How to "buy" honest reviews. Experimental evidence of the impact of prices on online reviews*

Przemyslaw Jeziorski† Konstantina Michelidaki‡

October 26, 2023

Abstract

This paper documents the causal relationship between transaction prices and subsequent product reviews given by customers. Drawing on evidence from a Large Scale Randomized Field Experiment that randomly adjusted nightly rates for short-term rentals, we ascertain that an exogenous 10% price decrease (increase) leads to a subsequent increase (decrease) in star ratings by a quarter of the standard deviation. The increase in reviews is especially pronounced for users with more diffused beliefs about product quality and for less affluent users with greater demand elasticity. In stark contrast, the effects are reversed for a niche group of highly experienced and affluent users. Our findings align with the notion that the impact of prices on reviews stems primarily from disconfirmation of quality expectations but also from the direct impact of price on consumer surplus. In parallel, we illustrate that positive reviews can uplift the demand curve. Consequently, firms might "invest" in favorable reviews by reducing prices in addition to improving product quality. Conversely, businesses could monetize positive reviews by increasing their prices. We conclude by reflecting on the ramifications of our findings for strategic pricing and overall market efficiency.

*We would like to thank Andrey Fradkin, Ganesh Iyer, and Miguel Villas-Boas for helpful discussion and suggestions. Both authors are co-founders and hold equity in Keybee Inc., the data supplier for this paper. Both authors received income from Keybee Inc. in excess of $10,000, which was unrelated to the research in this paper. Both authors and are on the board of directors of Keybee Inc. Konstantina Michelidaki is a CEO of Keybee Inc.

†przemekj@berkeley.edu, UC Berkeley, Haas School of Business

‡nadia@keybeehosting.com, Keybee Inc.

1 Introduction

Reputation systems based on customer reviews emerged as a prominent feature of modern online marketplaces. The main purpose of the reviews[1] is to increase market efficiency by lowering the information gap about product quality between buyers and sellers.[2] Indeed, numerous findings suggest that customer reviews contain useful information about product quality and can predict sales.[3] However, the literature identified several mechanisms impeding the information value of the reviews, including self-selection, review externalities, and fake reviews.[4] We show firms can "buy" favorable organic reviews by temporarily reducing prices.

To obtain the causal impact of prices on reviews, we used a naturally occurring field experiment in which an optimal pricing algorithm randomly perturbed millions of posted prices on a leading online platform offering short-term rentals. We show that an exogenous 10% decrease in transaction price leads to an increase in review rating by approximately a fourth of the standard deviation. We also demonstrate that the effect is more prominent when consumers have likely more diffused priors about product quality and when consumers are less wealthy.

Conversely, we discover a positive relationship between the most recent review and subsequent demand curve. In particular, we show that an exogenous one standard deviation increase in star rating of the most recent sale increases the price of a subsequent transaction between 20 to 50 percent.[5] Moreover, we find that, alongside of larger transaction price, a similar increase in rating shortens the waiting time for the next sale by 2 to 17 days. The causal interpretation of the link between reviews and future performance relies on experi-

[1] We use ratings and reviews interchangeably through the paper.

[2] Classical reference showing the detrimental role of asymmetric information is Akerlof (1978). Beyond adverse selection, asymmetric information may also lead to more nuanced frictions such as statistical discrimination, see Laouénan and Rathelot (2022).

[3] See Godes and Mayzlin (2004), Zhang, Chris, and Neveen (2004), Clemons, Gao, and Hitt (2006), Chevalier and Mayzlin (2006), Cabral and Hortacsu (2010), and Anderson and Magruder (2012). For a broader discussion, refer to a survey by Dellarocas (2003).

[4] Lewis (2011) shows that online marketplaces are subject to a significant residual information gap despite the ubiquity of reviews. Dellarocas (2006) discusses incentives of the firm to manipulate reviews directly.

[5] The dynamic pricing algorithm reduces the price as the check-in date approaches to obtain over 90% capacity utilization; thus, transaction prices, not conversion rates, are the most informative measure of profitability in our setting.

mentally induced variation in past reviews via lagged price shocks, which validates previous results that show positive impact of reviews on sales relying on observational methods.

The impact of transaction price on subsequent reviews and the converse demand feedback loop has important implications for pricing strategy. Traditional static pricing theory prescribes that the firm should adjust its price if contemporaneous demand elasticity changes due to a change in review rating.[6] Nonetheless, since the prevailing price affects reviews, which subsequently shape future profits, optimal pricing becomes dynamic, even if other market features only require static pricing.[7] In particular, a firm may lower its price to *invest* in better reviews, taking into account the ability to *monetize* positive evaluations in the future (Dellarocas, 2003, discusses various ways for sellers to monetize high average rating). This mechanism affects market design and regulation since inferior firms with deep pockets may be able to achieve market dominance via product reviews.[8]

The data suggest multiple potential mechanisms that might establish a relationship between transaction prices and subsequent customer satisfaction expressed in reviews. One potential mechanism involves the decreased surplus (utility minus price) experienced by customers who continue to purchase the product despite its increased price (henceforth, inframarginal buyers). Should buyers integrate this surplus into their ratings, elevated prices would lead to decreased reviews, even when there is complete information about quality. This regularity is termed the *consumer surplus effect*. In support of the idea that consumer surplus influences customer satisfaction, we observe that a price rise has a particularly negative effect on reviews in the *value* category, which directly addresses surplus.[9] Additionally, we note that more affluent users, originating from areas with higher median income and lower unemployment, do not alter their reviews. Such consumers may perceive themselves as less affected by increased prices due to a more flexible budget than their more budget-sensitive peers. We also confirm that this variation does not simply result from a straw-man

[6]For example, Zhong (2021) shows that sellers raise prices after a discontinuous jump in their star rating.

