

A First Look at Online Reputation on Airbnb, Where Every Stay is Above Average ¹

Georgios Zervas
Questrom School of Business
Boston University

Davide Proserpio
Marshall School of Business
University of Southern California

John W. Byers
Computer Science Department
Boston University

Abstract

Judging by the millions of reviews left by guests on the Airbnb platform, this trusted community marketplace for accommodations is fulfilling its mission of matching travelers with hosts having room to spare remarkably well. Based on our analysis of ratings we collected for millions of properties listed on Airbnb worldwide, we find that nearly 95% of Airbnb properties boast an average star-rating of either 4.5 or 5 stars (the maximum); virtually none have less than a 3.5 star-rating. We contrast this with the ratings of roughly 700,000 hotels, B&Bs, and vacation rentals worldwide that we collected from TripAdvisor. We find that hotel and B&B average ratings are much lower—3.8 and 4.1 stars, respectively—with much more variance across reviews. TripAdvisor vacation rental ratings are more similar to Airbnb ratings, but only about 85% of properties have an average rating of 4.5 or 5 stars. We then consider properties cross-listed on both platforms. For these properties, we find that even though the average ratings on Airbnb and TripAdvisor are more similar than hotels and B&Bs, proportionally more properties receive the highest ratings (4.5 stars and above) on Airbnb than on TripAdvisor. Moreover, there is only a weak correlation in the ratings of individual cross-listed properties across the two platforms. Finally, we show that these differences are consistent when considering data from two different time periods: 2015 and 2018.

Keywords: Airbnb, reputation, reviews, trust, sharing economy

¹ Authors contact information: Georgios Zervas, zg@bu.edu; Davide Proserpio, proserpi@usc.edu; John Byers, byers@bu.edu.

1. Introduction

Online reviews are a significant driver of consumer behavior, providing a convenient mechanism for consumers to discover, evaluate, and compare products and services on the web. Yet users of existing review platforms generate distributions of star-ratings that are unlikely to reflect true product quality. Most empirical papers that have analyzed the distributions of ratings arising on major review platforms have arrived at a similar conclusion: ratings tend to be overwhelmingly positive, mixed with a small but noticeable number of highly negative reviews, giving rise to what has been characterized as a J-shaped distribution (Hu et al. 2009).

Considerable effort has been dedicated to understanding how these distributions arise. The abundance of positive reviews on online platforms has been linked to at least four different underlying phenomena in the literature: herding behavior, underreporting of negative reviews, self-selection, and strategic manipulation of reviews. All of these are plausible contributing factors on Airbnb. In the scenario of herding behavior, prior reviews and ratings could subtly bias the evaluations of subsequent reviewers (Salganik et al. 2006, Muchnik et al. 2013), using a mechanism similar to anchoring bias. The phenomenon of underreporting of negative reviews occurs naturally on sites that encourage two-sided reviewing, where reviewers fear retaliatory negative reviews (Dellarocas and Wood 2008, Bolton et al. 2013, Fradkin et al. 2019). Self-selection bias arises when consumers who are a priori more likely to be satisfied with a product are also more likely to purchase and review it (Li and Hitt 2008) Finally, strategic review manipulation, typically undertaken by firms or providers who seek to artificially inflate their online reputations, is also operative (He et al. 2020, Luca and Zervas 2016, Mayzlin et al. 2014). Despite these concerns, over 70% of consumers report that they trust online reviews. The trust consumers place in online reviews is reflected in higher sales for businesses with better ratings and lower sales for businesses with worse ratings (Chevalier and Mayzlin 2006, Luca 2016).

In this paper, we evaluate the reputation system of Airbnb, a peer-to-peer marketplace for short-term rental accommodation that has now facilitated tens of millions of bookings and serves as a poster child for the so-called sharing economy. Being a peer-to-peer platform, Airbnb relies on ratings and reviews: not only do they build trust and facilitate trade among individuals, but they also serve to help determine how listings are ranked in response to user queries. We focus on Airbnb because it has several unique attributes. First, while most review platforms studied to date predominantly evaluate products, goods and services, and professional firms, Airbnb reviews are much more personal and typically rate an experience in another individual's home or apartment. Therefore, the social norms associated with these intimate Airbnb transactions may not be reflected in previously observed rating

distributions or captured by previously proposed review generation models. Indeed, both Fradkin et al. (2019) and Proserpio et al. (2018) find that the social aspect of Airbnb transactions plays an important role in the way users rate each other. Second, trust can be especially difficult to build in peer-to-peer marketplaces comprising the sharing economy, where participants face information asymmetries regarding each other's quality. Information asymmetries arise because buyers and sellers in the marketplace typically know little about each other; moreover, unlike firms with large marketing budgets, few of these individuals have an outside source of reputation or the means to build it, such as by investing in advertising or related activities. Therefore, a distinguishing feature of reviews on peer-to-peer marketplaces like Airbnb is that for most marketplace participants, reviews on the platform are their only source of reputation.

We study Airbnb's reputation platform using a dataset we collected in 2015 and updated in March 2018, encompassing all reviews and ratings that are publicly available on the Airbnb website. Having Airbnb ratings and reviews for two periods gives us the opportunity to test the extent to which the findings obtained with early data continue to hold when using the more recent data. Overall, we find that our findings are consistent across periods.

Our first main finding is that property ratings on Airbnb are overwhelmingly positive. The average Airbnb property rating garners 4.7 stars, with 94% of all properties boasting a star-rating of either 4.5 stars or 5 stars in 2015, and 91% in 2018. While one can potentially dismiss such ratings as being highly inflated and therefore not carrying much information about the sellers' true quality, we find that the picture is more nuanced. For example, we find significant and distinctive variability in rating distributions when we examine ratings by accommodation type and especially when we examine ratings across US markets.

We then consider Airbnb ratings contrasted against those at another large travel review platform, TripAdvisor, which we collected in 2015 and updated in March 2020. We chose TripAdvisor both for its worldwide scope and scale and for its accommodation diversity, as the TripAdvisor website contains reviews for hotels, bed and breakfasts, and vacation rentals. We find that the average TripAdvisor hotel rating is 3.8 stars, which is much lower than the average Airbnb property rating. This suggests that while TripAdvisor ratings employ the same 5-star scale employed by Airbnb, TripAdvisor reviewers appear to have a greater willingness to use the full range of ratings than Airbnb reviewers. Considering reviews of TripAdvisor vacation rentals in isolation, however, we see that the distribution of ratings for those properties is much closer to (but still less skewed than) the distribution of ratings of all Airbnb properties. This comparison complicates the argument that Airbnb ratings are evidently more inflated than those on TripAdvisor: perhaps the texture and quality of Airbnb stays are in fact more comparable

to vacation rental stays. Alternatively, perhaps sociological factors are at work, whereby individuals rate other individuals differently or more tactfully than they rate firms such as hotels, independent of the platform.

