informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE

Vol. 00, No. 0, Xxxxx 0000, pp. 000–000

Submitted to *Management Science*

ISSN 0025-1909, EISSN 1526-5501

Quality Externalities on Platforms: The Case of Airbnb

Peter Coles
Airbnb, peter.coles@airbnb.com

Sonia Jaffe
Microsoft, sonia.jaffe@microsoft.com

Steven Levitt
University of Chicago, slevitt@uchicago.edu

Igor Popov
Airbnb, igor.popov@airbnb.com

Flavio Stanchi
Airbnb, flavio.stanchi@airbnb.com

Bradley Turnbull
Pinterest

bturnbull@pinterest.com

Abstract.

Platforms need a good measure of seller quality in order to screen out low-quality sellers and incentivize high quality. This is particularly important when buyers on a platform have limited information and are still learning about average seller quality on the platform. In this case, the quality of a seller in any given transaction may affect whether or how frequently a buyer returns to the platform, even if that buyer's next transaction is with someone else – an *externality* on other sellers. We propose an intuitive measure of this quality externality, applicable across a range of platforms. *Guest Return Propensity* (GRP) is the aggregate propensity of a seller's customers to return to the platform after a transaction (controlling for customer characteristics). This metric differs from more traditional metrics, like star ratings, in that it is entirely based on revealed preference – future actions buyers take. We validate the performance of this metric using data from Airbnb, a home stays and experiences platform. Using an instrumental variable causal inference approach to account for unobservable guest characteristics, we find that matching customers to listings with a one standard deviation higher GRP causes them to take 29% more subsequent trips. We also provide marketplace recommendations on ways to increase overall platform surplus by matching buyers to higher-GRP sellers.

Key words: Quality, Marketplace, Platform, Ratings, Externality, e-commerce

1. Introduction

Platforms' success depends on their ability to match users and align incentives. Online marketplaces often 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. Some platforms use purchase propensity as a proxy for quality; other platforms have user-generated ratings, which vary in their informativeness (whether the ratings differ meaningfully 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 full value a high-quality 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 seller that provides a good (bad) experience makes it more (less) likely that the buyer returns to the platform. Those subsequent transactions help other sellers, but are not internalized by the initial seller.[1] A platform, however, 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 seller's *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. 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 sellers.

We then present the empirical context. We describe our data from Airbnb, which coordinates millions of trips in numerous countries and regions. 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. Controlling for guest and trip characteristics, we find significant variation in GRP across listings. Even among very highly reviewed listings, there is variation in how much their guests return to the platform. In fact, GRP is only weakly correlated with guests' rating of that listing or with the number of guests the listing has hosted.[2] Moreover, the GRP measure is persistent, predicting return trips 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. However, a change in beliefs will have a larger effect on travel volumes for frequent travelers. In our empirical heterogeneity analysis we find suggestive evidence that this second effect dominates. We hypothesize that this reflects that experienced guests

[1] Buyers can create an analogous externality on other buyers by affecting whether sellers return to the platform. For data reasons – there are not a lot of buyers who transact with enough sellers to estimate their effect – we focus on seller quality, but the same logic carries through for sellers learning about buyer quality.

[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 affected 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.

are more likely to be frequent travelers so, though the relative effect on beliefs is smaller, their beliefs influence more potential future trips and therefore more future spending on the platform. Thus, from the platform's point of view, experienced guests are attractive targets for high-GRP listings, even if their beliefs are less affected by a listing's GRP.

To gauge whether a listing's GRP has a causal effect on guests' subsequent trips, we use an instrumental variable strategy. Frequent bookings, changes in availability, and ongoing search experiments often result in two users who visit the site on the same day and enter the same search terms being shown different inventory. The size of our sample allows us to leverage these quickly changing choice sets for identification. First, we calculate the GRP for each listing based on trips to them from 2011-2017. To evaluate the impact of a guest engaging with these listings, we then instrument for the GRP of the listing booked for a trip in 2018 with the average GRP across the first page of search results, using only variation generated by multiple guests searching for listings in the same market, for the same travel day, with comparable trip lead time. This approach should enable us to learn whether a listing's GRP has a causal effect on guests' subsequent trips, rather than the correlation just reflecting unobserved guest characteristics.

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 0.60 additional future trips. This effect is over twice as large as 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 purchase probabilities by about 1% (.00038 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 for managers of other platforms. Airbnb could use our measure of listing quality to raise overall seller surplus by removing low-quality listings from the platform, or by directly trying to help low-quality listings improve. Alternatively, it could incentivize higher listing quality and redirect guests to higher-quality listings either by implementing a Pigouvian tax on low-quality listings or by highlighting higher-quality listings in guests' search results.

The remainder of the paper is organized as follows. Section 2 discusses the related literature. Section 3 outlines the theory, discussing both consumer learning and the externality it generates.

Section 4 describes the Airbnb context and data we use in the analysis. Here, we show how we construct the GRP metric for each listing. Section 5 presents the empirical results, both the reduced-form analysis and the instrumental-variable approach using the search data. In Section 6, we discuss potential strategies the platform could use to raise the average GRP of the listings that guests visit. Section 7 concludes.

2. Related Literature

This paper builds on the literature exploring quality and ratings metrics used by online marketplaces.[3] To our knowledge, Nosko and Tadelis (2015) were the first to highlight the role of quality-driven own-side platform externalities, and this paper was a key inspiration for our own. 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 buyer return rate to eBay to validate this measure of quality. We build upon their work in one key way: Rather than using return rates as a validation for a rating-based metric, the residual return rates are themselves our central metric, so that the externality is captured in our quality concept directly. We tested for correlation with a feedback propensity measure akin to Nosko and Tadelis, and found this to be positive but minimal.[4] Notably, GRP has the advantage that it can be calculated for any platform that involves interactions between buyers and sellers, regardless of rating mechanism, or even if there is no rating mechanism at all. In short, we see GRP as a useful, platform-agnostic, intuitive, and 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.[5] 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. There is a vast

[3] (e.g., Dellarocas 2003, Cabral and Hortacsu 2010, Mayzlin et al. 2014, Fradkin et al. 2015, Luca 2017).

[4] While a "positive review" at eBay is well-defined by the nature of the rating system, we can't perfectly replicated the concept for Airbnb because ratings are not binary. Instead, we construct a similar measure of the number of 5 star reviews divided by total bookings. The correlation of GRP with this version of the Nosko and Tadelis "effective percent positive" metric is .040.

[5] (e.g., Rochet and Tirole 2003, 2006, Evans 2003, Armstrong 2006, Hagiu 2007, Rysman 2009, Weyl 2010, Bardey et al. 2014)

literature on consumers learning about product quality.[6] Shapiro (1982) models consumer learning and firm quality investments in equilibrium. Taking these concepts to an empirical setting, Landon and Smith (1998) estimate the role 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.[7] We apply these principles to the platform setting to understand how intermediaries can mitigate these commons problems. The problem from the platform's perspective also relates to the vast literature firms wanting to measure and maximize consumer lifetime value

3. Theory

We start with a basic model of consumers' behavior and learning, and then highlight how the platform objective function compares to the payoffs of individual sellers. We present only the findings here, with proofs and underlying math in Appendix A.

3.1. 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 start with a prior over the distribution of the good's quality. Each period, a consumer may have an opportunity to potentially consume the good; that opportunity has a value draw corresponding to how much the consumer values a zero-quality good in that period.[8] If the consumer purchases the good, she gets a quality draw and her payoff for that period is the value of a zero-quality good plus a concave function of the quality. The consumer then updates her beliefs about the distribution of quality in the market. The next period if she has an opportunity to purchase, her decision depends on those updated beliefs. Given a consumer's beliefs, she will have a cutoff for the minimum value for a zero-quality opportunity such that she will purchase that period.

[6] (e.g. Kim 2020, Ching et al. 2013, and citations therein)

[7] (see also, Winfree and McCluskey 2005, Barnett 2007)

[8] Alternatively, the draw can be thought of as the observable component of quality of the available seller, where observable and unobservable quality are uncorrelated.

LEMMA 1. *For myopic consumers,*[9] *the more confidence consumers have in their beliefs or the more variation there is across sellers in quality, the less an individual's beliefs about the average quality will be influenced by a given quality draw.*

Lemma 1 suggests that quality externalities may be smaller on a platform like Etsy[10] – 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.

LEMMA 2. *When a consumer has more past quality draws (more experience with the market) then*

1. *The mean of the consumer's beliefs is less affected by an additional draw.*
2. *The consumer's threshold cutoff for the value of an opportunity where they chose to purchase is less sensitive to a quality draw (as long as the cutoff is decreasing in the quality draw).*[11]

Though a given individual's beliefs will be more affected 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. The effect on an individual's probability of purchase – as distinct from beliefs – depends on both the effect on their beliefs and on the probability that they have an opportunity to purchase. Therefore, the effect of quality on an experienced traveler's purchase probability could be larger than for inexperienced consumers, even if the effect on their beliefs is smaller.

3.2. 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

[9] If consumers are forward-looking, they take into account the value of the of what they learn about the quality distribution from making a purchase. For a given set of beliefs, their threshold cutoff for the value of a zero-quality opportunity will be lower because they also value the information received from purchasing. Overtime as a consumer learns more about the market, the value of the information will decrease and a forward looking consumer will behave more like a myopic one.