[7]See Stenzel, Wolf, and Schmidt (2020) for theoretical analysis.

[8]The mechanism is similar to *predatory pricing* and *dominance* in the markets with learning-by-doing, see Cabral and Riordan (1994) Cabral and Riordan (1997); or building *brand equity* via costly advertising and promotion, see Dubé, Hitsch, and Manchanda (2005) and Borkovsky, Goldfarb, Haviv, and Moorthy (2017).

[9]For reference, Airbnb describes value as "did the guest feel that the listing provided good value for the price?".

supposition that wealthier users inherently give more favorable reviews. This confirmation is derived from the lack of a relationship between financial capability and the overall star rating.

The second possible mechanism is more nuanced and results from an *advantageous selection* of buyers as the posted price increases.[10] Contrary to the consumer surplus effect, this mechanism could forge a positive correlation between prices and reviews. Specifically, if a minor price increment excludes a substantial number of buyers who were not ideally suited for the product (henceforth, marginal buyers), such an increase might boost reviews. Reviews increase because, after removing the marginal consumers, the residual buyers may be, on average, more satisfied despite the price increase. Overall, the impact of price on reviews depends on the number of marginal buyers removed by the price hike compared to the negative impact of the hike on the remaining buyers.[11]

The third potential mechanism relies on imperfect information about product quality. The considerable size of the information gap would not be surprising since short-term rental accommodations are prone to more significant quality dispersion than hotels. In light of such information discrepancy, beyond the online listing's content and associated reviews, the advertised price might also function as a signal of quality (see Wolinsky, 1983; Erdem, Keane, and Sun, 2008; Guo and Jiang, 2016; Wang and Van Der Lans, 2018). As a consequence, asymmetric information is expected to exacerbate the negative impact of price on reviews because consumers subjected to higher prices often anticipate a greater level of quality – we call this a *signaling effect*.[12] Consistently with this mechanism, the data reveals that even the ratings related to objective measures, such as *location* or *check-in instructions*, decrease significantly after the price hike. If consumers infer the quality of location or check-

[10]Pioneering work theoretical work demonstrates that the selection of reviewers matters was done by Li and Hitt (2008); Hu, Pavlou, and Zhang (2017); Brandes, Godes, and Mayzlin (2022).

[11]An anecdotal example of advantageous selection from the academic profession concerns instructor ratings. Some faculty believe that increasing the class difficulty (the price of taking the class) may result in better instructor reviews. The presumed increase in ratings may occur because higher difficulty encourages students who are not the best match to drop the course. However, by a similar principle, increasing the difficulty without additional academic benefits would upset students who chose to remain in the class.

[12]Extending the student teaching evaluation example from Footnote 11, it is conceivable that an increase in the perceived difficulty of the class could lead to students expecting larger academic benefits. However, if the class becomes harder without additional gratification, the reviews should reflect the gap between expectations and revealed academic benefits.

in instructions from the posted price, a price increase may lead to ex-post regret. Moreover, the prevailing price has little effect on reviews for users with less diffused priors. We measure prior diffusion using tenure on the platform, number of prior purchases, and travel distance. The mechanism is an instance of Expectancy-Disconfirmation (see Olson and Dover, 1979; Oliver, 1980; Swan and Trawick, 1981).

Considering the three effects collectively, our analysis indicates that the signaling effect exceeds the consumer surplus effect. This inference is drawn from the observation that customer experience accounts for more variation in the individual treatment effects than customer wealth. Regarding the direction of the relationship, the consumer surplus and signaling effects together are considerably larger than the advantageous selection effect, as evidenced by the negative average impact of prices on reviews. However, our theoretical framework anticipates that if advantageous selection outweighs consumer surplus and signaling effects, we should expect a positive relationship between prices and reviews. While the overall average treatment effect is negative, our data shows that roughly 30% of individual treatment effects yield positive point estimates, even if these are of lesser magnitude than the negative effects. These individual effects are subject to statistical noise; nevertheless, we identify a statistically significant positive average treatment effect in a modest 3% fraction of particularly experienced and affluent customers. This observation implies that optimal pricing, when considering reviews, is conducive to price discrimination. For instance, proprietors of premium listings, when selling to seasoned Airbnb users, might refrain from offering discounts and capitalize on advantageous selection.[13]

Establishing a relationship between price and reviews is challenging without an experiment. Prices and reviews are simultaneously determined by unobserved product characteristics, such as quality or other factors influencing customer satisfaction.[14] Various sources trig-

[13]Drawing a parallel with class ratings: instructors attentive to reviews may benefit from amplifying a course's rigor beyond its typical threshold, mainly when students already exhibit familiarity with the course content, leading to less diffused priors. This approach is exceedingly feasible in environments where students are anticipated to manage added challenges with minimal setbacks.

[14]In alignment with the literature in Industrial Organization, we adopt a broad definition of quality. Specifically, quality refers to any factor that causes the demand curve to shift upwards. As an illustration, if all other factors remain constant, an accommodation offering more amenities possesses a higher quality than one with fewer amenities. Similarly, the quality of a given accommodation might change based on conditions. For example, accommodation at the beach has a higher quality during warm weather.

ger variation in quality, emerging from cross-product differences, demand seasonality, macro business cycles, localized effects like gentrification, and firm-level adjustments in product quality. Each of these is prone to influence both prices and reviews.

