GRP and rating booked. For the first three columns, "Past Trips" refers to the number of trips the user took prior to the search; for the last column it is the number of trips prior to the trip's check-in date; for both zero is the reference group. All regressions control for market x check-in-date x how far in advance the search happened, the search date, the hour of the search interacted with the timezone the searcher was searching from and any search filters the searcher used.

Table 10: Effect of search results on whether a booking occurs

	All	All
Avg GRP shown	0.001***	0.002***
	(0.0002)	(0.0004)
Avg Jan 1 Rating	0.002***	0.002***
	(0.0004)	(0.001)
Avg price of results	−0.00001***	−0.00003***
	(0.00000)	(0.00000)
1 Past Trips	0.005***	0.004***
	(0.0001)	(0.0002)
2 Past Trips	0.015***	0.014***
	(0.0002)	(0.0002)
3+ Past Trips	0.035***	0.032***
	(0.0001)	(0.0002)
Unknown Guest	−0.053***	
	(0.0001)	
Observations	27,011,597	14,483,541
R^2	0.101	0.124
Adjusted R^2	0.037	0.012

*p<0.1; **p<0.05; ***p<0.01

Note: This table shows the effect of the average GRP and rating shown on whether a search results in a booking. Searches are matched to bookings by that user that happen within 6 days of the search and have the same destination market and check-in date +/− 1 day. The first two columns include searches by unidentified searchers (none of which result in a booking). In the second column the excluded group for the number of past trips is unknown searchers (whose number of past trips is unknown). The last column is limited to identified searchers.

when we look at the affect of booking characteristics on subsequent guest behavior, we have a slightly selected sample. However, if we assume that the guests who were less likely to book (and only did so because there were listings with a higher rating or GRP) are also less likely to take additional trips, then our results are a lower bound on the effect on an unselected sample.

Guest Experience

Guests with different amounts of past experience with Airbnb will both have more information about the platform and are likely to be people who travel more frequently. The first column of Table 11 repeats the IV analysis from the third column of Table 8; the subsequent columns show the same regression separately for guests for whom this is their first, second, third, or fourth (or more) trip. The effect of GRP is larger for guests with more experience on Airbnb: the higher frequency of travel outweighs the smaller learning effect.[35] Interestingly, though we lose power, the point estimates suggest that the effect of rating might be largest for first time guests.

Table 11: Effect of GRP and rating booked on subsequent trips, by guest experience (IV)

	By Trip Number				
	All	1st	2nd	3rd	4+
GRP booked	0.340***	0.080**	0.220**	0.301	0.873***
	(0.043)	(0.040)	(0.109)	(0.192)	(0.200)
Rating booked	0.190**	0.250***	0.152	0.603	−0.638
	(0.092)	(0.079)	(0.225)	(0.416)	(0.423)
Avg price of results	−0.0001**	0.00002	−0.00002	0.0002*	−0.0002
	(0.00003)	(0.00002)	(0.0001)	(0.0001)	(0.0001)
F-Stat	32899	13879.8	5217.4	3358.2	7424.2
Observations	3,878,265	1,630,636	730,491	430,529	1,086,609
R^2	0.555	0.640	0.774	0.842	0.704
Adjusted R^2	0.143	0.056	0.058	0.058	−0.003

*p<0.1; **p<0.05; ***p<0.01

Note: This table repeats the IV analysis of the effect of listing GRP on subsequent guest trips for subsamples of guests based on their number of prior trips with Airbnb. The first column repeats the results from the last column of Table 9, for reference. Columns (2)-(5) do the same analysis for guests for whom it is their first, second, third or fourth (or more) trip. All regressions control for market x check-in-date x how far in advance the search happened, the search date, the hour of day of the search interacted with the timezone the searcher was searching from and any search filters the searcher used.

What about Dollars?