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

[11] While this is the case in general, it is possible, that for quality draws way above the prior, the increase in posterior variance from a higher q draw can push the threshold up more than the increase in the posterior mean pushes it down.

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.[12] 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.

PROPOSITION 1. *A seller's private value of improving quality in period 1 is the social value of doing so times the seller's market share.*

If either the platform has a high discount rate 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 proportional to the change in the probability that the buyer returns in the subsequent period. At the other extreme, if the platform is very patient, the value to the platform of a seller's quality is equal to the effect on the number of subsequent trips the buyer takes. 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.[13]

4. 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, an 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; as of July 2019 there were over six million active Airbnb listings[14] in over 220 countries and regions. Figure 1 shows the geographical distribution of Airbnb listings as of July 31, 2019. Airbnb data also allow us to observe individual buyer and seller actions 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

[12] 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.

[13] 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 we think this simplification is reasonable.

[14] This is the number of listings that were viewable on the site.

accommodation quality is heterogeneous and not well-observed ex-ante. Airbnb could incorporate quality information into its marketplace design, making this a relevant setting for platform policy with respect to quality.

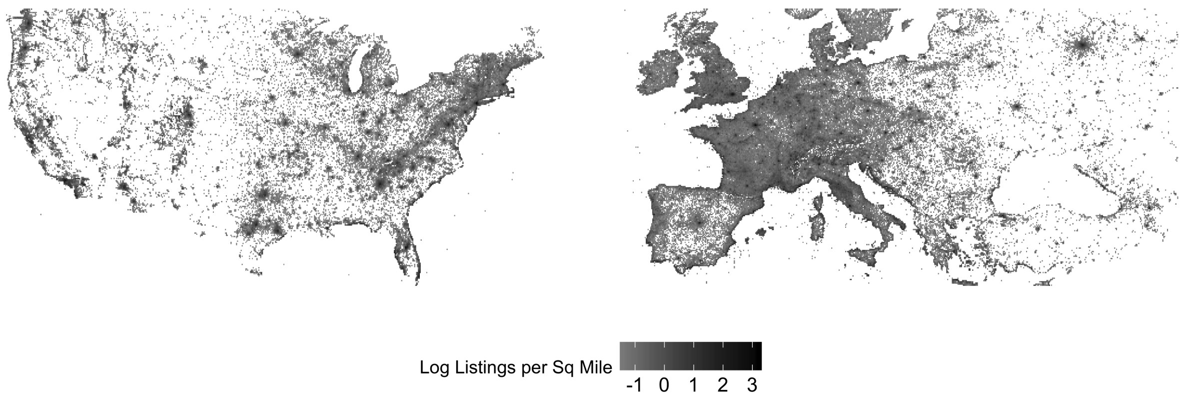


Figure 1 Log density of listings per square mile, as of July 31, 2019.

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 longitude by .1 latitude, using the approximation of 53 miles per degree latitude. Source: Airbnb data.

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 for all trips between January 1, 2011 and July 31, 2019 are in Table 1. We categorize listings into markets, which can include multiple, adjacent cities. The average market has roughly 1.3K listings that have hosted a trip and hosted over 40K trips in this timeframe. The average rating of listings as of January 1, 2018 was 4.59. For users who have taken a trip, the average number of trips taken is 2.7; about 55% are female and the average age is 37, where gender and age are based on self-reported data.

We are particularly interested in whether guests return to Airbnb after taking a trip. Figure 2 shows the fraction of guests that return within 400 days by trip number, for trips that occurred at 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 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 indicated via a search filter. We also drop subsequent searches for the same user and market with a check-in

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

Statistic	Mean	St. Dev.	Min	Median	Max
Markets					
Listings	1,332	3,915	1	221	85,500
Trips	43,017	141,515	1	4,869.5	2,599,220
Listings					
Rating	4.59	0.58	1.00	4.77	5.00
Trips	23	47	1	6	3,264
Guests					
Age	36.69	12.76	18	33	103
Female (0/1)	0.55	0.50	0	1	1
Trips	2.72	3.75	1	1	1,039
Trips w/in 400	0.84	1.83	0	0	693
Trips					
Number of Guests	2.6	1.9	1	2	150

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, 2018.

date within two days, since these are in some sense not the user's first search.[15] We are left with approximately 9.3 million searches per month in 2018. 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 80% of searchers are not people the site recognizes as having previously booked a trip.[16] On average they search for trips 86 days before the searched-for check-in date. The average rating of search results is slightly higher than the average across listings.

[15] Dropping bots reduces noise since those never result in a booking. Focusing on the first search allows us to ignore how results from the initial search may influence filters used or whether a person makes a subsequent search.

[16] 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.

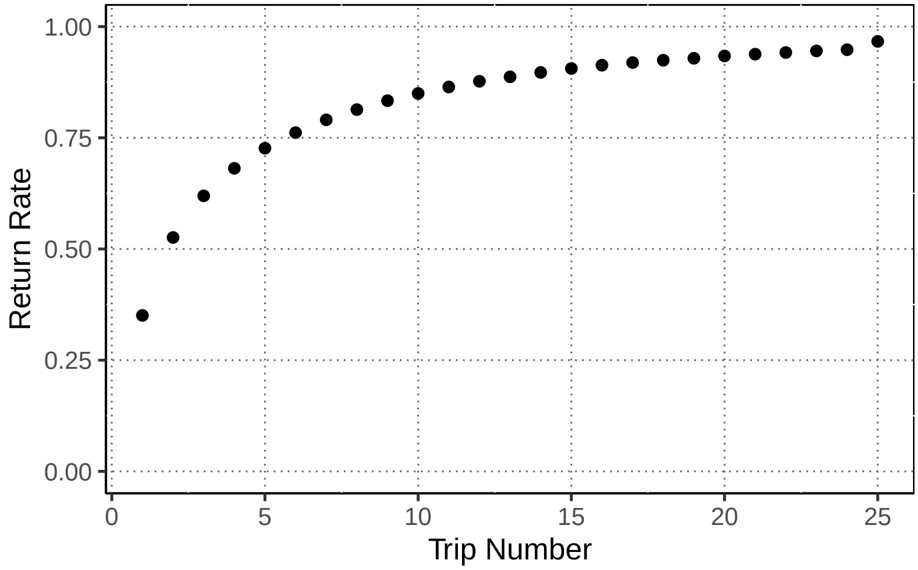


Figure 2 Fraction of guests that return to Airbnb markets

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.

The construction of the guest return propensity (GRP) variable is explained below. We only have GRP and rating measures for a subset of listings, so for listings without a measure we use the average of that measure across all listings.

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 identify take an average of 2.0 trips following the search, spending \$920 dollars. If, within 6 days of that initial search, the 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 1.4% of searches result in a booking. For searches by identified users (the only ones that could potentially result in a booking), the booking rate is 3.5%.

Table 2b summarizes some variables that are only available for the booked searches. The users whose searches we link to a booking spend an average of \$448 on the linked trip. After the linked trip the average guest books and takes 2.3 trips, spending \$815.

Table 2 Summary search variables

(a) Searches (Jan 2018)

Statistic	Mean	St. Dev.	Min	Median	Max	N
New Traveler	0.80	0.40	0	1	1	51,374,488
Days Advance Searched	86	82	0	59	730	51,373,947
Num Nights	7	18	1	3	731	51,365,379
Num Guests	3	3	0	2	33	51,374,488
Entire Only	0.06	0.23	0	0	1	51,374,488
Search Results						
Number of Listings	3,565	8,914	0	651	119,507	51,374,482
Avg Price	220	319	0	121	2,639	51,267,979
Avg Jan 1 Rating	4.745	0.107	1.000	4.764	5.000	51,369,798
Avg GRP shown	0.138	0.185	−3.550	0.148	3.557	51,374,488
Outcomes						
Booking (0/1)	0.014	0.117	0	0	1	51,374,488
Trips After	1.970	4.073	0	1	787	20,406,912
Spent After	920	2,623	0	189	879,849	20,406,912

(b) Booked Searches (Jan-June 2018)

Statistic	Mean	St. Dev.	Min	Median	Max	N
GRP booked	0.149	0.701	−3.550	0.215	3.557	2,487,030
Rating booked	4.743	0.272	1.000	4.812	5.000	3,505,380
Total price booked	448	656	1	265	72,945	3,890,084
Trips After Check-in	2.335	5.170	0	1	689	3,870,893
Spent After Check-in	815	2,822	0	175	875,338	3,870,893

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.

4.1. Constructing a Quality Measure

Using the data described, we construct a quality measure for each listing.[17] As discussed in Section 3.2 and A.2, we measure listing quality as the *Guest Return Propensity*: the number of trips a guest takes after staying with that host (controlling for guest characteristics). For each trip taken prior to 2018 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

$$y_{ilt} = \alpha_0 + \alpha_{m(l)} + \alpha_1 X_i + \alpha_2 Z_{it} + \varepsilon_{ilt},$$

where $\alpha_{m(l)}$ is the market the listing is in, X_i is a vector of guest characteristics – their gender and age bin[18] – and Z_{it} is a vector of characteristics of the trip taken by guest i in period t – the date, number of guests, the (binned) number of nights, whether the guest was verified, whether the trip was their 1^{st}, 2^{nd}, 3^{rd}, or 4^{th}+ trip. For each listing we then calculate a *listing effect*

$$\begin{aligned} f_l =& \frac{1}{n_l} \sum_{i,t} \varepsilon_{ilt} \\ =& \frac{1}{n_l} \sum_{i,t} \left(y_{ilt} - (\hat{\alpha}_0 + \hat{\alpha}_{m(l)} + \hat{\alpha}_1 X_i + \hat{\alpha}_2 Z_{it}) \right), \end{aligned}$$

where n_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. 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 to be correlated with the effect being estimated.[19] 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

[17] 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.