Due to this simultaneity, conducting a straightforward regression of reviews on prices results in a spurious correlation, which obfuscates the causal relationship under investigation. Navigating through this spurious correlation entails overcoming two hurdles. Firstly, the presence of potential product-level quality shocks means that even estimates based on panel data might harbor bias.[15] Secondly, it is difficult to determine the bias direction due to the ambiguous relationship between price and quality.[16] This paper explores multiple estimators, revealing price endogeneity when relying on observable product characteristics and time and product fixed effects. The exposition culminates with an instrumental variables estimator that employs exclusively experimental price variation.[17]

The closest paper to this study is the pioneering work of Luca and Reshef (2021). Employing panel data secured from the Yelp Transactions Platform, they demonstrate that price increments correspond to decreased Yelp restaurant ratings. Their methodology leans on observational data, complemented by restaurant-time fixed effects. Expanding upon their insights, we discern similar effects using experimental price variation. Beyond the negative relationship between prices and reviews, our research identifies segments with reversed effects. Furthermore, we introduce a theoretical framework outlining multiple rationales for the observed effects and clarifying the unearthed positive associations. To evaluate our framework,

[15] Gu and Lin (2006) identifies a correlation between incoming and most recent ratings for a particular product, even after controlling for the overall mean rating. This implies that negative and positive ratings cluster, potentially indicating product-level quality fluctuations.

[16] A prevalent observation suggests that price often correlates positively with quality as highlighted by Berry (1994). Such a relationship could introduce a downward bias in our estimates, meaning we might isolate an attenuated negative effect of prices on reviews. On the other hand, the association between price and quality can also be inverse. For instance, Yao, Wang, and Mortimer (2022) explains the concept of *shrinkflation*, wherein a price increase driven by supply factors may occur alongside a decrease in product size. Pertaining to Airbnb rentals, anecdotal evidence indicates that hosts endeavor to gain a competitive advantage during low demand (and low prices) by enhancing their listings. Enhancement can be achieved by introducing additional amenities or providing services like complimentary airport transportation. This scenario may be particularly problematic because the data might suggest a negative relationship between prices and reviews, even when no causal connection exists.

[17] An additional, more technical source of bias in panel-data estimators is the typical problem encountered in a dynamic panel with fixed effects, as described in Arellano and Bond (1991). Bias arises if the regressors are predetermined (not strictly exogenous). Regrettably, previous prices will likely breach strict exogeneity as they correlate with concurrent and lagged quality.

we deploy unique data on review frames, emphasizing varied facets of customer experiences. We also scrutinize several factors that shape customer preconceptions and the extent of their price sensitivity. We corroborate Luca and Reshef (2021)'s conclusion regarding experience acting as a modifier of the effects; moreover, we investigate other influential factors related to customer affluence, providing evidence of the consumer surplus and advantageous selection effects.

An extensive literature shows that reviews and word of mouth are a proxy for customer satisfaction (see Resnick and Zeckhauser, 2002; Dellarocas and Wood, 2008; Zhong, 2021; Nosko and Tadelis, 2015; Babić Rosario, Sotgiu, De Valck, and Bijmolt, 2016). Greater customer satisfaction should lead to better reviews, which suggests that firms may improve their reviews by raising customer satisfaction. For example, the most direct, albeit costly, way to achieve this goal is to improve product quality (see Hunter, 2020; Ananthakrishnan, Proserpio, and Sharma, 2019). We show that firms may achieve better customer satisfaction by lowering prices.

The paper contributes to the literature that studies the consumer deliberation process when leaving a review (see Dellarocas and Narayan, 2006). For instance, several studies show the potential bias of customer satisfaction measures that may stem from rating inflation or self-selection (see Peterson and Wilson, 1992; Moe and Schweidel, 2012; Gao, Greenwood, Agarwal, and McCullough, 2015; Filippas, Horton, and Golden, 2022). Several other papers propose solutions to the bias. For instance, Hui, Klein, and Stahl (2022) postulate that customers are more likely to review the more they learn about the seller during the transaction. Also, Burtch, Hong, Bapna, and Griskevicius (2018) and Fradkin and Holtz (2022) demonstrate that consumers may be convinced to leave a review by monetary incentives, suggesting substantial monetary reviewing costs. Moreover, Askalidis, Kim, and Malthouse (2017) show that consumers may leave more accurate reviews if chosen randomly and not exposed to previous reviews. Similarly, Karaman (2021) shows that review solicitation may have a similar effect. Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee (2009) finds evidence for conformity bias – other consumers find reviews more helpful if they confirm their prior opinion about the product.

In the next section, we describe the short-term rental industry.

2 Short-term rental industry

This paper focuses on the short-term rental (STR) industry.[18] STR is a business model in which landlords (henceforth called hosts) rent furnished properties for a short time, compared to long-term rentals – primarily unfurnished properties rented annually. In 2021, the global market value of the STR industry was predicted to be $78 billion.[19] Depending on the local market, STR landlords must obtain operating licenses and adhere to various particular regulations. For instance, many jurisdictions define short-term rental as any rental below 30 days. Thus, if the rental period is shorter than 30 days, the STR tenants (henceforth called guests) often lack tenant rights and must abide by tailored STR contracts. The closest competitors to STRs are hotels (see Zervas, Proserpio, and Byers, 2017).