Our final set of findings compares properties rated on both Airbnb and TripAdvisor in an effort to quantify cross-platform effects while controlling for heterogeneity in the kinds of properties listed on each platform. Linking these properties is itself a technically difficult procedure, as we describe in Section 3.1, and results in linkages of several thousand properties, most of which TripAdvisor classifies as B&Bs or vacation rentals. We find that differences in ratings persist even when we consider this restricted set of properties that appear jointly on both platforms. Specifically, we observe that 14% more of these cross-listed properties have a 4.5-star or higher rating on Airbnb than on TripAdvisor. To explain these differences, we first consider a theory proposed in prior work, i.e., that bilateral reviewing systems, as used in Airbnb, inflate ratings by incentivizing hosts to provide positive feedback so they are positively judged in return (Dellarocas and Wood 2008, Bolton et al. 2013). In fact, using a different methodology from ours, a recent study (Fradkin et al. 2019) reports on experiments they conducted on Airbnb to investigate determinants of reviewing bias, in which they implicate various factors, including fear of retaliation and underreporting of negative experiences, to varying degrees. In our work, we contrast cross-listed properties on Airbnb with TripAdvisor, which does not use a bilateral reviewing system and provides modest corroborating evidence for bias observed in this study. We then consider the extent to which ratings of linked properties on Airbnb and TripAdvisor are correlated and find only a weak (positive) correlation between the two sets of ratings. This suggests that TripAdvisor and Airbnb reviewers have distinctive preferences in ranking and rating accommodations.

Our observational analysis sheds more light on the reputation system used at Airbnb, and our collected datasets facilitate the further study of a root cause analysis to examine the managerial, marketing, and sociological implications of the high ratings seen in this sharing economy platform. Unlike most previous observational studies, which attempt to examine one review corpus in isolation, our linkage of datasets spans two competing platforms, from the sharing economy and the travel economy, respectively. This linkage affords a new opportunity to investigate questions regarding market structure within the travel review ecosystem as well as the future of the sharing economy more broadly.

2. Airbnb and TripAdvisor platforms

2.1 Airbnb and its reputation system

Airbnb, founded in 2008, describes itself as a trusted community marketplace for people to list, discover, and book unique accommodations around the world. Millions of properties in about 200 countries can now be booked through the Airbnb platform, which has quickly become the de facto worldwide standard for short-term apartment and room rentals. Airbnb hosts offer their properties for rent for days, weeks, or months, and Airbnb guests can search for and book any of these properties, subject to host approval.

Hosts, guests, and properties each have their own dedicated webpage on Airbnb publishing user-generated content. We specifically study the ratings of the reviews written by users that Airbnb publishes on these pages. Airbnb's bilateral reviewing system allows hosts and guests to review and rate each other at the conclusion of every trip, on a scale from one to five. The text of reviews written by guests is published both on their user page and on the page of the property they stayed at; reviews written by hosts appear on the user page of both the host and of the guest that stayed at one of the host's properties.² Unlike most other major travel review platforms, such as TripAdvisor and Expedia, Airbnb does not publish the star-ratings associated with individual review. However, Airbnb does prominently display summary statistics for each property, including the total number of reviews it has accumulated and its average rating rounded to the nearest half-star (provided the property has at least three reviews). Finally, starting in 2017, hosts have the additional option to publicly respond to reviews; these responses are displayed below the reviews they address and, unlike reviews, are not associated with any star-rating.

² We note that all reviews on Airbnb are solicited and published subsequent to a verified trip, so we believe that review fraud by users, a problem that plagues other review platforms (including TripAdvisor), is not a significant factor on Airbnb.

Airbnb’s reputation system went through a major iteration in the year prior to our initial collection efforts. Prior to July 2014, Airbnb published reviews immediately upon submission, which meant that for each transaction, the user to submit a review *last* could take into account their counterparty's previously submitted review. In July 2014, to limit strategic considerations in providing feedback (e.g., to limit retaliatory reviewing), Airbnb changed its reputation system to an embargo model, simultaneously revealing reviews only once both parties supplied a review for each other, or until 14 days had elapsed from the conclusion of the trip, whichever occurred first. After 14 days, no further reviewing of a completed trip is allowed. While we were not able to study the before-and-after effects of this change, evidence from a field experiment on Airbnb suggests that it did not have a major impact on rating scores (Fradkin et al. 2019).

2.2. TripAdvisor and its reputation system

Founded in 2000, TripAdvisor is one of the world's largest websites for travel-related reviews. As of 2019, TripAdvisor contained 760 million reviews of 8.3 million accommodations (hotels, B&Bs, inns, vacation rentals), restaurants, experiences, airlines, and cruises.

TripAdvisor’s unilateral reputation system allows anyone³ to write a review about any property listed on the platform, and it does not require its user to have a TripAdvisor account to do so.⁴ TripAdvisor displays both individual review ratings and average ratings rounded to the nearest half-star aggregated at the property level for any property with at least one review. Also, beginning in 2009, TripAdvisor began to allow firms and property owners to respond to reviews. In Table 1, we present the key similarities and differences between the reputation systems of Airbnb and TripAdvisor.

3. Datasets

In this study, we combine information that we collected from both Airbnb and TripAdvisor during two different time periods, 2015 and 2018 (2020 for TripAdvisor). In so doing, we can compare the evolution of ratings on the two platforms over these two periods and test the extent to which the findings obtained with early data continue to hold when using the more recent data.

³ A proof of stay is not required by TripAdvisor for a user to write a review, unlike Airbnb.

⁴ Reviews for users that decide not to create an account are listed anonymously as reviews written by “A TripAdvisor Member”.