If the platform takes a percentage cut instead of a flat fee, it will care about the dollars spent by guests, not just the number of times they return. Table 12 looks at the effect of the GRP and rating booked on the amount (in dollars) that the guest spends on subsequent trips on Airbnb, again instrumenting for the GRP and rating booked with the average GRP and average rating of the search results. Here rating plays a larger role, but the effect of GRP is still significant – statistically and economically. A one standard deviation increase in GRP leads to $111 more spent subsequently. The standard deviation across listings of the rating

[35] In this sample a first-time guest returns an average of 0.9 times and a fourth-time guest returns an average of 4 times, so the point estimates suggest that, contrary to our expectations, the percentage increase is also larger for more experienced guests, but the estimates are not precise enough to rule out the opposite.

is 0.58 so a standard deviation increase in rating leads to $250 more spent subsequently, slightly larger than the effect of GRP.[36]

Table 12: Effect of GRP and rating booked on subsequent spending, by guest experience (IV)

	By Trip Number				
	All	1st	2nd	3rd	4+
GRP booked	111.718***	18.995	68.864	141.485	381.593***
	(22.994)	(17.716)	(51.220)	(87.897)	(102.240)
Rating booked	430.002***	152.824***	227.085**	369.038*	737.493***
	(48.983)	(35.509)	(105.728)	(190.266)	(216.745)
Avg price of results	0.421***	0.192***	0.315***	0.616***	0.839***
	(0.014)	(0.011)	(0.030)	(0.055)	(0.057)
F-Stat	32899	13879.8	5217.4	3358.2	7424.2
Observations	3,878,265	1,630,636	730,491	430,529	1,086,609
R^2	0.473	0.589	0.724	0.817	0.723
Adjusted R^2	−0.015	−0.075	−0.150	−0.096	0.061

*p<0.1; **p<0.05; ***p<0.01

Note: This table uses the same IV approach as Table 11 to show the effect of listing characteristics on a guest's subsequent spending (in dollars) on the platform. The first column shows the effects for all guests (controlling for guest experience). Columns (2)-(5) do the same analysis for guests for whom it is their first, second, third or fourth (or more) trip. All regressions control for market x check-in-date x how far in advance the search happened, the search date, the hour of the search interacted with the timezone the searcher was searching from and any search filters the searcher used.

Since higher rated listings are more expensive (and higher GRP ones are not), it is perhaps not surprising that rating has a large effect on dollars spent even though GRP has a larger effect on trips. If booking more highly rated places also causes guests to return to more highly rated, and therefore more expensive places, the effect on spending would be disproportionately larger than the effect on the number of trips. [37]

[36] A version of Table 12 using winsorized spending is Appendix Table A3; the results are similar.

[37] In Appendix Table A4, we look at this directly by using the price and rating of the place booked subsequent to the trip of the booked search as the left-hand side variable. The sample size is smaller because this only includes booked searches where the guest made an additional booking after the trip associated with the search. As expected, the rating booked increases both the rating and the price of the subsequent listing booked. The effect of GRP is negligible. The price initial listing has a small effect on the price of the next booking.

4 Implications for the Platform

Having seen that there is substantial variation across listings in the externality that they exert on the rest of the platform, the next question is what the platform can do with that information. We consider four ways the platform can raise listing quality or redirect guests to higher-quality listings.[38] Often when there are externalities in a market, the efficient solution is a Pigouvian tax to make agents internalize the externality. However, platforms also have the option to incorporate quality into the ranking of search results, to inform users about listing quality, to remove low quality hosts from the platform, or to try to help low-quality hosts improve.

Taxing the externality

The platform could pay hosts when their guests return or fine them if the guests do not return within a certain time.[39] This would help high-quality hosts over low-quality hosts, and give all hosts an incentive to raise their quality. Since guests' returns are stochastic, the tax would also create additional variance in a host's revenue.

How much host quality would improve depends on how convex the costs to improving quality are, where costs include both the cost of figuring out what things make guests return and the cost of doing those things. Unless the platform could also give hosts guidance as to how to improve their guest return propensity, the changes might be fairly small. However, even if the externality for a given transaction is unchanged, Pigouvian taxes can improve efficiency by affecting prices (and quantities). A revenue-neutral tax/subsidy for quality would cause low quality hosts to increase prices and high quality host to decrease prices, shifting demand to hosts with better externalities.[40]

To get a sense of the magnitude of the effects of a Pigouvian tax on quantities and surplus, imagine an incremental implementation of a tax/subsidy in proportion to quality. If quality is measured in trips, and each trip has a value to the platform as a whole (including sellers) of V, then the revenue-neutral tax would be $\tau_j = \tau(\overline{GRP} - GRP_j)V$. If quality does not change, the change in platform surplus is the change in each host's number of guests, Δd_j,

[38] We are implicitly assuming that a listing's quality is not affected by the number of guests it gets. Our estimates cannot speak to whether average and marginal quality are the same – if a listing gets more guests it could gain experience and improve or be too busy and lose quality. Nevertheless, we think the platform will likely want guests to go to higher-quality listings.