STRs have a long history and were typically executed via informal markets. Today, most short-term rentals occur via organized online platforms like Airbnb, VBRO, Expedia, and Booking.com. These platforms facilitate the interaction between hosts and guests, from search, booking, and payment to check-out and feedback. STR platforms provide a contractual framework and arbitrate disputes. A significant difference between hotels and STRs is the degree of ownership concentration. The majority of hosts are private individuals offering one or a few residences for rent.[20] In comparison, a hotel typically offers multiple rooms and could be a part of a larger chain. For this reason, the degree of asymmetric information about product quality can be more substantial for a short-term rental than for a hotel. To mitigate the asymmetric information the platforms have introduced elaborate review systems. The actual design of these systems varies. For example, Booking.com allows to rate hosts but does not allow rating guests. In comparison, VRBO and Airbnb allow double-blind guest and host ratings to avoid reciprocation and retaliation.[21] Possibly due to extensive asymmetric information, reviews became the leading driver of the demand for STRs. The reviews have the potential to directly influence customer purchase decisions. Reviews are

[18] Another synonymous name for short-term rentals is vacation rentals.

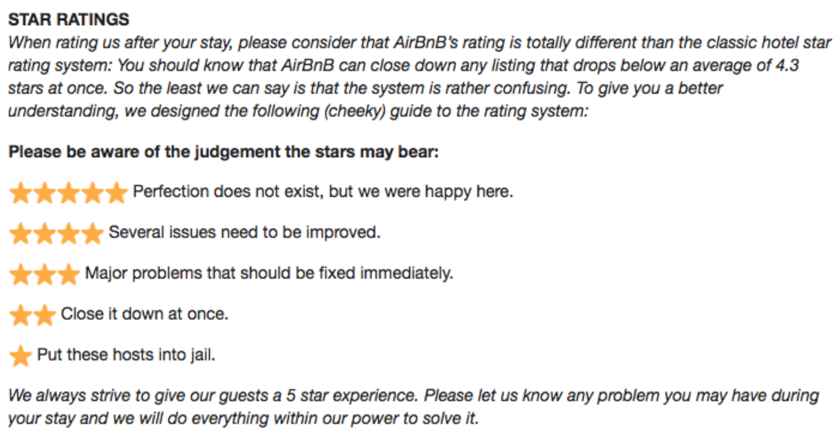
[19] Source "Vacation Rental Market Size, Share & Trends Analysis Report...," https://www.grandviewresearch.com/industry-analysis/vacation-rental-market, accessed 4/22/2022.

[20] According to Airbnb website accessed on 4/22/2022, the platform contains 6M listings and 4M+ hosts, see https://news.airbnb.com/about-us/.

[21] See Fradkin, Grewal, and Holtz (2021)

also a significant driver of the platform's ranking and search algorithm.[22]

We postulate that overall rating is driven by "value," quality net of price, and raw quality experience, disregarding the price.[23] This study concerns Airbnb – the largest short-term rental platform. Airbnb solicits feedback using several frames in which you rate customer experience using a star rating. Figure 1 describes the star system communicated to the users by Airbnb. Our deliberation process builds on earlier models in which reviews are a function of raw quality (see Hu, Pavlou, and Zhang, 2006; Li and Hitt, 2008).


STAR RATINGS

When rating us after your stay, please consider that AirBnB's rating is totally different than the classic hotel star rating system: You should know that AirBnB can close down any listing that drops below an average of 4.3 stars at once. So the least we can say is that the system is rather confusing. To give you a better understanding, we designed the following (cheeky) guide to the rating system:

Please be aware of the judgement the stars may bear:

★★★★★ Perfection does not exist, but we were happy here.

★★★★ Several issues need to be improved.

★★★ Major problems that should be fixed immediately.

★★ Close it down at once.

★ Put these hosts into jail.

We always strive to give our guests a 5 star experience. Please let us know any problem you may have during your stay and we will do everything within our power to solve it.

Figure 1: Description of Airbnb star rating. Source and copyright Airbnb Inc., fair use doctrine.

The "overall rating" is the modal frame displayed next to the listing name in search results. It is an unaided frame that asks, "How was your stay at X's place?" Airbnb's star description communicates a scale of the star rating. In particular, they communicate that a 4.3 rating may lead to the listing suspension. This scale is Airbnb's business-as-usual at the time encompassing our study. We report the scale to provide the context in which our data was generated. The description of the rating system is obtained directly from the Airbnb website, and we do not directly manipulate this review system besides randomizing prices.

In addition to the unaided frame, the platform uses several aided frames that focus on distinct parts of the customer experience. Conveniently, from an empirical standpoint, multiple review frames will allow us to understand customers' thought processes when submitting

[22]The authors of this paper obtained scraped data on online search rankings and discovered a strong correlation between rating and search position even after controlling for property fixed effects.

[23]Imagine booking an extremely cheap short-term rental. After arrival, the room had a broken TV, but you still gave it 5 stars because of the bargain price. On the other hand, imagine paying extra for a sea-view room in a luxury resort. Despite the exorbitant price, you still gave the room 5 stars because of the magnificent sunset views.

feedback. Airbnb solicits reviews using 6 aided categories:

- Accuracy
- Check-in
- Communication
- Location
- Cleanliness
- Value

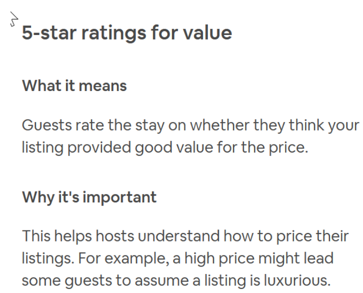


Figure 2: Screenshot from the Host Dashboard that explains the goal of the value rating solicitation. Source and copyright, Airbnb Inc., fair use doctrine.