For Airbnb, we collected two datasets about Airbnb properties listed worldwide at airbnb.com. For each property, published reviews accumulate over time, even as the other details of the property (e.g., photographs, rental price, amenities) vary over time. Therefore, a snapshot of Airbnb at time t observes the current time-varying attributes at t and the cumulative review history up through time t . The first dataset collected in 2015 contains information about 381,297 listings available on the platform in 2015 and the reviews for those listings during the period 2008-2015. The second dataset collected between 2015 and 2018 (the last scrape being in March 2018) contains updated information and reviews about any of the 381,297 listings that continue to exist up to 2018, and information and reviews about over 2.6M additional new listings.⁵

For every property, we store its unique ID, its location, the number of reviews, and the currently displayed average star rating. Since Airbnb does not display an average rating for properties with less than three reviews, we remove those properties from our study. This procedure leaves us with about 200,000 and 1 million Airbnb properties for the 2015 and 2018 datasets, respectively.

We collected similar information from TripAdvisor. The first dataset contains information about over 500,000 hotels and B&Bs, and over 170,000 vacation rentals listed worldwide on the platform in 2015. The second dataset contains information about the same properties in 2020 (92% of the hotels available in 2015 were still available in 2020; however, only 23% of the vacation rentals were still available in 2020). For every property, we store its unique ID, its location, the number of reviews, and the currently displayed average star rating. After removing properties with fewer than three reviews (to maintain consistency with the collected Airbnb properties), our dataset contains 412,223 hotels and 54,008 vacation rentals in 2015, and 427,747 hotels and 24,547 vacation rentals in 2020.⁶

Our units of analysis are (average) property ratings on Airbnb and TripAdvisor, rounded to the nearest half-star. Throughout the paper, in the interest of clarity and to improve the paper's readability,

⁵ In an earlier version of this paper, we only used data up to 2015. This dataset was then updated over the years (until March 2018) by searching for existing and new properties available for rent on Airbnb and replacing old records with new records. This collection process makes it difficult to identify properties that are active (i.e., available for rent) and received reviews after 2015. However, as we discuss in the Web Appendix, we rely on Airbnb reviews timestamp to identify active listings in a specific year and provide robustness checks that support the validity of our results.

⁶ We are left with more hotels in 2020 than in 2015 because many properties accumulated reviews during the time period 2015-2020.

when we compare Airbnb and TripAdvisor data, we refer to the most recent datasets period as 2018 (the earliest year between Airbnb and TripAdvisor most recent datasets).

3.1 Discovering properties cross-listed on Airbnb and TripAdvisor

To link properties listed on both sites, we undertook the following heuristic procedures to perform matching in the absence of a consistent cross-platform identifier. We start with the approximate latitude and longitude provided for each TripAdvisor and Airbnb property. We then computed pairwise distances between TripAdvisor and Airbnb properties, discarding all pairs that exceeded a 500-meter distance cutoff as non-matches.

Next, for every TripAdvisor property, we iterated through all Airbnb properties in a 500-meter radius and computed a string similarity between the respective property names and descriptions on both sites. For vacation rentals on TripAdvisor, we additionally matched on the property manager name. If the resulting string similarity was above a high threshold, we kept this pair as a possible match.⁷ Finally, for each Airbnb property, if there existed a unique TripAdvisor property amongst the possible matches, we kept this pair.⁸ Using the 2015 datasets, this process generated 2,234 matches between 1,959 unique TripAdvisor properties and 2,234 unique Airbnb properties with at least three reviews. Using the most recent datasets, we find 1,843 matches between 1,549 TripAdvisor properties and 1,843 Airbnb properties with at least three reviews. Note that because we do not have updated information on some of the TripAdvisor vacation rental properties and because properties on each platform accumulated reviews in the past few years, the overlap between cross-listed properties in 2015 and 2018 is 1,248 pairs.⁹

⁷ Our results are not sensitive to the threshold used.

⁸ This procedure links each Airbnb property to at most one TripAdvisor property, but it allows for multiple Airbnb properties to be linked to the same TripAdvisor property. This is quite common, as Airbnb listings are typically at the granularity of individual rooms, whereas TripAdvisor listings are at the granularity of the property (e.g., B&B).

⁹ In the Web Appendix, we replicate our findings using this subset of listings and obtain similar results.

To evaluate the quality of our matches, we manually inspected a few hundred of them and were satisfied that only in a handful of cases were properties incorrectly associated.¹⁰ Moreover, these errors did not appear to be systematic in any way and therefore would not introduce bias in our analyses.

4. Distributions of ratings: Airbnb vs. TripAdvisor

We begin our empirical evaluation by presenting basic statistics about the distribution of ratings on Airbnb and TripAdvisor. The top panel of Figure 1 displays the distribution of Airbnb property ratings in our worldwide dataset. We find that these ratings are overwhelmingly positive, with over half of all Airbnb properties boasting a top 5-star rating, and 94% (91%) of properties rated at 4.5 stars or above in 2015 (2018). These ratings seem unusually positive, but are they truly? To the extent that Airbnb is an accommodation platform that directly competes with hotels (Zervas et al. 2017), a comparison between Airbnb and hotel ratings can be informative. The second panel of Figure 1 shows the distribution of all hotel ratings in our TripAdvisor dataset, computed using the same methodology as the Airbnb ratings. The distribution of TripAdvisor property ratings is clearly much less extreme, and this is independent of whether we use the 2015 or 2018 dataset. In 2015, for example, only 4% of hotels carry the top 5-star rating, and only 26% are rated 4.5 stars or above. This difference is also reflected in the means of the two distributions: 4.7 stars for Airbnb and 3.8 stars for TripAdvisor.

Product heterogeneity is one potential explanation underlying these differences. To compare against a more similar baseline, we exploit the fact that TripAdvisor, which is best known as a hotel review platform, also contains reviews for B&Bs and short-term vacation rentals. The third and fourth panels of Figure 1 plot the rating distributions of these property types on TripAdvisor, which are arguably more similar to the stock of Airbnb properties than hotels. These distributions visually and statistically yield less extreme differences: the average TripAdvisor B&B rating is 4.2 stars in 2015 and 4.1 stars in 2020, while the average TripAdvisor vacation rental rating is 4.6 stars in 2015 and 4.7 stars in 2020. Yet differences remain in the tails of these distributions, with only 56% (52%) of B&Bs and 84% (88%) of vacation rentals rated at or above 4.5 stars in 2015 (2018), compared to 94% (91%) for Airbnb. A basic observation we can draw from this analysis is that average ratings, even within a

¹⁰ Similar techniques based on heuristic and checks have been used in different settings, i.e., matching advertising spending and TripAdvisor hotel reviews (Hollenbeck et al. 2019).

platform, are clearly influenced by product mix. A simple linear regression backing these findings confirms that these differences are statistically significant.