[18] 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.

[19] This seems potentially relevant for teacher quality – a common application of shrinkage estimators in economics (e.g. Chetty et al. 2014).

number of trips they have and the time (quarter) that they were first created.[20] 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

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

We also winsorize the adjusted listing effects at 3 standard deviations from the mean to get our final measure of GRP. Lastly, for some of our analysis we use only listings with at least 20 guest-trips on which to base the estimation.

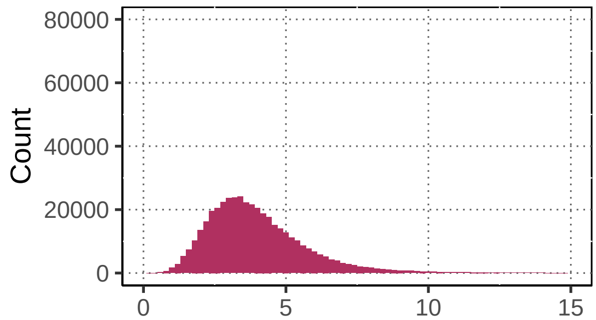
Figure 3a shows the distribution of the raw averages of the number of times a listing's guests return. It is limited to the 428k listings that had at least 20 trips prior to 2018. The average across listings is 4.1 return trips.[21] Figure 3b and 3c show the listing effects after adding guest controls or guest and trip controls. Figure 3d shows the residualized, shrunk, and winsorized listing effects, as described above. The standard deviation of the raw effects is 2.23; after all controls it is 1.71; and for the final effects it is 0.66. 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 moderately correlated with the unshrunk estimates and the raw averages. The difference between the raw average and the unshrunk estimates is due to the guest and trip-level controls.[22] However, GRP's correlation with other characteristics of the listings is very low. The

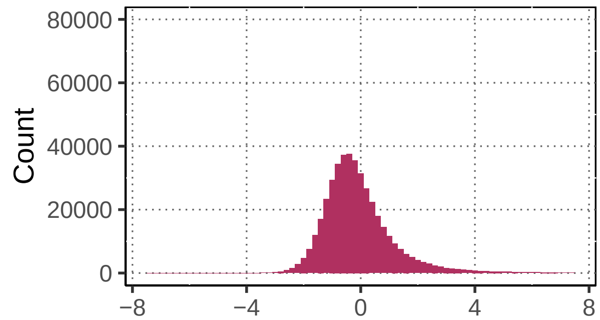
[20] If two listings have the same number of guests, but one has been around twice as long, the newer one is clearly the more attractive listing. Rather than trying to quantify this trade off, we group listings by age interacted with the number of guests.

[21] The average across trips is 4.3 subsequent trips. Unlike the entire sample summarized in Table 1, these trips all happened prior to 2018, so the guests have had more time to accumulate additional trips.

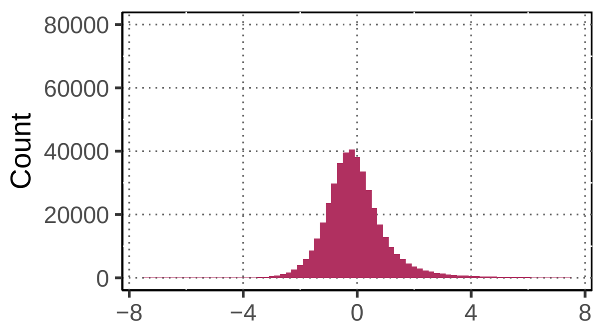
[22] When we estimate GRP using only time-fixed guest characteristics, we find that the correlation between this estimate and our final estimate of GRP with a full set of controls is .98. However, if we also drop the number of past trips a guest has, the correlation drops to .73, suggesting that past trips is the most important guest characteristic.



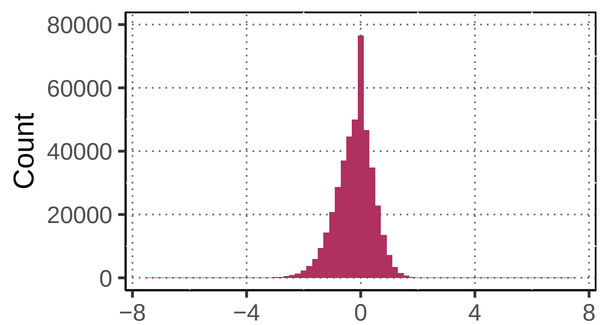
(a) Raw averages



(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 428k listings with at least 20 guests prior to 2018. The x-axis is the average number of time's a listing's guests return, the y-axis is the count of listings with that average. Figure (a) is the raw average in the data. 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 1349, 2226, and 1830 observations, respectively, that fall outside of the x-axis range. Figure (d) shows the listing effects (GRP) after shrinking towards a group mean to account for small samples, and winsorizing them at 3 standard deviations from the mean.

correlations with rating, price, number of guests, and a variety of other listing metrics are all below 0.05.[23] We do not know why the correlation with ratings is so low. It may be that ratings are based on factors guests think are specific to a given listing, whereas whether a guest returns is not affected

[23] We look at the correlation with the average guest rating the listing host gives because in theory hosts could influence a guest's return probability through the review the host leaves instead of the quality of the experience. The correlation is negligible at .0076.

Table 3 Correlation of Metrics across Listings

(a) Correlations among listing effects:

	Guest Return Propensity	Unshrunk	Raw Average
Guest Return Propensity	1	0.543	0.317
Unshrunk	0.543	1	0.800
Raw Average	0.317	0.800	1

(b) Correlation between GRP and other metrics:

Metric	Corr.	Metric	Corr.
Rating	0.0407	Occupancy Rate	0.046
Price	-0.0038	Host verified	-0.0059
Number of Guests	0.0286	Avg Guest Rating Given	0.0076
Number of Ratings	0.0364	Fraction of Guests Recommended	0.0301
#Ratings / #Guests	0.0406	Still active 1/1/2018	0.0184

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 428k listings with at least 20 guests prior to 2018. The bottom panel gives the correlation between the guest return propensity (GRP) and other listing characteristics. The price is the average price for October 2017 as of September 1, 2017. The occupancy rate is the number of nights the listing was booked prior to 1/1/2018 divided by the number of nights the listing was made available. The other variables are all measured as of 1/1/2018. 'Avg guest rating given' and 'fraction of guests recommended' refer to the rating that the host leaves for the guests at the listing and what fraction of them the host recommends as guests.

by things they think are listing-specific, but precisely by the factors they think generalize to all listings.[24]

Hosts on Airbnb can manage multiple listings. We compare the within host variance in listing GRP to the overall variance for listings managed by hosts with multiple listings. For listings with at least 20 guests, the ratio of within host variance in GRP to overall variance is about 71%. This is higher than we would have expected and suggests that GRP is more a characteristic of the listing,

[24] 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 GRP is no higher.

rather than the host; though not characteristics of the listing that are easily observable, as shown by the the low correlations in Table 3.

5. GRP as a Driver of Future Guest Returns

In the prior section, we constructed the GRP metric to measure the degree to which a listing's guests return to the platform. In this section, we first demonstrate that guests staying (in 2018 or later) at a listing with high GRP (calculated based on pre-2018 trips) predicts more future bookings for that guest. We then use an instrumental variables approach to demonstrate that this relationship is causal: a guest is more likely to return to the platform *because* of staying in a high GRP listing (compared to staying in a low GRP one).[25]

Importantly, this causal relationship is demonstrated out of sample: guest return propensities (GRP) are calculated using the post-trip behavior of each listing's trips from 2011 through 2017, and we examine whether GRP predicts subsequent guest behavior using trips starting in 2018.

5.1. Correlation of GRP with Future Bookings

In this subsection we explore how a guest booking a higher GRP listing is strongly predictive of that guest returning to the platform, focusing on trips where a 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).[26] 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 x days after staying at a listing, conditional on not having returned between the stay at that listing and that point in time. The probability of return is substantially higher for a few months for guests that stayed at a listing in the top GRP quartile; the difference persists for just under a year.

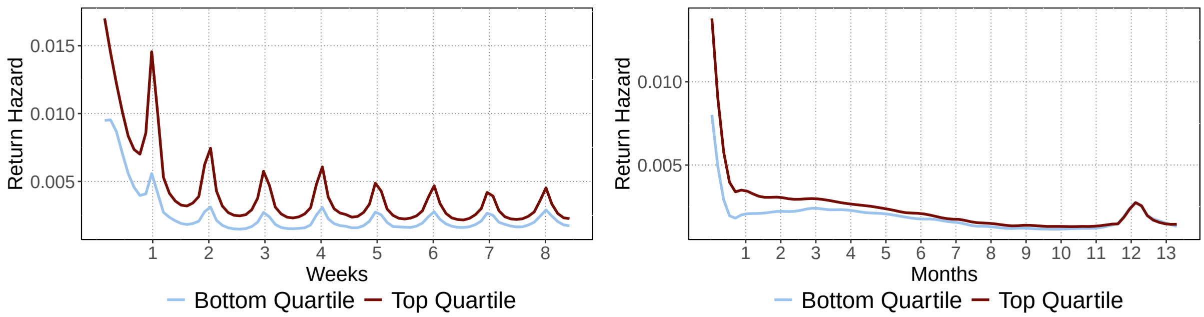
Table 4 shows the regression results analogous to Figure 4. The first column is the raw correlation and the later columns control for ratings and guest and trip characteristics.[27] These are the coefficients from Cox hazard regression; since guest return propensity is normalized to have a standard deviation across listings of 1,[28] the interpretation is that switching to a listing with a 1

[25] In theory, hosts could get a high GRP by using the two-sided nature of the platform to reject guests who they think will be unlikely to return (Fradkin 2018), but booking rejection rates are below 10% in the period we study.