The ratings in aided categories are not immediately visible to future consumers next to the listing name. Nonetheless, they are displayed on the "reviews" sub-page of the online listing. The purpose of the aided frames is to provide information along the specified dimensions. The first 5 categories aim to solicit various aspects of accommodation quality. For instance, the "location" category aims to improve "listing description" accuracy and asks, "Was the guest aware of safety, transportation, points of interest and special considerations like noise or other situations that might affect their stay?."[24] The last category, "value," has a different goal. According to Figure 2, the "value" rating aims to improve pricing. For example, low "value" ratings may suggest that the posted prices are too high. According to this description, a mere presence of the "value" ratings suggests a causal relationship between

[24]According to the platform "Host Dashboard", accessed on 3/14/2022.

prices and reviews. In addition, examining the "value" ratings allows us to determine the importance of consumer surplus during the feedback deliberation process.[25]

In the next Section, we present a stylized model of consumer preferences and deliberation that highlights the drivers of customer reviews.

3 Model

In this Section, we develop a stylized theory model of customers' deliberation. The model is mathematically parsimonious by design because its primary goal is to formalize the subsequent discussion of the empirical relationship between prices and reviews. In particular, the model highlights possible mechanisms of review generation and hypothesizes pathways via which transaction prices may affect customer satisfaction ratings. We reference these pathways when discussing empirical estimates.

Consider a continuum of customers and a single product. The product is a short-term rental accommodation for specific calendar dates at a specific property. Utility of the accommodation is given by

$$u = \alpha q - p,$$

where p is the price, and q is the vertical quality. Parameter α represents price-quality trade-off and varies across potential customers. Large values of α represent customers with a higher willingness to pay for quality. We assume that α is distributed with a CDF denoted by $F(\cdot)$.

The consumers may purchase the accommodation or outside option with utility $u_0 = 0$. By design, the utility model abstracts from many prominent features of the short-term rental market, such as horizontal differentiation[26] and competition. This decision allows us to highlight fundamental mechanisms linking transaction prices with subsequent reviews without introducing extensive analytical complexity.

As explained in Section 2, customer reviews are solicited using various frames. Consider

[25]Interestingly, Lewis and Zervas (2019) show that overall rating impacts demand but does not elicit supply-side price response – the impact of "value" rating on prices in an open question.

[26]Bondi and Stevens (2019) demonstrates that horizontal differentiation is an important factor driving reviews.

frame f with a review production function $R^f(u, q)$. We postulate that the review depends on consumer surplus u, capturing the "value" component, and q, capturing dependence on raw quality. It is reasonable to expect that R^f is weakly increasing in both components. The relative importance of u and q depends on a specific circumstance for generating feedback, which is synonymous with the review frame in our setting. This formulation deliberately ignores the multidimensional nature of product quality solicited by various review frames. However, varying degree of dependence on u and q enables us to isolate 3 distinct cases: (i) aided "value" frame, which places relatively more emphasis on u, (ii) aided quality frames, which places relatively more weight on q, and (iii) unaided "overall" quality frame that is potentially a more balanced function of u and q. We shall omit the f superscript for the remainder of the paper to simplify the notation.

Consider a scenario in which the buyer is aware of the quality q before the purchase. Further, consider a frame f that does not depend on consumer surplus, u, but is entirely determined by quality, q. It is straightforward to notice that transaction price does not affect the review in such a frame. Thus, under complete information, the dependence of review on the transaction price would indicate that consumer surplus is related to review deliberation. We expect this to occur in the "value" frame, but it may or may not occur in aided quality and unaided frames.

There are two mechanisms via which price influences the review. Consider a price increase from p to p'. Infra-marginal consumers with $\alpha > \frac{p'}{q}$ purchase the product before and after the price increase. If a review frame R depends on u, infra-marginal consumers will leave a lower review under price p'. We denote this as *consumer surplus effect* because it results from a fixed, infra-marginal consumer changing their review rating because of a lower surplus.

The second mechanism is *advantageous selection* of buyers. In most markets, only consumers who purchased the product may complete the review. A price increase selects buyers (and reviewers) with a higher preference for quality, hence truncating the distribution of α. Consumers having α, such that $\frac{p'}{q} > \alpha > \frac{p}{q}$, would have purchased with the lower price, p, but do not purchase with the higher price, p'. Consequently, an increase in price removes buyers with the lowest u, potentially increasing the reviews. The final impact of a price hike on reviews is a combination of the consumer surplus effect and advantageous selection.

Higher prices may increase or decrease reviews depending on the relative strength of both effects.

We examine the consumer surplus effect and advantageous selection mechanisms formally. To further simplify the argument, we fix product quality to 1 and drop the direct dependence of R on q. We also assume that R is differentiable in u. The average review is given by

$$\int_p^\infty R(\alpha - p)\frac{f(\alpha)}{1-F(p)}d\alpha$$

Taking the derivative with respect to p, we obtain that review drops as price increases only if

$$-R(0)\frac{f(p)}{1-F(p)} + \int_p^\infty \left[R(\alpha - p)\frac{f(\alpha)f(p)}{(1-F(p))^2} - R'(\alpha - p)\frac{f(\alpha)}{1-F(p)}\right]d\alpha < 0$$

After some manipulations, we obtain

$$\int_p^\infty \left\{\big[R(\alpha - p) - R(0)\big]H(p) - R'(\alpha - p)\right\}f(\alpha)d\alpha < 0,$$

where $H(p)$ is the hazard rate of the marginal consumer. The first term inside the integral represents advantageous selection after removing marginal consumers. The larger the hazard rate $H(p)$, the more marginal consumers are removed after raising p, and more mass is allocated to infra-marginal consumers. The infra-marginal consumers leave higher reviews since $R(\alpha - p) - R(0) > 0$. The second term represents the consumer surplus effect for infra-marginal consumers. To obtain the negative impact of price hikes on reviews, it suffices that the consumer surplus effect is larger than advantageous selection on average. If we impose the inequality point-wise, we obtain a simple ratio, linking the review-sensitivity of infra-marginal consumers to the hazard rate of marginal consumers

$$\frac{R'(\alpha - p)}{R(\alpha - p)} > H(p), \forall \alpha > p.$$

Consider a specific case where $R'(\alpha - p) \approx 0$ if $\alpha - p > \bar{u}$. In this scenario, consumers with large enough surplus base their reviews predominantly on the raw quality, q, formally