[39] It would probably be more practical to give the host part of the payment at the time of the trip and the rest only if the guest returns.

[40] If there are fix costs to being on the platform, a tax on low quality could also cause low-quality hosts to leave the platform. This is related to the effects of screening discussed below.

times the number of trips a guest who stays with that host takes, times the value of a trip:[41]

$$\Delta S = \sum_j \Delta d_j (1 + GRP_j) V.$$

The changes in quantities will depend on how much hosts pass through their cost changes to prices and on how responsive demand is to price. If we take a super-simplified case where the equilibrium prior to the tax is symmetric, then marginal costs, c_j, the own- and cross-derivatives of demand $\left(\frac{\partial d_j}{\partial p_j}, \frac{\partial d_j}{\partial p_k}\right)$, and own- and cross- pass-through rates $\left(\frac{\partial p_j}{\partial c_j}, \frac{\partial p_j}{\partial c_k}\right)$, will all be the the same across listings. In this case, the effect on welfare is

$$\frac{\partial S}{\partial \tau} = -J \cdot \frac{\partial d_j}{\partial p_j}(1 + DR_{jk}) \left(\frac{\partial p_j}{\partial c_j} - \frac{\partial p_j}{\partial c_k}\right) var(GRP) \cdot V^2, \tag{2}$$

where J is the number of listings and $DR_{jk} = -\frac{\partial d_k}{\partial p_j} \big/ \frac{\partial d_j}{\partial p_j}$ is the diversion ratio (the fraction of the consumers that j loses when it raises its price that become consumers of k).[42]

The tax is revenue neutral, so average cost change across all listings is zero. Therefore, as the number of listings in the relevant market gets larger, the average change in the competitors' costs is close to zero. Similarly, the changes in competitors' prices will cancel out. Because of this, the effect of demand diversion (DR_{jk}) and cross pass-through ($\frac{\partial P_i}{\partial c_j}$) will be negligible. Moreover, if sellers are pricing optimally then the margin equals the semi-elasticity of demand $\frac{V}{p_j} = p_j - c_j = \frac{d_j}{\frac{\partial d_j}{\partial p_j}}$. In this case, the fraction increase in surplus is approximately[43]

$$\frac{\partial S}{\partial \tau}\frac{1}{S} \approx \frac{\partial p_j}{\partial c_j} var(GRP). \tag{3}$$

We see from Equation (2) when the benefits of a Pigouvian tax will tend to be larger. When demand is more responsive to price (for a given number of listings), more consumers will shift to higher quality listings. Therefore, the tax is more beneficial when demand is more responsive. Similarly, when more consumers switch to other listings rather than leaving the platform when a price increases, (i.e. the diversion ratio is higher), the benefit from the tax is larger. In addition, the more listings pass the cost change through to prices, the more

[41] If guests who do not book a given trip still return in the future, then this term should in include $\Delta d_0(1 + GRP_0)$, where $\Delta d_0 = -\sum_j \Delta d_j$

[42] As τ moves away from zero, the equilibrium becomes less symmetric, so this formula holds for all τ only if demand is linear and pass-through is constant over the relevant range of costs.

[43] We simplify

$$\frac{\partial S}{\partial \tau} \approx -J \cdot \frac{\partial d_j}{\partial p_j} V \cdot \frac{\partial p_j}{\partial c_j} var(GRP) \cdot V = -J \cdot d_j \frac{\partial p_j}{\partial c_j} var(GRP) \cdot V,$$

and use the fact that total surplus is the number of trips $J \cdot d_j$ times the surplus per trip V.

the tax will affect prices and shift demand; so, the benefit from the tax is higher when pass-through is higher.[44] Lastly, and unsurprisingly, the welfare gains from the tax are larger when there is higher variance in quality and higher value of each trip.