Next we consider the possibility that while the overall distribution of Airbnb property ratings is highly positive, there exist specific market segments that have a less skewed distribution. To better understand potential heterogeneity underlying the distribution of property ratings, we segment properties by various attributes. First, in Figure 2 we plot the distribution of Airbnb ratings by accommodation type in 2015 and in 2018. We find evidence of limited variation in ratings: apartments and shared rooms have higher ratings than B&Bs and small hotels, but in all cases the fraction of ratings that are 4.5 stars or higher is at least 90%. Second, in Figure 3 we plot the distribution of property ratings by geographic market for six major US cities, analogous to the worldwide comparison between Airbnb properties and TripAdvisor hotels in the top two panels of Figure 1. For Airbnb, while we find evidence of considerable variation in the relative frequency of 4.5- and 5-star Airbnb ratings across cities, the fraction of ratings at or above 4.5 stars is consistently high. In contrast, while the distribution of TripAdvisor ratings by city also reveals considerable variation by city, TripAdvisor ratings continue to be less extreme than Airbnb ratings. Referring back to Figure 1, we found an overall difference of nearly 1 star between Airbnb ratings and TripAdvisor hotel ratings, and this is true independent of the observation period. Figure 3 shows that while this difference persists, there is considerable variation by city. For example, among the 6 cities we plot, the difference is highest in Los Angeles (1.2 stars) and considerably lower in cities like Boston and New York (0.6 to 0.7 stars). An interesting direction for future research is to conduct a root cause analysis of the differences between Airbnb and hotel ratings by city: are they due to differences in tourism demand? Are they due to differences in the supply of both Airbnb properties and hotels? Or, are they due to differences in travelers (demographics, taste, etc.)?

5. Comparing properties listed on both platforms

To better understand the source of these cross-platform differences in property ratings, we next consider those *cross-listed* properties that we linked using the methods described in Section 3.1. Recall

that the use of cross-listed properties allows us to control for differences in ratings arising from property heterogeneity across Airbnb and TripAdvisor. In addition, the study of cross-listed properties opens up other research questions that we are just beginning to explore and outline here. We first provide descriptive evidence for how ratings of cross-listed properties differ across platforms, consider possible explanations for these differences, and close with future directions for exploration.

5.1 The distribution of ratings for cross-listed properties

Our analysis thus far considered properties listed on each platform individually. We now limit our analysis to cross-listed properties to address a possible confounding effect: that differences in ratings could systematically arise because distinct sets of properties are listed on the respective platforms. We present the distributions of ratings for cross-listed properties in Figure 4. Here we find that the distributions of Airbnb and TripAdvisor ratings mirror the distributions shown in the top and bottom panels of Figure 1, with 14% more properties rated 4.5 stars or above on Airbnb, and a 0.1-star difference in the means of the distributions (these results apply to both 2015 and 2018). As was the case in our previous platform-wide analysis, we again observe that even cross-listed properties are rated higher on Airbnb than on TripAdvisor.

This comparison of cross-listed properties suggests that property heterogeneity alone is unlikely to fully explain the Airbnb-TripAdvisor rating gap. While a variety of factors could be causally responsible for this bias, one in particular from the platforms literature stands out. Specifically, several empirical papers (Dellarocas and Wood 2008, Cabral and Hortacsu 2010, Bolton et al. 2013) find that bilateral reputation mechanisms create strategic considerations in feedback giving, which in turn cause underreporting of negative reviews due to fears of retaliation. As Airbnb uses a bilateral review system (guests can rate hosts and hosts can rate guests), whereas on TripAdvisor only guests rate host properties, this platform difference is operative in practice. Indeed, in a controlled experiment on Airbnb, Fradkin et al. (2019) observed bias arising from bilateral reviewing on Airbnb, although, interestingly, the size of this bias was rather small. While higher ratings on Airbnb are consistent with reciprocity bias, we should not rule out differences in ratings due to reviewer self-selection—that is, a separation of reviewers across the platforms based on their distinct tastes. Indeed, recent theoretical work (Zhang and Sarvary 2014) has shown that in the presence of multiple review platforms, reviewers may split up according to their unique tastes. Future research can try to answer the following questions:

can the higher ratings of Airbnb we observe be explained by differences in preference between Airbnb and hotel travelers? Or, can the higher ratings of Airbnb we observe be explained by differences in reviewing behavior between Airbnb and TripAdvisor travelers?

5.2 How well do Airbnb ratings predict TripAdvisor ratings?

Our analysis thus far has focused on understanding differences in the distributions of ratings across TripAdvisor and Airbnb. Beyond a comparative evaluation of the two platforms, there is also considerable interest in understanding why differences in the relative rankings of properties across the platforms occur (Sun et al. 2010, Shani and Gunawardana 2010). For example, consider two properties listed on both Airbnb and TripAdvisor. Suppose that on Airbnb, property A has a higher rating than property B. Is the same true on TripAdvisor? More broadly, to what extent do ratings on one platform predict ratings on the other? To answer this question, we focus on cross-listed properties and regress the Airbnb rating of each property on its TripAdvisor rating. Note that even though TripAdvisor ratings are on average lower, they could still in principle perfectly predict Airbnb ratings (and vice versa). For example, TripAdvisor and Airbnb users could have similar tastes but a different interpretation of the 5-star rating scale, with TripAdvisor reviewers grading on a stricter curve. Therefore, differences in the means of these distributions do not predetermine the outcome of this analysis. The results of this regression are shown in Table 2 and Table 3 for the 2015 and 2018 datasets. While a significant positive association exists between the ratings of cross-listed properties across the two platforms, the adjusted R^2 of the model is low, suggesting that ratings on one platform explain only a small degree of variation in ratings on the other.

One concern with this analysis is that we are comparing properties across different geographic markets and price segments. However, most travelers limit their search for accommodation to a specific location within a target budget. Therefore, while ratings are not predictive overall, they could have more explanatory power within tightly defined market segments. For instance, it could be the case that TripAdvisor users prefer higher-priced accommodations, while Airbnb users are more price conscious. Yet when comparing properties within each price segment, users' relative preferences are the same. Motivated by this observation, we incorporate city and price-quantile dummy variables in the second and third columns of Tables 2 and 3. We see only a modest increase in the adjusted R^2 . Overall these results suggest that while on average better-rated properties on Airbnb are better-rated on TripAdvisor, there is a great deal of unexplained variation in the joint distribution of ratings across the two platforms,

even within tightly defined market segments. Perhaps, differences in consumer behavior and preference between the two platforms can help explain the remaining variation.