$R(\cdot, q) \approx R(q)$. This situation is more plausible when the effect of a price increase on surplus is minimal, particularly for consumers who exhibit lower price sensitivity. In the context of short-term rental (STR) accommodations, we might anticipate this behavior for luxury listings or among consumers in the luxury segment or those with greater wealth. In these instances, advantageous selection may dominate as price increase excludes more price-conscious consumers. In other words, as price increases, the pool of buyers may shift towards the luxury segment, which places a premium on inherent quality. One may expect prices to positively influence reviews if this effect is large. Overall, the direction of the effect is an empirical question.

Next, we consider a scenario with asymmetric quality information. The STR platform communication presented in Figure 2 indicates the presence of such asymmetric information. The platform states that "...a high price might lead some Guests to assume that listing is luxurious." Such a statement presumes that (i) the quality of accommodation is unknown, and (ii) the market is in a (partially) separating equilibrium where the posted price is a signal of quality. We consider a fully separating case, as in Wolinsky (1983). Fully separating equilibrium generates a deterministic consumer belief quality schedule, $Q(p)$,[27] where p is the posted price. We postulate that $Q(p)$ is weakly increasing.

In the following analysis, we abstract from general equilibrium effects by which quality schedule $Q(p)$ changes. Instead, we aim to highlight incentives for unilateral price deviations. Due to the modest size of each host and their limited ability to influence market-level beliefs, we expect the schedule in our data sample to remain unaffected by our pricing manipulation.[28]

As before, we set the true quality equal to 1 and consider a price increase from p to p'. We also assume that the actual quality is revealed after the purchase so that the review is given by $R(\alpha q - p)$. Let $G(p) = \frac{p}{Q(p)}$ be the lowest α at which the product is bought. The

[27]Extension to stochastic posterior beliefs is possible but not pursued. Such an extension delivers similar qualitative insights with extra mathematical complications; see Cooper and Ross (1984) for helpful discussion.

[28]Field experiment design is compatible with assuming away general equilibrium effects. In particular, the experiment utilizes minor idiosyncratic deviations from the equilibrium price, which are unlikely to ripple in general equilibrium.

average review is given by

$$\int_{G(p)}^{\infty} R(\alpha - p)\frac{f(\alpha)}{1 - F(G(p))}d\alpha$$

Taking the derivative with respect to p, we obtain that review drops as price increases, only if

$$-R(G(p) - \alpha)\frac{f(G(p))}{1 - F(G(p))}G'(p)+ \int_{G(p)}^{\infty}\left[R(\alpha - p)\frac{f(\alpha)f(G(p))G'(p)}{(1 - F(G(p)))^2} - R'(\alpha - p)\frac{f(\alpha)}{1 - F(G(p))}\right]d\alpha < 0$$

After imposing the equilibrium constraints $Q(p) = q = 1$ and that $G(p) = p$ we obtain

$$\int_{p}^{\infty}\left\{\big[R(\alpha - p) - R(0)\big]G'(p)H(p) - R'(\alpha - p)\right\}f(\alpha)d\alpha < 0, \qquad (1)$$

The effect depends on $G'(p)$, which embodies the notion that "...a high price might lead some guests to assume that listing is luxurious." The result depends on how a price adjustment affects market share. In an extreme instance of an inferior good, a price increase may result in a greater market share. This occurs if $G'(p) < 0$. In such a case, the positive impact of higher prices on beliefs about quality is larger than its negative impact on utility. Such upward-sloping demand may occur for some luxurious accommodations that rely predominantly on the price to signal quality. In that case, the negative impact of prices on reviews is guaranteed. The negative effect should be particularly pronounced in "value" review frames that put more relative weight on consumer surplus.

In the case of $1 > G'(p) > 0$, the effect of beliefs mutes customer selection compared to complete information. In this instance, we may observe the negative impact of price on reviews, even if the impact would have been positive with complete information.

The case $G'(p) > 1$ cannot occur when $Q(p) = q = 1$ and if the expectations are upward sloping, $Q'(p) \geq 0$. To see that, examine the inequality by plugging the formula for $G'(p)$ to obtain

$$\frac{Q(p) - pQ'(p)}{Q^2(p)} > 1.$$

The inequality implies a contradiction of $Q'(p) < 0$.

The immediate observation is that asymmetric information introduces additional negative pressure on reviews, which may either guarantee an adverse effect or, in more marginal cases, flip the direction of the effect. The possible interactions of private information and the negative impact of price hikes on reviews rationalize why platforms collect and communicate the "value" frame and how this frame may be diagnostic of overpricing.

In concluding our theoretical exploration, we now outline our empirical strategy derived from multiple testable predictions formulated from our theoretical model:

1. The relationship between the transaction price and reviews is theoretically ambiguous. Depending on specific economic primitives, a positive and negative relationship can emerge. One can estimate the average treatment effect on reviews when exogenously manipulating transaction prices to discern the average direction of this effect. Perhaps more interestingly, the varied nature of the effect can further be highlighted by estimating the distribution of conditional treatment effects.

2. When price has a large effect on consumer surplus, its influence on reviews will be negative. This hypothesis can be confirmed by investigating how price impacts reviews among customer segments that show high price elasticity, e.g., demand for less upscale properties or purchases of customers with lower income brackets.