[26] 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.

[27] The sample size drops because some trips do not have guest characteristics or are to unrated listings.

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



(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 2018, the returns are shown for guests starting in 2018. Because we cannot control for market in the graphs, we use a measure of GRP that does not control for market. The cyclicality in (a) is because guests tend to start trips on the same day of the week their previous trip started on.

standard deviation higher GRP corresponds to a guest being $\exp(.157) \approx 1.170$ 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.[29] Surprisingly, the increase in return probability from a higher GRP host is slightly larger for returning guests than for first time guests. Ratings also matter, but the association is much smaller; the standard deviation of the listing rating is .58 so a guest is $\exp(.056 * .58) \approx 1.033$ times as likely to return if they stay at a listing with a one standard deviation higher rating.

The platform cares not just about whether a guest returns, but how many times and how much money they spend. To examine this, for each trip we calculate how many trips the guest booked (and took) after that trip and how many dollars were spent on those subsequent trips. Tables 5 shows the regression results. On average, a 1 standard deviation higher GRP listing predicts guests taking 0.20 additional trips and spending \$43 additional dollars on Airbnb. The association tends to be larger for guests with more experience; because these guests travel more and spend more, even if their beliefs about Airbnb are less affected, small changes in their propensity to use it when traveling

[29] 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. An alternative would be to do a linear regression for the probability of return. Table A1 in the Appendix shows these results. The coefficients are smaller, suggesting a 3-5% increase in probability of return. We do not know if this is due to the poor fit of the linear model or the importance of the market controls. Both of these are addressed in the analyses of trips taken after, which we focus on below.

Table 4 Association between listing GRP and hazard of return

	All	All	1st	2nd	3rd	4+
GRP Booked	0.272***	0.157***	0.142***	0.141***	0.151***	0.164***
	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)	(0.001)
Rating at booking		0.056***	0.128***	0.086***	0.064***	0.019***
		(0.002)	(0.004)	(0.005)	(0.006)	(0.003)
Controls	0	1	1	1	1	1
Observations	18,443,232	17,869,298	5,823,153	2,948,578	1,931,525	7,166,042
R^2	0.006	0.159	0.021	0.020	0.022	0.089

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

Note: This tables show the coefficients from Cox Hazard regressions, analyzing the association of listing GRP with 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.

lead to larger changes in the number of trips and dollars spent on the platform.[30] The association with ratings is less clear. In the overall sample the coefficients negative though small. Since the standard deviation of rating across listings[31] is 0.58, a one standard deviation increase in the rating is associated with only .026 fewer trips and $3.8 fewer dollars.[32]

[30] Listing 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 role of listing 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. With information on referrals, the platform could adjust our metric to include effects on new guests.

[31] Like GRP, the standard deviation across trips is smaller than across all listings, it is 0.27.

[32] Because the variance in spending is so high, we repeat the analysis with winsorized spending. The results, shown in Appendix Table A2, are similar, though the coefficient on GRP is slightly smaller and that for ratings is somewhat larger and consistently positive. The association of GRP with spending is still highest for those with the most experience.

Table 5 Association between Listing GRP and Trips Taken and Dollars Spent after

	All	All	1st	2nd	3rd	4+
			Trips After			
GRP Booked	0.851***	0.200***	0.044***	0.070***	0.100***	0.334***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.003)	(0.003)
Rating at booking		−0.044***	0.049***	0.024***	−0.005	−0.122***
		(0.005)	(0.003)	(0.006)	(0.008)	(0.010)
R^2	0.001	0.164	0.061	0.086	0.101	0.159
			Dollars Spent After			
GRP Booked	9.050***	43.449***	9.808***	14.237***	21.467***	70.042***
	(0.773)	(0.809)	(0.554)	(0.950)	(1.342)	(1.678)
Rating at booking		−6.514***	33.052***	39.733***	25.866***	−51.680***
		(2.427)	(1.600)	(2.790)	(3.984)	(5.215)
R^2	0.000	0.075	0.041	0.063	0.075	0.073
Controls	0	1	1	1	1	1
Observations	23,616,685	22,786,307	6,530,211	3,584,997	2,462,306	10,208,793

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

Note: This table shows the association of the GRP and rating of a listing with the number of trips a guest takes (top half) and amount of money a guest spends (bottom half) subsequently. The first column looks only at the association with 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.

5.2. Using Searches as an Instrument

The evidence above suggests that listing quality differences as measured by GRP are meaningful, persistent, and associated with increased future bookings. But it is still possible that the difference is actually due to the guest side – certain guests could be (unobservably) more likely to return and take more trips and also tend to choose certain, higher GRP 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 choices for where to stay are 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, room type searched for, the time of day interacted with the timezone in which the search happened, and dropping searches that used other 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 6 shows the reduced-form association of the average GRP, rating, and price with trips taken subsequently. The first two columns of Table 6 show the reduced-form association of the average GRP and rating of the listings shown with the number of trips the searcher takes after *searching*. The first column is for all (first) searches; 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, 2018. When we limit to booked searches the sample size shrinks dramatically, so we use searches January – June 2018.[33] Column (3) uses the same sample of booked searches as Column (2) but looks at the association with the number of trips the searcher takes after *check-in* and controls for the number of past trips the guest had taken at that time. For the IV, where we are 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.[34] 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, so Column (3) of Table 6 implies that raising the average GRP of all the listings shown by one standard deviation of listing GRP results in 0.496 more trips. The effect of a one standard deviation increase in rating

[33] Since GRP is measured as of the end of 2017, the fraction of search results with a GRP measure available decreases over time, as the fraction of listings created since the beginning of January, 2018 increases. To the extent that quality changes over time, the GRP measure also becomes less accurate. For these reasons, we use only the first half of 2018.

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

Table 6 Association of average GRP and rating of search results with number of subsequent trips

	Trips After Search		Trips After check-in
	January Searches	Booked Searches	
Avg GRP shown	0.0001	0.494^{***}	0.496^{***}
	(0.0001)	(0.028)	(0.027)
Avg Jan 1 Rating	0.008^{***}	0.166^{***}	0.213^{***}
	(0.0002)	(0.047)	(0.044)
1 Past Trips	0.004^{***}	0.621^{***}	0.432^{***}
	(0.0001)	(0.013)	(0.012)
2 Past Trips	0.009^{***}	1.127^{***}	0.808^{***}
	(0.0001)	(0.015)	(0.014)
3+ Past Trips	0.013^{***}	3.559^{***}	2.784^{***}
	(0.0001)	(0.011)	(0.010)
Unknown Guest	-0.026^{***}		
	(0.0001)		
Observations	49,044,627	2,386,653	2,386,653
R^2	0.067	0.396	0.394
Adjusted R^2	0.031	0.119	0.116

*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 guests that made searches in January 2018. The remaining columns are limited to searches that resulted in a booking; to maintain a large enough sample we use searches in January - June 2018. Column (2) repeats the analysis of Column (1) of trips taken *after the search* for the IV-subsample. Column (3) shows 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 two 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 log of the average prices of listings shown, the hour of the search interacted with the timezone the searcher was searching from and any search filters the searcher used. Standard errors are clustered at the listing level.

is about a quarter as large ($.58 * .213 \approx .12$). The standard deviations in the average of the search results are smaller, 0.19 for GRP and .11 for ratings. So raising the average GRP (rating) shown by one standard deviation (of the average shown) only results in 0.09 (.02) additional trips. Across a variety of measures of subsequent trips, we see a strong relationship between the GRP of displayed search results and the number of subsequent trips the searcher takes.

Having seen the relationship between the GRP of search results and subsequent rips, we turn to the IV to see that the effect works via the effect on the GRP of the listing booked. The first two columns of Table 7 show the effect of the average GRP and rating shown on the GRP and rating booked (the first stage of the IV). The estimated effect of average GRP shown on average GRP booked is 0.82; for ratings the analogous effect is 0.42. The cross effects are negligible. It is expected that the effect of the listings shown on the listing booked is less than 1 because 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).

The last column of Table 7 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 0.599 additional trips by that guest. On average, in this sample, a guest takes 2.1 subsequent trips, so this is an increase of 29%.[35] The coefficient on the rating is somewhat smaller and recall that there is also less variance across listings in their rating so the coefficient of .48 implies that a one standard deviation increase in the rating of a listing leads to 0.28 additional trips by the guest. In Appendix Table A3, we do the same analysis for whether a guest returns to the platform instead of the number of times they return. The coefficient is smaller (.018 vs .599), partially because guests who return take multiple trips and partially because GRP was constructed based on number of subsequent trips guests took, not whether they returned to the platform.

At this point, it is reasonable to reflect on the identifying assumption for our IV regressions, namely 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[36] (2) the date the search

[35] The 0.599 coefficient is not directly comparable to the standard deviation of 0.66 trips that GRP had before we normalized it to 1, because that is based on trips in an earlier time window.