5.3 From ratings to rankings

We next turn our attention to analyzing the rankings of cross-listed properties on the two sites. This non-parametric comparison serves as a robustness check, since consumers could interpret ratings relatively rather than absolutely, preferring a 5-star property to 4-star property, but not necessarily ascribing much meaning to the magnitude of the star difference.

While different consumers will use different ranking heuristics, we focus on what we consider to be a reasonable, but far from universal, ranking algorithm. First, within each city we rank properties by their star-rating. Then, among properties with the same star-rating, we break ties by using the number of reviews, which is typically prominently displayed on review platforms. This choice coincides with the intuition that a 5-star property with 100 reviews is likely to contain a smaller margin of error than a 5-star property with one review, making it a less risky choice for a consumer. Finally, we break ties among properties with the same rating and number of reviews lexicographically. This is a conservative approach, as it implies that properties tied by star-ratings and number of reviews will be ranked in the same way across the two platforms. We then compute the Kendall rank correlation for the four major cities (Los Angeles, Miami, New York, and San Diego). Such correlations are generally small, ranging from 0.05 (San Diego) to 0.33 (Los Angeles) in 2015, and they slightly increase in 2018 (see Table 4). While there are some cross-city differences in terms of correlations in 2015, these largely disappear by 2018. Identifying the causes of these differences remains an interesting future research direction: Are they due to differences in city characteristics, differences in the travelers that visit these cities, or differences in the type of supply (Airbnb properties and hotels) offered in these cities?

Overall these results support our regression analysis. Airbnb and TripAdvisor reviewers exhibit little agreement: TripAdvisor and Airbnb ratings are only weakly correlated, with the relative rankings of properties varying to a significant degree across the two sites.

One limitation of our work is that the cross-listed properties we have discovered constitute a small fraction of the inventory available on either site. That said, any cross-listed properties that our heuristics did not successfully match would not alter the relative order of the properties we have discovered. However, the unlikely possibility remains that there is much higher correlation in the ranks and ratings of cross-listed properties we have not matched.

6. Managerial implications and future directions

Our findings have important implications for platforms and consumers. For platforms, our results show that the design of a reputation system and the nature of the transactions might have a strong impact on the reviews and ratings posted on the platform; moreover, these effects cannot be easily explained by previously studied phenomena such as retaliation, reciprocal ratings, or fake reviews. This is because starting in mid-2014, Airbnb made changes to its reputation system, making it harder for its users to game the reputation system and because fake reviews on Airbnb are very costly.¹¹ Both Fradkin (2019) and Proserpio et al. (2018) arrive at similar conclusions, providing additional evidence suggesting that the social interaction part of an Airbnb stay might play an important role in driving up user ratings.

More generally, while review and booking platforms strive to maintain unbiased ratings, our results provide strong evidence suggesting that achieving this goal is hard. Despite the fact that platform design has improved over the past decade or so, there are platform characteristics that can affect ratings and that cannot be easily changed (e.g., whether transactions involve a social component, or whether transactions are between peers or between a peer and a firm). This means that higher ratings like those observed on Airbnb might be an idiosyncratic feature of Airbnb and similar platforms. Indeed, similar rating distributions have been observed on other peer-to-peer platforms such as Uber.¹²

The implications for consumers are less clear. Our findings suggest that consumers looking for a property are likely to be exposed to different ratings and rankings for the same properties on different platforms, which in turn might affect their final choices. Whether these final choices are sub-optimal and therefore welfare-reducing is an open question. It could also be the case that consumers with similar tastes in accommodations gravitate toward the same platform, in which case platform diversity can be welfare-enhancing.

Turning to future research directions, our work opens more questions than it can answer. First, while we show that ratings on Airbnb are more positive than those on TripAdvisor, we are not able to pin down a specific explanation for this finding, and more research is needed to better understand the

¹¹ Only guests with a past stay can write a review, and the minimum transaction cost on Airbnb is \$10 plus Airbnb service fees.

¹² See, for example, <https://theatlantic.com/charts/4y8zfUM9>

causes behind our results. For example, researchers can investigate the following questions: are Airbnb travelers simply more positive than hotel travelers? And, if so, why? Or, is it easier to leave a negative review when the other party is a hotel as opposed to a person?

Second, our findings have the potential to be of consequence for platforms and consumers. Platforms like Airbnb and Uber continue to prosper and grow despite the distribution of ratings we observe, which seems to suggest that consumers are satisfied with their choices; so far this seems to be a win-win situation. This leads to the following research questions: are consumers harmed from the bias we document? Or, are ratings and reviews still informative and consumer choices optimal?

Finally, our work underscores the importance of platform design and the effect that different designs have on reviews and ratings: are higher Airbnb ratings due to a better platform design that facilitate better matches between hosts and guests? We expect to see more research looking at the effect of different designs on different outcomes such as consumer choices, platform utilization, and consumer reviews and ratings.

7. Conclusion

Reflecting on Airbnb in context, research in other online marketplaces indicates that positive ratings are critical to entrepreneurial and platform success when they play such a prominent role in ranking and user selection. One widely referenced experiment in online entry-level labor markets (Pallais 2014) demonstrates that a single detailed evaluation can substantially improve a worker's future employment outcomes. At the platform level, recent work (Nosko and Tadelis 2015) shows that eBay buyers draw inferences about the eBay marketplace at large based on their experiences with specific sellers, and that buyers who have a poor experience with any one seller are less likely to return to eBay. The same has been found to be true on Airbnb (Jaffe et al. 2017). These studies suggest that attaining high ratings is likely essential to an individual entrepreneur's success on the Airbnb platform. As a result, hosts may take great pains to limit negative reviews. These include rejecting guests that they deem unsuitable or risky, or resetting a property's reputation with a fresh property page once a property receives too many negative reviews. Airbnb itself may also seek to limit negative reviews, for

example by adding frictions to make it harder for users to provide a low rating, or by removing negative reviews.¹³ We look forward to exploring these and other possible explanations in our future work.