For a back-of-the-envelope calculation, we can use Equation (3) and our estimate of 0.34 for the effect of a standard deviation increase in quality on guests' subsequent trips. If a listings' pass-through rate were one-half,[45] and implementation and brand-image concerns were not an issue, a Pigouvian tax on quality could increase the surplus per trip by about 6%.

Search results

While the platform may consider Pigouvian taxes infeasible or undiplomatic, the platform can directly shift quantity from low to high quality hosts, without taxes or price changes, via its search algorithm.[46] Like taxes, including quality in the search rankings would give hosts an incentive to improve quality, if they are aware that it affects the search ranking. Promoting high-quality listings would also raise the surplus for given levels of quality. The more search rankings affect which listings guests book, the more incorporating quality into the search algorithm will raise efficiency and incentivize hosts to improve.

To think about the platform's trade offs in which listings to promote to users (e.g. put on the first page of results) we need to think about another dimension of listing quality. Listings vary not only in the propensity of their guests to return to the platform, but also in their propensity to be booked when shown to guests – their *purchase propensity* or 'attractiveness'. In reality, there is horizontal heterogeneity – a listing's booking rate will depend on which other listings it is shown with. Listings may also vary in what fraction of their bookings are 'new' bookings as opposed to bookings 'taken' from other listings. However, since our focus is on the effect of incorporating guest return propensity into the rankings, we assume that, conditional on what the platform knows about a searcher, attractiveness is uni-dimensional.

If the platform is just trying to maximize current bookings, it will show the most attractive listings (those with the highest booking-propensity) that meet the search criteria. If the platform is maximizing the number of bookings in the long term, it will trade-off the listing's attractiveness, A, and its guest return propensity, GRP. Abstracting from guest

[44] Since a listing's cost moves in the opposite direction as the average of other listings' costs, if cross price pass-through is positive, that attenuates the listing's price change.

[45] If we do not think sellers are necessarily pricing optimally, we can choose the elasticity of demand and margin separately. If $\varepsilon = -4$, and $\frac{V}{P} = .2$, then the fractional gain in surplus gain is $4 \cdot .2 \cdot .5 \cdot .12 \cdot = .05 \cdot$. The formula is linear in the elasticity, the margin, and the pass-through rate, so it is easy to see how the result would change for different values of these parameters.

[46] Changes in the search algorithm can be thought of as the 'nudge' alternative to the 'paternalistic' banning of hosts.

heterogeneity, the platform will pick the listings with the highest total expected bookings

$$b(A_l, GRP_l) = A_l(1 + GRP_l - GRP_0),$$

where GRP_l is the expected number of trips that a guest takes after staying with that host, and GRP_0 is the number of trips that a guest who does not book this trip will end up taking in the future.[47]

The gains from maximizing $b(A_l, GRP_l)$ instead of just A_l are (1) decreasing in the correlation between GRP and A, (2) increasing in the variance of GRP, and (3) decreasing in the variance of A.[48] If we use ratings as a proxy for attractiveness, then in the case of Airbnb the correlation between A and GRP is low. The variance in GRP is about 0.11 trips. Our data do not speak directly to the variance of A (since ratings are not in the units of booking probability). However, if we think about the gain from re-ordering only the 20 most attractive listings by the true ranking $b(A_l, GRP_l)$, we can say that the variance of A among those 20 is decreasing in the total number of listings in the market (as long as the distribution of A is log-concave).[49] So the value to paying attention to GRP as well as attractiveness is increasing in the number of listings in a market.

In a large market, if the top 20 listings have about the same attractiveness, then the value of re-ordering those listings would depend just on the variance of GRP and the amount that booking probability changes with ranking. If GRP is normally distributed, the top 10 values will, on average, be 0.77 standard deviations above the mean (and the bottom half will be 0.77 s.d. below). Ursu (2016) find that Expedia results ranked 1-10 are about 4 times as likely to be booked as those ranked 11-20. If on average listings in the top half are four times as likely to be booked, the optimal order would result in an additional $.77 \cdot (4-1)/(4+1) \cdot .34 \approx 0.16$ trips per booking.[50] If we use 2 or 5 times as likely, instead of 4, we get 0.09 or 0.17 trips, respectively.

[47] This manipulation of search results is in the interest of guests, helping them book the unobservably higher quality listings. On the host side it will help some hosts over others, but it improves efficiency in the sense that it puts hosts' individual incentives more inline with the social surplus.