[36] 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.

Table 7 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			0.599^{***}
			(0.043)
Rating booked			0.484^{***}
			(0.136)
Avg GRP shown	0.817^{***}	0.007^{***}	
	(0.004)	(0.001)	
Avg Jan 1 Rating	0.017^{***}	0.416^{***}	
	(0.006)	(0.002)	
1 Past Trips	0.0004	0.003^{***}	0.431^{***}
	(0.002)	(0.001)	(0.009)
2 Past Trips	−0.003	0.004^{***}	0.808^{***}
	(0.002)	(0.001)	(0.011)
3+ Past Trips	0.0003	0.009^{***}	2.780^{***}
	(0.001)	(0.0004)	(0.011)
F-Stat	.	.	30814.4
Observations	2,384,401	2,384,401	2,384,401
R^2	0.369	0.400	0.391
Adjusted R^2	0.080	0.125	0.111

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

Note: The first two columns show the effect of search results on the GRP and rating of the listing booking, the first stage for the IV regression. 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. "Past Trips" refers to the number of trips the user took prior to the search. 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, the log of the average prices of listings shown, and the room type searched for and drops searches that use other filters. Standard errors are clustered at the listing level.

happened (3) the timezone the searcher was in interacted with the hour of day the search happened (4) the room-type searched for – and dropping searches that used other filters, the remaining variation in the GRP and rating of listings shown – from listings being added, listings being booked, and experiments in Airbnb's search algorithm[37] – is uncorrelated with guest characteristics. Are there any channels by which guest or search characteristics may be passed through, so that two otherwise identical searches systematically lead to substantively different GRP in search results? Examples might include the device of the user, or a user's prior booking history causing the search algorithm to show different listings. We know that Airbnb's search algorithm did not use GRP directly and the correlation between GRP and all the variables we have investigated is sufficiently low that we do not see how they could drive our results. Nevertheless, we acknowledge the limitation that our IV is not as bullet-proof as, say, a coin flip, and that these latent factors may theoretically influence the results, even if we believe them to be implausible.

In addition to affecting what listing a searcher books if the search results in a booking, the search results can also affect whether the search results in a booking. Table 8 looks at the effect of search results on this latter, extensive margin effect. The effects are fairly small: for known guests, a one standard deviation increase in the average GRP shown increases the probability of booking by $.002 * .19 \approx .00038$ percentage points (e.g. from 3.5% to 3.538%). Since the baseline booking rate for known guests is about 3.5%, this is an increase of 1%. The fact that these effects are non-zero means that for our IV analysis, we have a slightly selected sample, because those shown higher GRP listings were slightly more likely to book. However, the selection is very small and 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 IV results are a lower bound on the effect on an unselected sample – if we were able to remove those searches that only resulted in a booking because of the higher GRP shown, our estimates for the effect of GRP on subsequent trips would increase.

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 9 repeats the IV analysis from the third column of Table 6; the subsequent columns show the same regression separately for guests for whom this is their first, second, third, or

[37] 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.

Table 8 Effect of search results on whether a booking occurs

	All	Known Guests
Avg GRP shown	0.0001	0.002^{***}
	(0.0001)	(0.0003)
Avg Jan 1 Rating	0.008^{***}	0.015^{***}
	(0.0002)	(0.0005)
1 Past Trips	0.004^{***}	0.003^{***}
	(0.0001)	(0.0001)
2 Past Trips	0.009^{***}	0.008^{***}
	(0.0001)	(0.0002)
3+ Past Trips	0.013^{***}	0.011^{***}
	(0.0001)	(0.0002)
Unknown Guest	-0.026^{***}	
	(0.0001)	
Observations	49,044,627	18,136,876
R^2	0.067	0.108
Adjusted R^2	0.031	0.023

*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 for searches in January 2018. 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 column includes searches by unidentified searchers (none of which result in a booking); the second column is limited to identified searchers. For both, the excluded group is those with no past trips.

fourth (or more) trip. Given the large number of control variables, we are somewhat underpowered for analyzing effects in the smaller groups, but the results suggest that the effect of GRP is larger for guests with more experience on Airbnb: the higher frequency of travel outweighs the theoretically smaller learning effect.[38] Interestingly, the results suggest that the rating might have a significant effect for first time guests. A one standard deviation higher-rated listing causes them to take another .26 trips.

[38] In this sample a first-time guest returns an average of 0.9 times and guest on their fourth or higher trip returns an average of 4.2 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.

Table 9 Effect of GRP and rating booked on subsequent trips, by guest experience (IV)

	By Trip Number				
	All	1st	2nd	3rd	4+
GRP booked	0.599***	0.063	0.172	0.267	1.090***
	(0.043)	(0.045)	(0.105)	(0.200)	(0.116)
Rating booked	0.484***	0.451***	0.074	−0.214	0.990**
	(0.136)	(0.145)	(0.330)	(0.838)	(0.394)
F-Stat	30814.4	8535.5	3711.3	2005.6	13428.3
Observations	2,384,401	660,925	388,286	263,039	1,072,151
R^2	0.391	0.634	0.746	0.803	0.515
Adjusted R^2	0.111	0.039	0.046	0.036	0.021

*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 7, 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, the log of the average prices of listings shown, and the room type searched for and drops searches that use other filters.

What about Dollars? If the platform takes a percentage cut instead of a flat fee, it might care about the dollars spent by guests, not just the number of times they return. Table 10 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 much larger role, and the effect of GRP is significant, but smaller. Booking a listing with a one standard deviation higher GRP causes a guest to spend an additional \$84; a one standard deviation higher rating leads to $.58 * 1035 = \$600$ more spent subsequently.[39]

[39] A version of Table 10 using winsorized spending is Appendix Table A4; the results are similar.

Table 10 Effect of GRP and rating booked on subsequent spending, by guest experience (IV)

	By Trip Number				
	All	1st	2nd	3rd	4+
GRP booked	83.815***	−12.542	24.351	−28.894	141.871***
	(19.478)	(23.322)	(60.771)	(109.601)	(54.453)
Rating booked	1,034.847***	376.846***	496.377***	738.990**	1,550.767***
	(62.961)	(75.951)	(179.027)	(371.982)	(271.408)
F-Stat	73924.7	8535.5	3711.3	2005.6	13428.3
Observations	2,384,401	660,925	388,286	263,039	1,072,151
R^2	0.328	0.601	0.715	0.743	0.479
Adjusted R^2	0.019	−0.048	−0.070	−0.255	−0.051

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

Note: This table uses the same IV approach as Table 9 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, the log of the average prices of listings shown, and the room type searched for and drops searches that use other filters.

In some sense it is not surprising that GRP is not as good a predictor of subsequent spend because it was constructed based on the number of subsequent trips guests took after staying at a listing. If we wanted to focus on listings that caused guests to spend more money on the platform subsequently, we would instead construct a metric based on guests' dollars spent after staying at a listing. GRP is also uncorrelated with price, so people staying at higher GRP listings will not spend more. Ratings are slightly more correlated with price, but it is surprising the extent to which exogenously staying at a more highly rated listing seems to cause guests to book more expensive listings for subsequent trips.[40]

6. Implications for the Platform

Having seen that quality externalities drive user activity on Airbnb, we believe platform managers may use GRP as a new tool to improve matching outcomes and platform surplus. Specifically, there may be ample opportunities on other platforms to re-engage users more effectively by incorporating repurchasing propensity metrics into search and marketplace design. GRP adds to an existing toolkit of quality metrics that include average ratings, click through rate, and effective percent positive.

When will GRP be most useful to platforms? While ratings are a common measure of quality for consumers and platform managers, many platforms do not solicit ratings. GRP may be especially useful in this context. On long-term rental platforms, for example, renters do not review properties they rent, and on dating sites, users don't typically review the people they date. Re-engagement propensities can help these platforms understand which rentals and dates exert a strong externality on the platform in the absence of traditional star ratings.

Even in cases where other quality metrics are available, however, we show that direct measurement of quality externalities conveys new information. This new signal will be most informative for platforms that are (1) selling a relatively homogenous product (2) have heterogeneous quality and (3) have frequent transactions to learn from. The quantitative importance of quality externalities, and therefore the opportunities to leverage GRP, may be significantly greater or weaker on other marketplaces based on these characteristics.[41] A metric similar to GRP could also be used for content platforms where the platform may want to direct users to content creators that make it more likely the user will stay on the platform after consuming their content. Even as platforms get more sophisticated at measuring the aspects of quality that may be reflected in photos or natural language descriptions, there will likely still be unobservable aspects of quality that are usefully captured with a revealed-preference metric such as GRP. We believe GRP is especially relevant for platforms that need to understand the quality of its power users, which generate an outsized portion of platform revenue.

[40] In Appendix Table A5, we look at this directly by using the price and rating of the next listing booked subsequent to the trip of the booked search as the left-hand side variable, though this analysis is not a clean IV since the instruments affect the probability of booking again. 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 price of the initial listing also has a small positive effect on the price of the next booking. GRP actually has a small negative effect on the price of the subsequent booking.

[41] Airbnb listings have heterogeneous quality, but transactions, on average, are relatively infrequent. GRP can be even more precisely measured on marketplaces where users on both sides of the market tend to transact frequently after joining.