¹³ See, for example, <https://qz.com/1333242/airbnb-reviews/>.

References

- Bolton, Gary, Ben Greiner, Axel Ockenfels. 2013. Engineering trust: reciprocity in the production of reputation information. *Management Science* 59(2) 265-285.
- Cabral, Luis, Ali Hortacsu. 2010. The dynamics of seller reputation: Evidence from eBay. *The Journal of Industrial Economics* 58(1) 54-78.
- Chevalier, Judith A., Dina Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43(3) 345-354.
- Dellarocas, Chrysanthos, Charles A. Wood. 2008. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science* 54(3) 460-476.
- Fradkin, Andrey, Elena Grewal, and David Holtz. 2019. Reciprocity in two-sided reputation systems: Evidence from an experiment on Airbnb. Working paper.
- He, Sherry, Brett Hollenbeck, and Davide Proserpio. "The Market for Fake Reviews." Available at SSRN (2020).
- Hollenbeck, Brett, Sridhar Moorthy, and Davide Proserpio. 2019. Advertising strategy in the presence of reviews: An empirical analysis. *Marketing Science* 38(5) 793-811.
- Hu, Nan, Jie Zhang, Paul A. Pavlou. 2009. Overcoming the j-shaped distribution of product reviews. *Communications of the ACM* 52(10) 144-147.
- Jaffe, Sonia, Peter Coles, Steven Levitt, Igor Popov. 2017. Quality externalities on platforms: The case of Airbnb. Working paper.
- Li, Xinxin, Lorin M. Hitt. 2008. Self-selection and information role of online product reviews. *Information Systems Research* 19(4) 456-474.
- Luca, Michael. 2016. Reviews, reputation, and revenue: The case of yelp.com. Tech. rep., Harvard Business School.
- Luca, Michael, Georgios Zervas. 2016. Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science* 62(12) 3412-3427.
- Mayzlin, Dina, Yaniv Dover, Judith Chevalier. 2014. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review* 104(8) 2421-2455.
- Muchnik, Lev, Sinan Aral, Sean J. Taylor. 2013. Social influence bias: A randomized experiment. *Science* 341(6146) 647-651.
- Nosko, Chris, Steven Tadelis. 2015. The limits of reputation in platform markets: An empirical analysis and field experiment. Tech. rep., National Bureau of Economic Research.
- Pallais, Amanda. 2014. Inefficient hiring in entry-level labor markets. *American Economic Review* 104(11) 3565-3599.

Proserpio, Davide, Wendy Xu, Georgios Zervas. 2018. You get what you give: theory and evidence of reciprocity in the sharing economy. *Quantitative Marketing and Economics* 16(4) 371-407.

Salganik, Matthew J., Peter Sheridan Dodds, Duncan J. Watts. 2006. Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762) 854-856.

Shani, Guy, Asela Gunawardana. 2010. Evaluating recommendation systems. F. Ricci, L. Rokach, B. Shapira, P. B. Kantor, eds., *Recommender Systems Handbook*. Springer, 257-297.

Sun, Mingxuan, Guy Lebanon, Kevyn Collins-Thompson. 2010. Visualizing differences in web search algorithms using the expected weighted Hoeffding distance. *WWW 2010* 931-940.

Zervas, Georgios, Davide Proserpio, John W. Byers. 2017. The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Journal of Marketing Research* 54(5) 687-705.

Zhang, Kaifu, Miklos Sarvary. 2014. Differentiation with user-generated content. *Management Science* 61(4) 898-914.

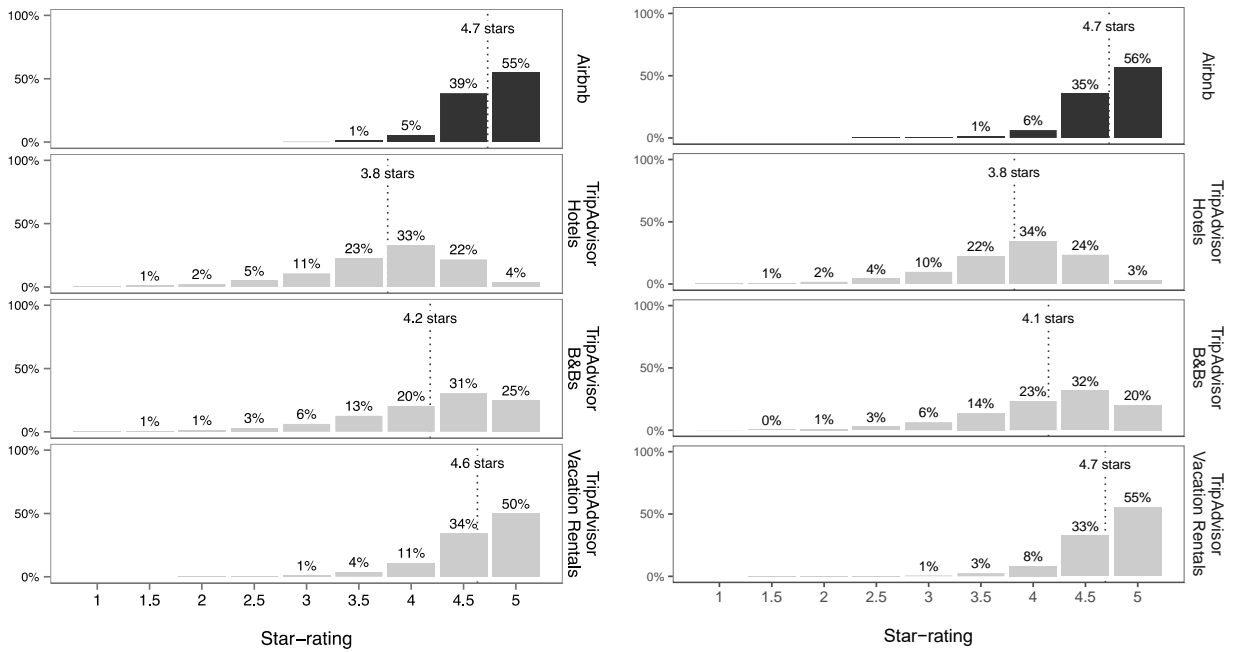


Figure 1: Distribution of property ratings on Airbnb and TripAdvisor in 2015 (left) and in 2018 (right). The dotted lines show the distribution means.