3. Under high asymmetric information, the impact of price on reviews will be negative. To validate this theory, one can examine the relationship between price and reviews for customers likely to have more diffused quality priors.

4. The model outlines clear expectations regarding the interplay between signaling and consumer surplus effects. These effects act as substitutes, given that G' and $-R'$ are combined additively in equation (1). For example, in scenarios where signaling or consumer surplus effects are large, either effect alone can ensure a negative influence of reviews on the price. Testing this prediction is possible through the analysis of interactions between variables that modulate asymmetric information and income.

Expanding on the last point, it is helpful to recognize that both the effect of consumer surplus and asymmetric information can produce similar empirical results. Either factor might lead to a negative relationship between price and reviews, and conversely absence of any of these factors can result in a positive relationship. Therefore, if empirical results show a negative effect, it could indicate substantial surplus effects, significant asymmetric information, or both. However, as pointed out earlier, we can gain insight into the relative influence of these effects by evaluating and ranking the significance of various moderators. An apparent moderator of the surplus effect is income. Thus, particularly insightful uncertainty moderators are those with a weaker association with income, such as experience on the platform.

In the next section, we describe the specific setting of our investigation and the data.

4 Data

The data was obtained from Keybee, a company that manages short-term rental properties. Compared to long-term rentals, short-term rentals require significantly more resources to manage. Extra requirements may include replying to daily guest messages, calendar management across platforms, and scheduling turn-overs. Keybee is a management company that property owner hires to take over their STR platform accounts, including Airbnb, VRBO, and Booking.com. For this study, we have obtained data from Airbnb traffic.

Notably, Keybee does not directly market or sell the accommodations and utilizes the STR platforms to generate all reservations. Keybee acts in place of hosts controlling their platform accounts and receives a commission on rental revenue. Accounts are managed through programmatic access granted by official partnerships with the platforms. When connecting the Airbnb listing to Keybee, the online ad on Airbnb does not change; that is, Keybee's branding is not added to the listing, and the listing is still branded with the host's name and photo. As a result, guests are unaware of Keybee when interacting with the platform, such as when making reservations or providing feedback.[29] In other words, Airbnb is the context of this study, and Keybee should be considered a mere data provider. Specifically, the customer experience context, including the review system, is comparable to

[29] Keybee offers web and mobile applications to hosts. Hosts may use this software to order Keybee services, such as cleaning and house maintenance, check their unified calendar across platforms, and obtain business analytics. Guests do not directly use Keybee software.

other studies of the Airbnb marketplace.

The pricing infrastructure of STR platforms enables elaborate dynamic pricing. For example, the host can set the nightly rate separately for each calendar date and change these prices as the reservation date approaches. The complexity of this decision is similar to Airline pricing. To explain the pricing further, consider a fixed day of the year, say December 1st, 2020. Guests looking for 1-night reservations from December 1st to December 2nd may arrive at the platform any time before December 1st. Under full dynamic pricing, the price of this given reservation would vary depending on the date of the reservation query. For example, a consumer searching for December 1st on November 1st may see a different price than a consumer searching for the same date on November 2nd. Thus, pricing data is a triple panel reflecting a listing, reservation date, and price-setting date. In the remainder of the paper, we refer to this triple as *pricing event*.

One of the services offered by Keybee is dynamic programmatic pricing. The proprietary algorithm performs daily updates of nightly rates for each unrented calendar day, that is within the next 2 years of today. The updates reflect three main factors: (i) the change in the elasticity of demand as the reservation date approaches, (ii) decreasing opportunity cost as the reservation date approaches, and (iii) the patterns of calendar utilization.[30] Conveniently, Keybee's pricing algorithm contains an exploration module, which perturbs the daily price of the rental by a random multiplicative shock distributed uniformly on the interval $[90\%, 110\%]$. Moreover, the perturbation is identically and independently distributed (IID) across all pricing events. In other words, nightly rates are subject to a random pricing shock that varies daily.

Obtaining a causal impact of a price on subsequent reviews is usually tricky. Prices are set as a function of cross-sectional and over-time demand shifters, which may jointly determine product reviews. For example, luxurious vacation rentals are more expensive and obtain better reviews. Similarly, vacation rentals on the beach are more expensive in the summer and may obtain better reviews when warmer. This simultaneity may positively correlate prices and reviews, obfuscating causal associations. In the case of positive correlation, one

[30]The calendar date with open adjacent dates has different market potential than calendar dates that do not have open adjacent dates. The lower value of monetizing gaps in the calendar occurs because the single-night (or N-night) gaps do not show in search queries for reservations longer than one (or N) night(s).

is likely to underestimate the effect of pricing on reviews, which may lead to overpricing.

The bias is challenging to sign because the negative correlation between prices and reviews is also plausible. In periods of low demand, the competition between short-term rentals sharply increases because supply is relatively inelastic. Our conversations with the hosts indicate that more competition leads to lower prices, which may lead to more quality provision. For example, hosts frequently improve their listings by adding amenities like toiletries or free airport transit to obtain a competitive edge. They may also accelerate discretionary maintenance, such as repainting. The increase in quality is frequently aimed at obtaining better review ratings. This pathway is particularly concerning since one may find a negative relationship between prices and reviews, even if the causal link is absent.

Conveniently, Keybee's exploration pricing provides the field experiment partialling out these two effects.

Descriptive statistics

Keybee provided us the data on Airbnb listing characteristics, pricing events, reservations, and reviews. The data contains a random selection of reservations from 11/29/2018 until 08/26/2021 to obfuscate the cash flow and protect the hosts' privacy. We excluded less than 5% of reservations that did not result in a review.[31] Additionally, we excluded observations where the reporting system failed to record a pricing date due to reporting issues with the Google Big Query engine.