[48] If the platform were just choosing the maximum A, then in expectation it would get

$$E[b(A_l, GRP_l)] = E[\tilde{A}](1 + \mu_{GRP}) + \rho \cdot var[\tilde{A}]$$

where $\mu_G RP$ is the mean of GRP_l, unconditional on A_l and ρ is the correlation between A_l and GRP_l.

[49] We can think of the top 20 as being drawn from a truncated distribution with a truncation point at the 21^{st} best. As long as the distribution of A is log-concave (as is the case for most common distributions), the variance of the truncated distribution is decreasing in the truncation point. Increasing the total number of listings leads to a first-order stochastically dominate distribution of the $21^s t$ best draw, so the variance of the distribution of the top 20 is necessarily decreasing.

[50] If only the top 10 listings had equivalent A, we care about the top 5 verse the next 5. Ursu (2016) find the top 5 are about 2.5 times as likely to be booked as the next five; the expected GRP of the top 5 is 0.74 standard deviations above the mean, so the gain would be $.77 \cdot (2.5-1)/(2.5+1) \cdot 0.34 \approx 0.11$ trips.

The average searcher in our sample took 2 trips subsequent to their booked search. The gain to reordering their search results would depend on the variance in booking probability and its correlation with GRP. For a large market, our ballpark calculations suggest incorporating the guest return propensity into search ranking could generate an increase in return trips on the order of 4.5-8.5%.

Informing guests

In addition to or instead of altering the search results based on GRP, the platform could report GRP to users, the way it does with ratings. However, it might be hard to explain to guests what GRP is measuring. The platform could show guests the GRP as an undefined 'platform quality' measure, but that might frustrate hosts who feel it is out of their control. The risk with presenting the information – in whatever form – is that doing so will change guests' expectations. To the extent that what is important for returns is how good a listing is *relative* to the guest's expectations, telling guests a listing's quality could have substantial unintended consequences.

Screening and Improving

Airbnb also has the option to remove low-quality hosts. The value of this will depend on the thickness of the platform-market and also on the extent and type of competition Airbnb has in the broader market. In our model of consumer learning, we assumed that the value of the outside option was known. If instead, the outside option is (partially) other peer-to-peer accommodation platforms, guests may update their priors not just about the quality of Airbnb, but about the quality of all such platforms. This changes Airbnb's incentives. It means they want hosts on all platforms to be high quality, so that people return to the peer-to-peer accommodation market as a whole. If Airbnb kicks off a bad host, that host may register on another platform.[51] That is still not good for Airbnb. So if a platform thinks that users are learning about the industry as a whole, they have an incentive to try to improve listing quality, rather than just remove low-quality listings.

Conversely, if the market as a whole is more mature or well-known, Airbnb may think that users are only learning about the quality of sellers on its platform. In that case, it has much more of an incentive to remove bad hosts. Not only is it potentially easier to screen hosts than improve them, but if a bad host went to another platform, that could be good for Airbnb. Diverting low-quality hosts to another platform increases the probability that guests conclude that the other platform is low quality, and switch to Airbnb.

[51] If hosts are multi-homing (selling through multiple platforms), then a host kicked of Airbnb could increase its number of transactions on another platform.

5 Conclusion

We propose measuring sellers' quality on a platform by the externality they impose on other sellers on the platform – how many times consumers who purchase from a given seller return to the platform. *Guest return propensity* is a 'revealed preference' measure of quality, based on what consumers do after after purchasing from a seller, rather than what they say about the seller. For Airbnb listings with at least 20 guests prior to 2016, the raw standard deviation across listings in the number of times their guests return is about 1.9; the average is 3.4. After accounting for observable guest and trip characteristics and adjusting for small samples, the estimated standard deviation of GRP is 0.55 trips. Our IV estimates, using variation in users' search results, suggest that a standard deviation increase in a listing's GRP causes a an increase in a guest's subsequent trips of 0.34 (17%).

We also highlight an important aspect of consumer heterogeneity. Even though we believe inexperienced buyers learn more about the platform, in many contexts, experienced buyers will be frequent buyers. The lower probability of an experienced guest leaving the platform due to one bad experience may be outweighed by the much higher cost to the platform if that guest leaves. This is analogous to the airline reward programs. Though experienced frequent flyers are probably less likely to switch airlines because of one bad flight, if they do, the revenue loss is much greater than if an infrequent consumer switches.