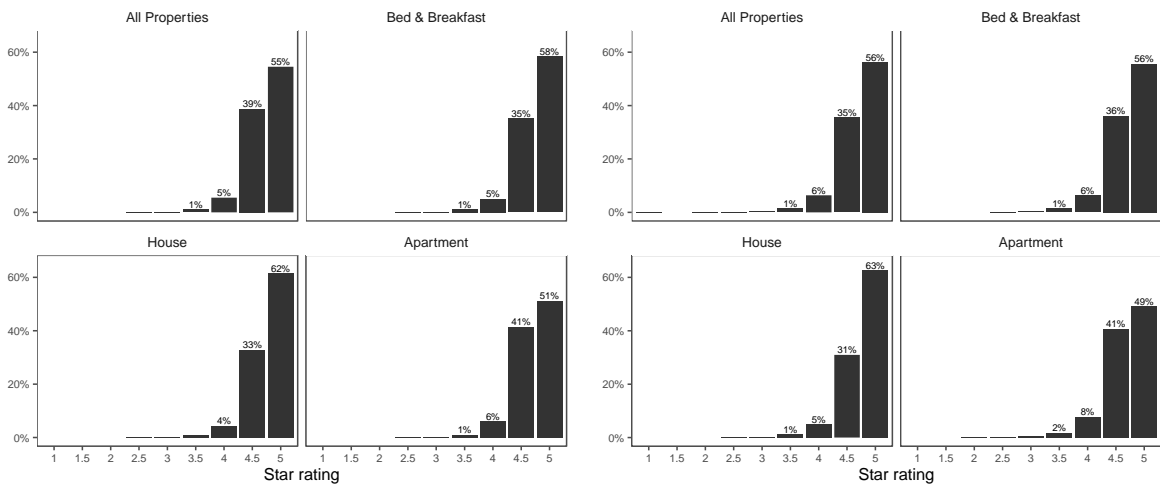


Figure 2: Distribution of Airbnb property ratings by accommodation type in 2015 and 2018.

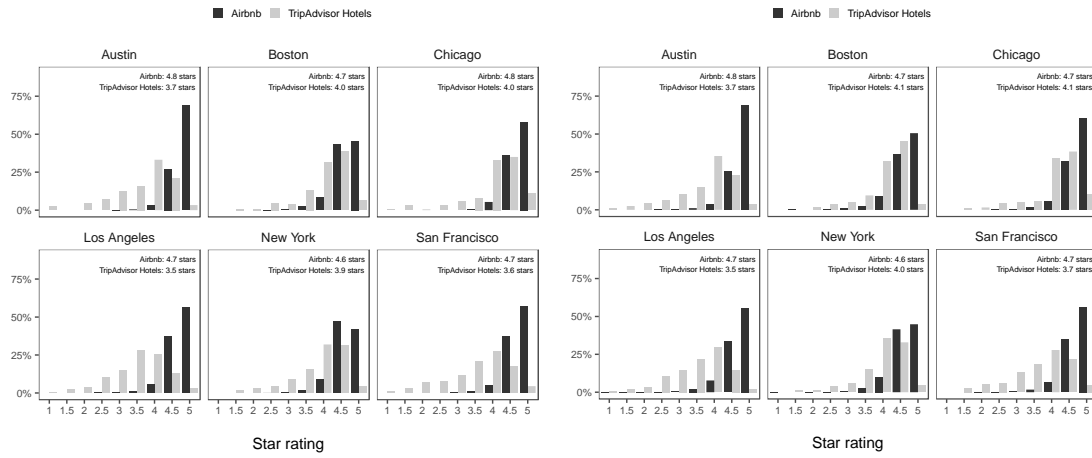


Figure 3: Distribution of Airbnb and TripAdvisor property ratings by US market in 2015 and in 2018 (right).

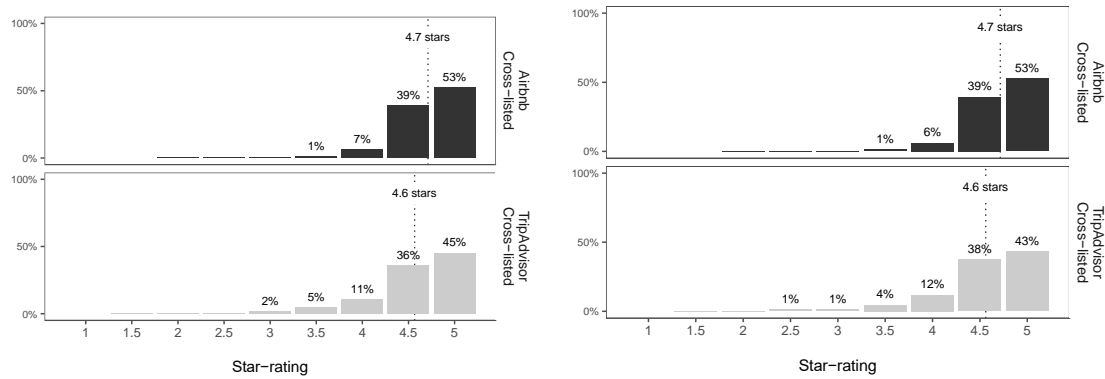


Figure 4: The distribution of ratings for properties cross-listed on both Airbnb and TripAdvisor in 2015 (left) and 2018 (right). The dotted lines show the distribution means.

Table 1: Comparing Airbnb and TripAdvisor platforms and their reputation system.

	Airbnb	TripAdvisor
Properties listed	Vacation rentals	Hotels, B&Bs, Vacation rentals
Reservations	On Airbnb	Anywhere
Reputation system	Bilateral	Unilateral
Who can write reviews	Users with a booking	Everyone
Sign-in required	Yes	No
Display avg. ratings	After 3 reviews	Always
Individual review rating	No	Yes
Review responses	Starting 2017	Starting 2009

Table 2: The relationship between Airbnb and TripAdvisor star-ratings in 2015.

	(1)	(2)	(3)
TripAdvisor Ratings	0.275*** (15.88)	0.244*** (13.45)	0.238*** (12.82)
City Dummies	No	Yes	Yes
Price Dummies	No	No	Yes
N	2234	2234	2234
R ²	0.18	0.55	0.55
Adj. R ²	0.17	0.22	0.22

Note: The dependent variable is the Airbnb star-rating of each linked property.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 3: The relationship between Airbnb and TripAdvisor star-ratings in 2018.