Next, we ask *how* a platform may use GRP to improve marketplace outcomes. We consider four ways the platform can raise listing quality or redirect guests to higher-quality listings.[42] 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, or to remove low-quality listings from the platform or to try to help them improve. The feasibility and relative efficacy of each of these approaches may very widely across online marketplaces, so we stress the importance of experimentation in the process of implementation.

Taxing the externality

The platform could reward hosts when their guests return within a certain time.[43] This would help high-GRP listings over low-GRP listings, and give all hosts an incentive to raise their GRP. How much such a tax or subsidy would cause host GRP to improve depends on how convex the costs to improving it 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 listing is unchanged, Pigouvian taxes can improve efficiency by affecting prices (and therefore quantities). A revenue-neutral tax/subsidy for GRP would cause low-GRP listings to increase prices and high-GRP listings to decrease prices, shifting demand to listings with better externalities.[44]

The Appendix shows that in a simple model of Pigouvian tax with no changes to GRP, the semi-elasticity of surplus with respect to the tax rate is approximately the own-cost pass-through rate times the variance in GRP

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

We see from Equation (1) that the benefits of a Pigouvian tax (as a share of producer surplus) will tend to be larger when there is more variation across listings in GRP and when pass-through is

[42] 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.

[43] As an example, a practical implementation may involve giving the host part of the payment at the time of the trip and the rest only if their guest returns to the platform.

[44] If there are fixed costs to being on the platform, a tax on low-GRP listings could also cause them to leave the platform. This is related to the effects of screening discussed below.

higher, since the more listings pass the cost change through to prices, the more the tax will affect prices and shift demand. For a back-of-the-envelope calculation, Equation (1) and our IV estimate of 0.599 for the effect of a standard deviation increase in GRP on guests' subsequent trips imply that if a listings' pass-through rates were one-half, and implementation and brand-image concerns were not an issue, a tax on GRP could increase the surplus per trip by about 18%.

This calibration highlights the significant gains that can be achieved by efficiently correcting for quality externalities. In practice, however, implementing Pigouvian taxes directly may often be infeasible. There is a delay in measuring guest returns and since they are stochastic, the tax would also create additional variance in a host's revenue. In cases where rejections are possible, users may also attempt to cherry pick matches with a high likelihood of return, especially if the calculation of inputs and methods of controlling for baseline probabilities are insufficient or unclear.

Search results

In cases where Pigouvian taxes are infeasible, a platform can also directly shift quantity from low to high GRP listings, without taxes or price changes, via its search algorithm.[45] Like taxes, including GRP in the search rankings – and telling hosts about it – would give hosts an incentive to improve GRP. And again, even if the GRP of each listing were unchanged, promoting high-GRP listings would also raise surplus by nudging guests towards higher GRP listings. The more search rankings affect which listings guests book, the more incorporating GRP into the search algorithm will raise efficiency and incentivize hosts to improve.

Consider a simple setting where in addition to GRP, there is another dimension of listing quality: their propensity to be booked when shown to guests, which we term *attractiveness*.[46] If the platform is just trying to maximize current bookings, it will show the most attractive listings 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, and the effect of a listing on the number of subsequent trips taken by a guest. The gains from maximizing this joint maximization instead of just maximizing attractiveness are (1) increasing in the standard deviation of the effect of a listing on subsequent trips (2) decreasing in the variance of attractiveness, and (3) decreasing in the correlation between the two.

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

[46] 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.

Our IV estimates suggest that a one standard deviation increase in GRP leads the guest to take .599 more subsequent rips. If we use ratings as a proxy for attractiveness, then in the case of Airbnb the correlation between attractiveness and GRP is low. Our data do not speak directly to the variance of attractiveness (since ratings are not in the units of booking probability). However, 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 (2018) finds that travel site 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 $.77 \cdot (4-1)/(4+1) \cdot .599 \approx 0.27$ trips per booking.[47] If we use 2 or 5 times as likely, instead of 4, we get 0.15 or 0.31 trips, respectively.

The average searcher in our sample took 2.1 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 7.1-15%.

It is worth noting, however, that we observe empirically that variation in GRP affects some users more than others. If travelers that are more responsive to GRP are more likely to book early search results, these gains would grow (and vice-versa). The possibility of these dynamics underscores the benefit of experimentation in deployment of new search ranking models.

Informing guests

A platform may also choose to report listing GRP to users, as many platforms do with ratings. If the information presentation is designed in a sufficiently compelling way, the platform may nudge users away from low-GRP listings in a way that increases platform surplus. If hosts or sellers understand how to improve GRP, they may make incremental investments in listing quality in a way that drives user returns to the platform.

In practice, it may be difficult to explain to both buyers and sellers on a platform what GRP measures, let alone how a buyer should weigh this information or how a seller might improve GRP.

[47] If only the top 10 listings had equivalent attractiveness, we care about the top 5 verse the next 5. Ursu (2018) 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 $.74 \cdot (2.5-1)/(2.5+1) \cdot 0.599 \approx 0.19$ trips.

With this in mind, we believe more promising alternatives may involve creating simple badges that draw attention to high-GRP listings. Flagging experienced or highly-rated sellers is common practice (e.g. Top Rated Sellers on eBay and Superhosts on Airbnb), and GRP can be used to create analogous classifications at the listing level.

In either case, experimentation will be crucial for implementation. The risk with presenting the information explicitly – 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 that interact with the direct effects of the design change.

Screening and Improving

A platform like Airbnb also has the option to remove low-quality supply. In cases where full removal is impractical, or marketplaces want to incrementally test this approach, suppression of visibility can be done independent of search ranking. This response could screen out low-quality power users, because practically those are the users for whom GRP can be precisely measured. Moreover, suppressing the presence of low-quality supply is more effective for marketplaces that are not supply constrained.

The value of screening out low-quality hosts will depend on the thickness of the platform-market and also on the extent and type of competition a platform 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 includes other 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 removes a low quality listing, that listing may appear on another platform.[48] 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. Improving listing quality would require more of an investigation into what factors drive quality, since we find it is uncorrelated with many easily observable variables.

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

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

incentive to remove low-GRP listings. Not only is it potentially easier to screen hosts than improve them, but if a low quality listing moved to another platform, that could be good for Airbnb. Diverting low-quality listings to another platform increases the probability that potential guests conclude that the other platform is low quality, and switch to Airbnb.

7. 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 2018, the raw standard deviation across listings in the number of times their guests return is about 2.2; the average is 4.1. After accounting for observable guest and trip characteristics and adjusting for small samples, the estimated standard deviation of GRP is 0.66 trips. Our IV estimates, using variation in users' search results, suggest that a standard deviation increase in a listing's GRP causes an increase in a guest's subsequent trips of 0.599 (29%).

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, so the bottom-line effect of exposing them to quality is greater. The lower probability that one bad trip causes an experienced guest to leave the platform 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;[49] the more

[49] 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.

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.

We hope that future research builds on these findings, applying new methods to new marketplaces. We find that high quality Airbnb listings exert material quality externalities onto the platform, but extensions to new settings can test the comparative statics of our theory. Future studies may also test the quality externalities that buyers exert on the platform through their effect on seller churn.

With a validated new quality metric, practitioners may incorporate GRP into experimentation and algorithmic frameworks in new ways as well. Platforms may incorporate GRP into their construction of surrogate indices to predict long-term outcomes, and indeed our results suggest that a surrogate index that excludes GRP may fail the surrogacy assumption. Platforms may also find success in more flexible prediction of both rebooking behavior and GRP using machine learning methods. Mining more rich data, such as text reviews or photo quality, may provide new views into previously unobservable components of host quality. Such approaches may help platforms get quicker signals on seller and user quality.

Appendix A: Theory

A.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, $\eta_t \sim G$, which is her value for a good of zero quality (a normalization).[50] If the consumer purchases the good, she gets a quality draw q_t and her payoff for that period is $\eta_t + u(q_t)$ for some concave function $u(\cdot)$. A consumer's prior for the subsequent period, f_t, is a function of the old prior and the quality draw.

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

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

This implies a threshold that depends on the beliefs that period, where the consumer purchases if and only if $\eta_t > \underline{\eta}(f_t)$, where $\underline{\eta}$ is a function of individual's beliefs about the distribution of quality.

For simplicity, assume that the prior is normal.[51] 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

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

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 in seller quality, the more weight the individual puts on a quality draw.

[50] Alternatively, η_t can be thought of as the the observable component of quality of the available seller, where observable and unobservable quality are uncorrelated.

[51] 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 derive interesting comparative statics with closed-form results.

LEMMA 3. *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 (lower $\frac{\partial \mu_t}{\partial q_t}$).*

Lemma 3 suggests that quality externalities may be smaller on a platform like Etsy[52] – where the artistic nature of the products may lead consumers to think that sellers differ a lot – than 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 updates her beliefs to a posterior

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

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, $\underline{\eta}(f_k) = \underline{\eta}(\mu_k, \sigma_k)$. Later posteriors are less affected by the quality draw.

LEMMA 4. *When a consumer has more past quality draws, that is $k > k'$, then*

1. *The mean of the consumer's posterior is less affected 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 $\underline{\eta}$ is decreasing in the quality draw,[53] 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*

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

When $q < \mu$, increasing the quality both increases the mean of the posterior and decreases its variance; these two changes both push $\underline{\eta}$ 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 $\underline{\eta}$ as increasing the mean. We cannot sign the derivative $\frac{\partial \underline{\eta}}{\partial q}$ 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 affected 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

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

[53] While this is the case in general, it is possible, that for quality draws way above the prior, the increase in posterior variance from a higher q draw can push the threshold up more than the increase in the posterior mean pushes it down.