The gains to other platforms from incorporating this quality measure will vary. The value of screening bad sellers will depend on the broader market that the platform operates in. When quality is very observable ex-ante, we would expect our measure of quality to not vary much and potentially be correlated with the measures the platform already uses for evaluation – such as propensity to purchase. The externalities are smaller in markets where consumers have strong priors or believe there to be a lot of variance across sellers, because buyers will change their beliefs less based on a given experience. By contrast, in thick markets, with lots of good options, platforms may be able to nudge users towards sellers who are higher-quality by our metric, and not particularly lower quality by other metrics. The importance of potential return business will also matter;[52] the more important return users are to a platform's business model, the more weight should be given to our measure of quality relative to purchase propensity.

[52] A platform for wedding photographers, for example, probably does not have a lot of return business; while there still may be an externality in terms of whether people recommend it to their friends, it cannot be captured with our measure. Frequent consumers who are not myopic will value learning about a platform, so their threshold for not returning may be lower, but the cost to the platform of a seller pushing them across that threshold is higher.

Appendix: Supplementary Tables

Table A1: Effect of listing GRP and rating on subsequent spending, by guest experience

	All	All	1st	2nd	3rd	4+
Quality Booked	346.285***	34.832***	21.030***	16.154***	19.720***	49.229***
	(0.515)	(0.710)	(0.998)	(1.368)	(1.754)	(1.300)
Rating at booking		43.736***	36.018***	35.982***	40.270***	54.824***
		(1.406)	(1.945)	(2.660)	(3.425)	(2.634)
All Controls	0	1	1	1	1	1
Observations	11,499,195	7,710,665	2,016,255	1,384,152	989,703	3,320,555
R^2	−0.152	0.164	0.102	0.115	0.130	0.177

*p<0.1; **p<0.05; ***p<0.01

Note: This tables uses winsorized spending and replicates the analysis from Table 6 of the effect of the GRP and rating of a listing on the dollars a guest spent on Airbnb after staying there. The first column looks only at the effect of GRP. The other columns add the listing rating and full set of controls. Columns (3)-(7) run the analysis separately for guests for whom the trip was their 1st, 2nd, 3rd or 4th or more trip on Airbnb.

Table A2: Effect of GRP by price residual quartile on price of next booking and trips taken

	Trips	Trips	Price	Price
Nightly Price	−0.001***	−0.0004***	0.249***	0.281***
	(0.00002)	(0.00003)	(0.001)	(0.001)
Quality x Price Q1	0.092***	0.197***	1.031***	0.146
	(0.004)	(0.006)	(0.146)	(0.166)
Quality x Price Q2	0.030***	0.120***	−0.252	−0.427**
	(0.005)	(0.006)	(0.166)	(0.186)
Quality x Price Q3	0.020***	0.107***	−0.977***	−0.470**
	(0.005)	(0.006)	(0.172)	(0.192)
Quality x Price Q4	0.013***	0.081***	−2.321***	−0.775***
	(0.005)	(0.006)	(0.179)	(0.200)
All Controls	0	1	0	1
Observations	11,498,914	7,710,559	5,107,111	4,227,932
R^2	0.007	0.247	0.138	0.150

*p<0.1; **p<0.05; ***p<0.01

Note: This tables show the effect of the GRP of a listing on the price residual of the next listing booked by a guest and the number of subsequent trips that guest takes. (Only including initial trips where the subsequent trip was booked after the check-in date of the initial trip.) The effect is shown separately by the quartile of the price residual of the initial booking. The first and third columns looks only at the effect of GRP. The second and fourth columns add guest, trip and market controls as well as the rating of the initial listing at the time the trip was booked.