	(1)	(2)	(3)
TripAdvisor Ratings	0.309*** (17.82)	0.227*** (9.78)	0.221*** (9.39)
City Dummies	No	Yes	Yes
Price Dummies	No	No	Yes
N	1843	1843	1843
R ²	0.21	0.63	0.63
Adj. R ²	0.21	0.24	0.24

Note: The dependent variable is the Airbnb star-rating of each linked property.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 4: TripAdvisor vs. Airbnb ranks correlation for cross-listed properties in 2015 and 2018.

	2015	2018
Los Angeles	0.33	0.38
Miami	0.23	0.33
New York	0.21	0.18
San Diego	0.05	0.25

Web Appendix

In this Appendix, we test the robustness of our findings by reporting additional results using different subsets of our datasets.

In Figure 5, we replicate Figure 4 in the paper, using the 1,248 pairs of properties common across the 2015 and 2018 datasets. Results are similar to those reported in the paper.

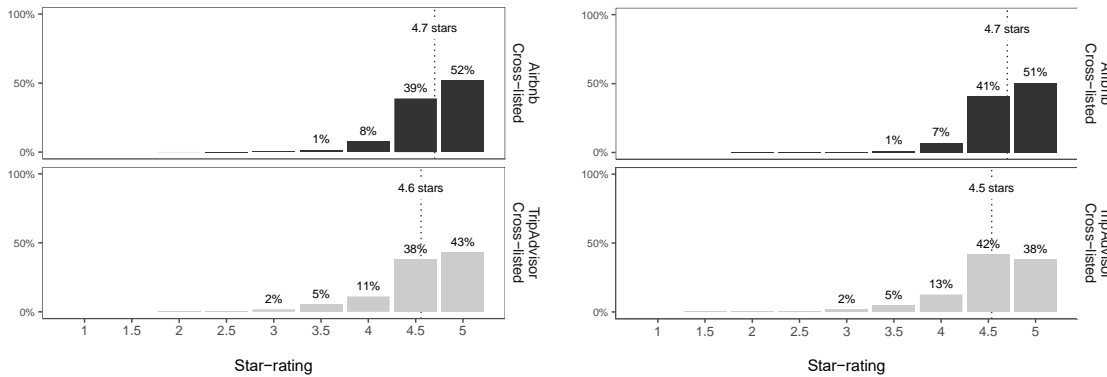


Figure 5: The distribution of ratings for properties cross-listed on both Airbnb and TripAdvisor in 2015 (left) and 2018 (right) using cross-listed properties that exist both in 2015 and 2018. The dotted lines show the distribution means.

To reduce concerns about whether Airbnb listings are indeed active and received reviews after 2015, in Figure 6 we replicate the right panel of Figure 5 above, using Airbnb listings that received at least one review in 2017 (which leaves us with 828 cross-listed pairs).¹⁴ Again, results continue to hold.

¹⁴ We obtain similar results using listings with at least 1 review in 2018.

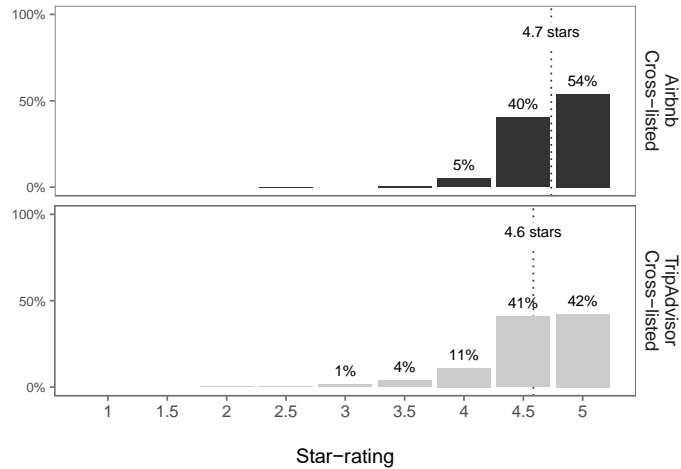


Figure 6: The distribution of ratings for properties cross-listed on both Airbnb and TripAdvisor in 2018 using only properties with at least a review in 2017. The dotted lines show the distribution means.

Finally, in Figure 7 we replicate the top panel of Figure 1, using listings that received at least one review in 2016, 2017, and 2018. We find that results are similar to those reported in the paper and that, if anything, Airbnb ratings seems to become slightly more positive over time.

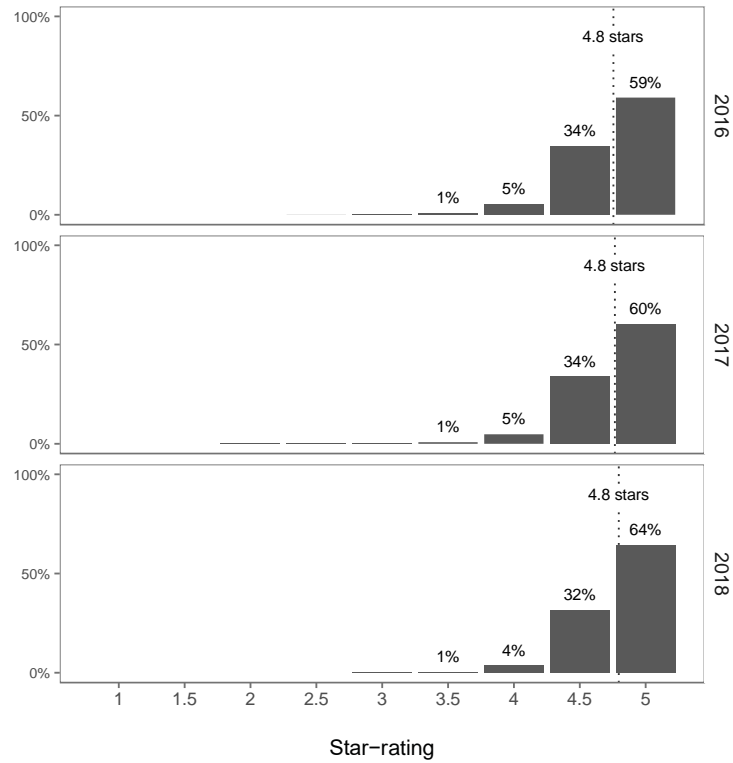


Figure 7: Distribution of property ratings on Airbnb using properties with at least a review in the years 2016, 2017, and 2018. The dotted lines show the distribution means.