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

$$\pi_i G'(\underline{\eta})\frac{\partial \underline{\eta}}{\partial q}.$$

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.[54] 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. This motivates the analysis by guest experience that we do in the empirical section of the paper.

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

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

where $\tilde{W}(f_k) = \frac{\pi_i}{1-(1-\pi_i)\delta} E_\eta[\max\{0, W(\eta|f_k)\}]$ 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.[55] Non-myopic consumers will have a lower threshold $\underline{\eta}$ 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.

A.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.[56]

[54] 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.

[55] 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.

[56] 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.

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

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

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

$$\begin{aligned} \omega_j &= d_j \pi_i \sum_{t=1}^{\infty} \delta^t E\left[G\left(\underline{\eta}\left(q_1 \ldots q_{k(t)} | \mu_0, \sigma_0, \tau_0, d_0\right)\right)\right] \\ &= d_j W. \end{aligned}$$

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 improve quality is also d_j of the social value.

Proposition 2. *The value to the seller of improving quality in period 1 is*

$$\frac{\partial \omega_j}{\partial q_j} = d_j \pi_i \sum_{t=1}^{\infty} \delta^t \frac{\partial}{\partial q_j}\left(E\left[G\left(\underline{\eta}_t\right)\right]\right) d_j$$

where $\underline{\eta}_t = \underline{\eta}_t\left(q_1 \ldots q_{k(t)} | \mu_0, \sigma_0, \tau_0, d_0\right)$. *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 affects 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

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

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

$$W(q_j) - W(0) = \sum_t E\left[G(\underline{\eta}_t(q_j|\cdot)) - G(\underline{\eta}_t(0|\cdot))\right],$$

which is the effect on the total number of expected trips across all periods. 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.[57]

A.3. 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.[58] 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 τ_t, and a consumer draws $(q, p)_{t+1}$, the consumer's posterior beliefs will be

$$\begin{aligned}
\mu_{t+1} &= \frac{\tau_t \mu_t + (q, y)_{t+1}}{\tau_t + 1}, \\
\tau_{t+1} &= \tau_t + 1, \\
\sigma_{t+1}^2 &= \frac{\tau_t}{\tau_t + 1}\left(\sigma_t^2 + \frac{((q, y)_{t+1} - \mu_t)\,((q, y)_{t+1} - \mu_t)^T}{\tau_t + 1}\right).
\end{aligned}$$

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

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 price p and correlation between price and quality, ρ, the distribution of quality is $q \sim N\left(\mu_t^q + (p_j - \mu_t^p)\rho, (1 - \rho^2)\sigma_q^2\right)$. 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 for the consumer.

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

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

[57] 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.

[58] It may also mean that some purchases may be more valuable if they transact at a higher price, which we discuss in Section 3.2.

When the correlation changes, we get

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

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, then the increase in expected quality (for a fixed p) decreases the marginal return to expected quality, so the second term is negative if the correlation is positive and the price is above the mean. Lastly, the third term is negative: a 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-rated listing.

Externality 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 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:

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

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

Appendix B: Implications

Pigouvian Tax

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, times the number of trips a guest who stays with that host takes, times the value of a trip:[59]

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

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

$$\frac{\partial S}{\partial \tau} = -J \cdot \frac{\partial d_j}{\partial p_j} (1 + DR_{jk}) \left(\frac{\partial p_j}{\partial c_j} - \frac{\partial p_j}{\partial c_k}\right) var(GRP) \cdot V^2, \tag{3}$$

where J is the number of listings and $DR_{jk} = -\frac{\partial d_k}{\partial p_j} / \frac{\partial d_j}{\partial p_j}$ is the diversion ratio (the fraction of the consumers that j loses when it raises its price that become consumers of k).[60]

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 $V = p_j - c_j = \frac{d_j}{-\frac{\partial d_j}{\partial p_j}}$. In this case, the fraction increase in surplus is approximately

$$\begin{aligned} \frac{\partial S}{\partial \tau} &\approx -J \cdot \frac{\partial d_j}{\partial p_j} V \cdot \frac{\partial p_j}{\partial c_j} var(GRP) \cdot V \\ &= -J \cdot d_j \frac{\partial p_j}{\partial c_j} var(GRP) \cdot V \\ \frac{1}{S}\frac{\partial S}{\partial \tau} &\approx \frac{\partial p_j}{\partial c_j} var(GRP). \end{aligned}$$

where the last line uses the fact that total surplus is the number of trips $J \cdot d_j$ times the surplus per trip V. 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 when pass through is .5 is $4 \cdot .2 \cdot .5 \cdot .627^2 \cdot \approx .16$. 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.

[59] 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 $\Delta d_0 = -\sum_j \Delta d_j$

[60] 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.

Search results

If Pigouvian taxes are infeasible, the platform can also directly shift quantity from low- to high-quality hosts, without taxes or price changes, via its search algorithm.[61] Like taxes, including quality in the search rankings – and telling hosts about it – would give hosts an incentive to improve quality. Even if quality of each listing were unchanged, promoting high-quality listings would also raise surplus by nudging guests towards higher quality listings. 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 examine tradeoffs a platform faces when surfacing listings to users, we consider a simple setting where in addition to GRP, there is another dimension of listing quality: their propensity to be booked when shown to guests, which we term *attractiveness*, A.[62]

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 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.[63]

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.[64] 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 $.627^2 = .39$ 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

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

[62] 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.

[63] 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.

[64] 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.

in the market (as long as the distribution of A is log-concave).[65] So the value to paying attention to GRP as well as attractiveness is increasing in the number of listings in a market.

[65] We can think of the top 20 as being drawn from a truncated distribution with a truncation point at the 21$^{\text{st}}$ 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$^{\text{st}}$ best draw, so the variance of the distribution of the top 20 is necessarily decreasing.

Appendix C: Supplementary Tables

C.1. Tables supplement Section 5.1

Table A1 Association between listing GRP and probability of return, using market FE

	All	All	1st	2nd	3rd	4+
GRP Booked	0.003***	0.013***	0.010***	0.013***	0.014***	0.014***
	(0.0002)	(0.0002)	(0.0003)	(0.0004)	(0.001)	(0.0002)
Rating at booking		0.003***	0.023***	0.011***	−0.0002	−0.013***
		(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
Controls	0	1	1	1	1	1
Observations	23,616,685	22,786,307	6,530,211	3,584,997	2,462,306	10,208,793
R^2	0.001	0.247	0.096	0.138	0.168	0.268

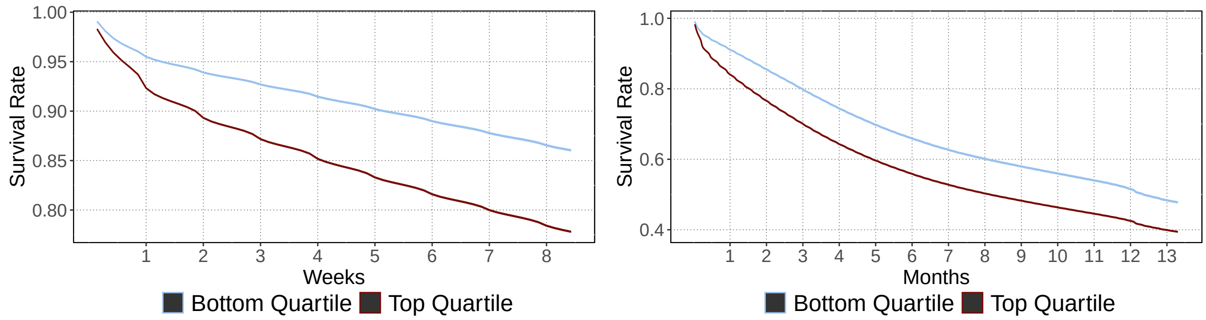
*p<0.1; **p<0.05; ***p<0.01

Table A2 Association between listing GRP and rating and winsorized subsequent spending, by guest experience

	All	All	1st	2nd	3rd	4+
GRP Booked	5.446***	28.653***	8.092***	11.588***	16.777***	43.627***
	(0.387)	(0.374)	(0.426)	(0.725)	(0.982)	(0.687)
Rating at booking		35.089***	36.003***	40.437***	35.001***	37.805***
		(1.123)	(1.228)	(2.129)	(2.916)	(2.134)
Controls	0	1	1	1	1	1
Observations	23,616,685	22,786,307	6,530,211	3,584,997	2,462,306	10,208,793
R^2	0.001	0.196	0.058	0.087	0.109	0.212

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

Note: This table uses winsorized spending and replicates the analysis from Table 5 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.



(a) First 60 days

(b) First 400 days

Figure 5 Fraction of guests that have yet to re-book on the platform 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 fraction of guests that have yet to return to Airbnb; the y-axis shows the fraction of guests that have yet to return to Airbnb *x* weeks or months after staying with one of those listings. The thickness of the lines represent 95% confidence intervals. Top and bottom confidence bounds are hard to distinguish, but differences in quartiles are statistically significant. A listing's GRP is measured by its guests prior to 2018, the returns are shown for guests starting in 2018. Because we cannot control for market in the graphs, we use a measure of GRP that does not control for market.