Table A3: IV:Effect of GRP and rating booked on winsorized subsequent spending

	By Trip Number				
	All	1st	2nd	3rd	4+
Quality booked	99.376***	19.945	64.081	79.651	285.970***
	(12.549)	(14.791)	(39.797)	(67.966)	(51.412)
Rating booked	396.961***	168.318***	245.938***	580.355***	786.608***
	(27.224)	(30.723)	(82.837)	(150.923)	(111.134)
Avg price of results	0.462***	0.244***	0.450***	0.745***	0.899***
	(0.012)	(0.014)	(0.035)	(0.064)	(0.044)
F-Stat	32741.7	13884.8	5237.7	3500	7347
Observations	3,869,881	1,626,456	729,047	429,669	1,084,709
R^2	0.520	0.604	0.744	0.821	0.693
Adjusted R^2	0.075	−0.040	−0.068	−0.072	−0.043

*p<0.1; **p<0.05; ***p<0.01

Note: This table replicates the analysis of the effect of listing characteristics on a guest's subsequent spending (in dollars) on the platform from Table 12, with a winsorized measure of subsequent spending. Each column is a subgroup of guests based on whether it was their first, second, third or fourth (or more) trip. All regressions control for market x check-in-date x how far in advance the search happened, the search date, the hour of the search interacted with the timezone the searcher was searching from and any search filters the searcher used.

Table A4: Effect of listing characteristics on rating and price of next trip booked (IV)

	Rating of Next Booking	Price of Next Booking
Quality booked	0.006	1.046
	(0.007)	(1.244)
Rating booked	0.209***	38.163***
	(0.014)	(2.673)
Avg price of results	0.00001**	0.021***
	(0.00000)	(0.001)
F-Stat	14595.1	16572.3
Observations	1,918,582	2,186,457
R^2	0.605	0.546
Adjusted R^2	0.010	−0.075

*p<0.1; **p<0.05; ***p<0.01

Note: This table uses the same IV approach as Table 11 to show the effect of listing characteristics on the price and rating of a guest's next booking. All regressions control for market x check-in-date x how far in advance the search happened, the search date, the hour of the search interacted with the timezone the searcher was searching from and any search filters the searcher used.

References

Armstrong, M. (2006). Competition in two-sided markets. *The RAND Journal of Economics*, 37(3):668–691.

Bardey, D., Cremer, H., and Lozachmeur, J.-M. (2009). Competition in two-sided markets with common network externalities.

Barnett, M. L. (2007). Tarred and untarred by the same brush: Exploring interdependence in the volatility of stock returns. *Corporate Reputation Review*, 10(1):3–21.

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

Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10):1407–1424.

Evans, D. S. (2003). The antitrust economics of multi-sided platform markets. *Yale Journal on Regulation*, 20:325.

Fradkin, A., Grewal, E., Holtz, D., and Pearson, M. (2015). Bias and reciprocity in online reviews: Evidence from field experiments on airbnb. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 641–641. ACM.

Hagiu, A. (2007). Merchant or two-sided platform? *Review of Network Economics*, 6(2).

King, A., Lenox, M. J., and Barnett, M. L. (2002). Strategic responses to the reputation commons problem. *Organizations, Policy and the Natural Environment: Institutional and Strategic Perspectives*, pages 393–406.

Landon, S. and Smith, C. E. (1998). Quality expectations, reputation, and price. *Southern Economic Journal*, pages 628–647.

Luca, M. (2017). Designing online marketplaces: Trust and reputation mechanisms. *Innovation Policy and the Economy*, 17(1):77–93.

Mayzlin, D., Dover, Y., and Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *The American Economic Review*, 104(8):2421–2455.

Morris, C. N. (1983). Parametric empirical Bayes inference: Theory and applications. *Journal of the American Statistical Association*, 78(381):47–55.

Nosko, C. and Tadelis, S. (2015). The limits of reputation in platform markets: An empirical analysis and field experiment.

Rochet, J.-C. and Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association*, 1(4):990–1029.

Rochet, J.-C. and Tirole, J. (2006). Two-sided markets: a progress report. *The RAND Journal of Economics*, 37(3):645–667.

Rysman, M. (2009). The economics of two-sided markets. *The Journal of Economic Perspectives*, 23(3):125–143.

Shapiro, C. (1982). Consumer information, product quality, and seller reputation. *The Bell Journal of Economics*, pages 20–35.

Ursu, R. M. (2016). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Browser Download This Paper*.

Weyl, E. G. (2010). A price theory of multi-sided platforms. *The American Economic Review*, 100(4):1642–1672.

Winfree, J. A. and McCluskey, J. J. (2005). Collective reputation and quality. *American Journal of Agricultural Economics*, 87(1):206–213.