C.2. Tables supplement Section 5.2

Table A3 Effect of GRP and rating booked on return probability, by guest experience (IV)

	By Trip Number				
	All	1st	2nd	3rd	4+
GRP booked	0.018***	0.001	0.018	0.050	0.020***
	(0.003)	(0.011)	(0.021)	(0.032)	(0.006)
Rating booked	0.087***	0.131***	0.054	0.016	0.099***
	(0.010)	(0.034)	(0.064)	(0.097)	(0.018)
F-Stat	73924.7	8535.5	3711.3	2005.6	13428.3
Observations	2,384,401	660,925	388,286	263,039	1,072,151
R^2	0.422	0.634	0.745	0.804	0.532
Adjusted R^2	0.157	0.040	0.041	0.039	0.055

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: This table uses the same IV approach as Table 9 to show the effect of listing characteristics on whether a guest returns to 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 the room type searched for and drops searches that use other filters.

Table A4 IV:Effect of GRP and rating booked on winsorized subsequent spending

	By Trip Number				
	All	1st	2nd	3rd	4+
GRP booked	50.131***	−11.614	2.753	−2.025	71.696***
	(9.259)	(18.892)	(44.852)	(79.736)	(25.945)
Rating booked	1,304.583***	409.499***	719.179***	955.218***	2,109.560***
	(29.954)	(61.460)	(139.009)	(247.572)	(84.746)
Avg Price	1.818***	0.550***	1.091***	1.588***	3.398***
	(0.017)	(0.040)	(0.094)	(0.172)	(0.064)
F-Stat	75357.6	9134.4	3981.8	2206.7	14788.1
Observations	2,384,401	660,925	388,286	263,039	1,072,151
R^2	0.387	0.613	0.726	0.792	0.521
Adjusted R^2	0.105	−0.017	−0.028	−0.018	0.034

*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 10, 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 A5 Effect of listing characteristics on rating and price of next trip booked (IV)

	Rating of Next Booking	Price of Next Booking
GRP booked	-0.003	-4.681^{***}
	(0.003)	(1.319)
Rating booked	0.284^{***}	75.808^{***}
	(0.011)	(4.445)
F-Stat	17346.6	13298.9
Observations	1,305,907	1,018,304
R^2	0.460	0.529
Adjusted R^2	0.001	0.005

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: This table uses the same IV approach as Table 9 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.

C.3. Prices

It is hard to cleanly measure guests' learning about prices, but we look at whether high GRP has a different effect on guest behavior when it is a cheaper versus 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.

Table A6 shows the correlation between GRP and subsequent trips by residualized price quartile, looking both at the number of subsequent trips taken and the price of the next trip booked.[66] 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. Looking at the price of subsequent bookings, we see suggestive evidence that high GRP at low-priced listings leads guests to book cheaper listings for subsequent stays whereas high GRP at high-priced listings leads guests to book more expensive listings on their next trip. These are both consistent with guests learning that the correlation between price and quality is lower than they previously believed (see Equation (2)). Guests at low-priced listings feel a higher price is not worth it if they do not get higher quality, whereas guests with higher williness to pay for quality feel they need to go to more expensive listings to really get high quality (assuming the correlation is still positive).

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

Table A6 Effect of GRP by price quartile: price of next booking and subsequent trips by price residual quartile

	Trips	Trips	Price	Price
Price Residual	-0.0004^{***}	0.0001^{***}	0.280^{***}	0.272^{***}
	(0.00002)	(0.00002)	(0.001)	(0.001)
GRP x Price Q1	0.114^{***}	0.230^{***}	-1.768^{***}	-1.415^{***}
	(0.003)	(0.003)	(0.110)	(0.112)
GRP x Price Q2	0.087^{***}	0.229^{***}	-0.513^{***}	-0.664^{***}
	(0.003)	(0.003)	(0.099)	(0.100)
GRP x Price Q3	0.061^{***}	0.182^{***}	0.324^{***}	-0.136
	(0.003)	(0.003)	(0.098)	(0.100)
GRP x Price Q4	0.056^{***}	0.163^{***}	1.423^{***}	0.470^{***}
	(0.003)	(0.003)	(0.106)	(0.108)
All Controls	0	1	0	1
Observations	23,485,213	22,785,179	8,050,719	7,795,026
R^2	0.001	0.164	0.101	0.118

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: These tables shows 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 by the quartile of the price residual. The first and third columns looks only at the effect of GRP. The second and fourth columns add rating and guest, trip and market controls as well as the rating of the initial listing at the time the trip was booked.

Table A7 Effect of GRP by price quartile: price of next booking and subsequent trips by price quartile

	Trips	Trips	Price	Price
Nightly Price	-0.001^{***}	-0.001^{***}	0.276^{***}	0.342^{***}
	(0.00001)	(0.00002)	(0.0005)	(0.001)
GRP x Price Q1	0.269^{***}	0.335^{***}	-0.455^{***}	-2.319^{***}
	(0.004)	(0.004)	(0.139)	(0.142)
GRP x Price Q2	0.111^{***}	0.225^{***}	-1.285^{***}	-2.167^{***}
	(0.003)	(0.003)	(0.115)	(0.117)
GRP x Price Q3	0.018^{***}	0.155^{***}	-1.637^{***}	-1.359^{***}
	(0.003)	(0.003)	(0.115)	(0.117)
GRP x Price Q4	0.030^{***}	0.156^{***}	-2.054^{***}	-0.685^{***}
	(0.003)	(0.003)	(0.126)	(0.128)
All Controls	0	1	0	1
Observations	23,564,302	22,788,412	10,587,719	10,209,790
R^2	0.011	0.164	0.177	0.196

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

Note: These tables shows 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 by the quartile of the initial booking. The first and third columns looks only at the effect of GRP. The second and fourth columns add rating and guest, trip and market controls as well as the rating of the initial listing at the time the trip was booked.

Table A7 shows how the correlation between GRP and subsequent trips differs by non-residualized price quartile. 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. The effect of GRP on subsequent trips differs slightly more across quartiles (.34 for the lowest price quartile and 0.16 for the highest). The coefficients for the subsequent price regression all get more negative, so a higher GRP is associated with a lower subsequent price-booked for all quartiles.

C.4. Other Tables

Table A8 Jan 2018 search results summary statistics by average GRP

Statistic	Mean	St. Dev.	N
Entire Only	0.05	0.23	25,687,244
Num Nights	7.12	19.62	25,682,919
New Traveler	0.79	0.41	25,687,244
Num Guests	2.31	2.53	25,687,244

(a) Below median average GRP for listings shown

Statistic	Mean	St. Dev.	N
Entire Only	0.06	0.24	25,687,244
Num Nights	6.73	15.69	25,682,460
New Traveler	0.81	0.40	25,687,244
Num Guests	2.98	2.76	25,687,244

(b) Above median average GRP for listings shown

Note: The median average GRP of listings displayed is 0.148.

Table A9 Jan 2018 booked search results summary statistics by average GRP

Statistic	Mean	St. Dev.	N
Entire Only	0.09	0.29	266,633
Num Nights	3.99	7.68	266,632
New Traveler	0.38	0.49	266,633
Num Guests	1.82	1.77	266,633

(a) Below median average GRP for listings shown

Statistic	Mean	St. Dev.	N
Entire Only	0.10	0.30	266,633
Num Nights	4.00	6.77	266,633
New Traveler	0.42	0.49	266,633
Num Guests	2.41	2.06	266,633

(b) Above median average GRP for listings shown

Note: The median average GRP of listings displayed is 0.153.

Table A10 Correlation of quality measures (computed excluding guest characteristics) across listings

	Guest Return Propensity	Unshrunk	Raw Average
Guest Return Propensity	1	0.545	0.319
Unshrunk	0.545	1	0.801
Raw Average	0.319	0.801	1

Correlation of Listing Effect with:

Rating	Price	Number of Guests	Number of Ratings	#Ratings / #Guests
0.0443	0.0046	0.0205	0.0291	0.0447
Occupancy Rate	**Host verified**	**Avg Guest Rating Given**	**Fraction of Guests Recommended**	**Still active 1/1/2018**
0.0372	-3e-04	0.0077	0.0289	0.0182

Acknowledgments

The opinions and views expressed herein are those of the authors, not Airbnb. The authors thank Kevin Murphy, Mike Egesdal, and seminar participants at Dartmouth, Microsoft, Northwestern, and the University of Chicago for helpful comments. They are grateful to Anya Marchenko, Ryan Parks, and Graham Tierney for their research assistance.

References

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

Bardey D, Cremer H, Lozachmeur JM (2014) Competition in two-sided markets with common network externalities. *Review of Industrial Organization* 44(4):327–345.

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

Chetty R, Friedman JN, Rockoff JE (2014) Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American economic review* 104(9):2593–2632.

Ching AT, Erdem T, Keane MP (2013) Learning models: An assessment of progress, challenges, and new developments. *Marketing Science* 32(6):913–938.

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

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

Fradkin A (2018) Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb. *Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb (March 21, 2017)* .

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

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

Kim Y (2020) *Customer retention under imperfect information*. Ph.D. thesis, The University of Chicago.

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

Landon S, Smith CE (1998) Quality expectations, reputation, and price. *Southern Economic Journal* 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, Chevalier J (2014) Promotional reviews: An empirical investigation of online review manipulation. *The American Economic Review* 104(8):2421–2455.

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

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

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

Rochet JC, 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* 20–35.

Ursu RM (2018) The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science* 37(4):530–552.

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

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