For every listing, we adjusted the posted daily rate for each successive night over a future 750-day window. Taking into account that some of these future dates were already booked, we conducted, on average, 282 price adjustments for each listing daily. While the experimental duration spanned 1,002 days, certain listings joined the dataset post 11/29/2018. Furthermore, a daily random exclusion saw a third of the listings omitted from the experi-

[31] No official data indicates how many Airbnb reservations lead to a review. Estimates from various bloggers place this number between 70% and 85%; hence, our review rate is likely higher than the Airbnb average. One contributing factor could be Keybee's automated communication, which encourages reviews. Another possibility is that Keybee listings are generally of superior quality since Keybee prioritizes professional hosts. The former mechanism is advantageous for us as it minimizes selection bias. However, the latter means our findings are more relevant to professional hosts than seasonal ones. It is crucial to note that the response rate is uncorrelated with experimental manipulation. See Dellarocas and Narayan (2006) for a deeper discussion on the incentives to leave reviews.

	Mean	SD	Min	Max	N
Total nightly fees	617.48	872.99	19.00	8,767.00	2,931
Price per night	277.51	370.23	19.00	2,922.33	2,931
Log-price per night	5.14	0.91	2.94	7.98	2,931
Number of nights	2.29	0.65	1.00	3.00	2,931
Bathrooms	1.86	1.10	1.00	8.00	2,931
Bedrooms	2.36	1.43	0.00	7.00	2,931
Review score (1-5)	4.77	0.67	0.00	5.00	2,931
Value score (1-5)	4.71	0.70	1.00	5.00	2,921
Check-in score (1-5)	4.90	0.46	1.00	5.00	2,921
Accuracy score (1-5)	4.81	0.62	1.00	5.00	2,923
Location score (1-5)	4.85	0.51	1.00	5.00	2,921
Communication score (1-5)	4.88	0.52	1.00	5.00	2,923
Cleanliness score (1-5)	4.81	0.60	1.00	5.00	2,923
Price random shock, 1st night	1.00	0.05	0.90	1.10	2,931
Price random shock, 2nd night	1.00	0.04	0.90	1.10	2,931
Price random shock, 3rd night	1.00	0.03	0.90	1.10	2,931
Low-price Arm	0.22	0.42	0.00	1.00	2,931
High-price Arm	0.16	0.37	0.00	1.00	2,931

Table 1: Descriptive statistics. Descriptive statistics derived from the estimation sample. An observation is a reservation.

ment. This methodology culminated in roughly 93,000 listing-days, producing a total tally of over 25 million price adjustments. However, despite this vast number of price adjustments, the daily conversion rates remained modest. In simpler terms, not every price change guaranteed a reservation (e.g., a listing for December 1st might not necessarily be booked precisely on November 1st). After implementing the above selection criteria, conducting random reservation sampling and removing reservation longer than 4 days, our final dataset comprised 2,987 reservations.

We restrict the sample to stays of 3 nights or less since our estimate loses statistical power for longer reservations – a drop in statistical power results from independent experimental manipulations applied at each posted daily rate. For longer reservations, the manipulations average out and have little effect on the total price. Conveniently, considering shorter bookings allows us to avoid modeling long-term discounts.

Table 1 provides descriptive statistics for the data set. The overall cost of the reservation consists of two components. The cost consists of *cleaning fee* and nightly fees. The cleaning fee is paid regardless of the number of nights booked and is fixed over time. Nightly fees per reservation average \$617, and there is considerable variation across listings. The cheapest recorded fee is \$19, while the most expensive fee is approximately \$9,000. The average nightly rate is \$277, and the average reservation length is 2.3 nights. The data set contains a variety of listings. For example, the average number of bedrooms is 2.36; however, our data contains studios (0 bedrooms) and 8-bedroom listings.

The second panel of Table 1 displays average review scores. The overall review score averages 4.77. According to our sources inside Airbnb, this is somewhat larger than the average review on the platform.[32] Interestingly, the aided review frames deliver higher average review ratings than the overall rating.

We examined pair-wise correlation coefficients between various review scores to investigate if the review frames solicit independent quality information. The "overall score" has a pair-wise correlation coefficient that averages 0.61 across frames and ranges between 0.75 and 0.43. Interestingly, the most considerable correlation is with the "value" score, which suggests that consumer surplus is essential to the overall post-purchase evaluation. Similarly, the "value" score has correlation coefficients that average 0.55 across frames and range between 0.75 to 0.36. The modest correlation suggests that consumers pay attention to the frames instead of submitting the same or similar score in each frame. Low correlation is also consistent with the noisy cross-frame valuation; however, this noise is less crucial because it will subsequently work against us, leading to attenuation. Finally, the most orthogonal score is obtained in a "check-in" frame, which solicits feedback on a narrow portion of the customer experience. Narrow focus and lower correlation again support that frames succeed in soliciting feedback along various quality dimensions.

[32]Our data set has a larger-than-average review score than Airbnb because most of our listings are professional hosts, who typically receive better satisfaction scores. In addition, outsourcing management tasks to Keybee generally increases service quality. Consequently, our estimates are conservative because they are subject to ceiling effects as described by Everitt (2010).

The third panel of Table 1 contains the descriptive statistics of the pricing experiment. As mentioned above, the data contains reservations shorter than 4 nights; thus, each reservation is subject to at most three pricing shocks. Denote a multiplicative shock to the t-th reservation night as ζ_t. If the reservation date is shorter than 3 nights, we set the absent shocks to 1 to balance the panel.[33] Setting missing shocks to 1 explains why the standard deviation of ζ_t decreases as t increases.

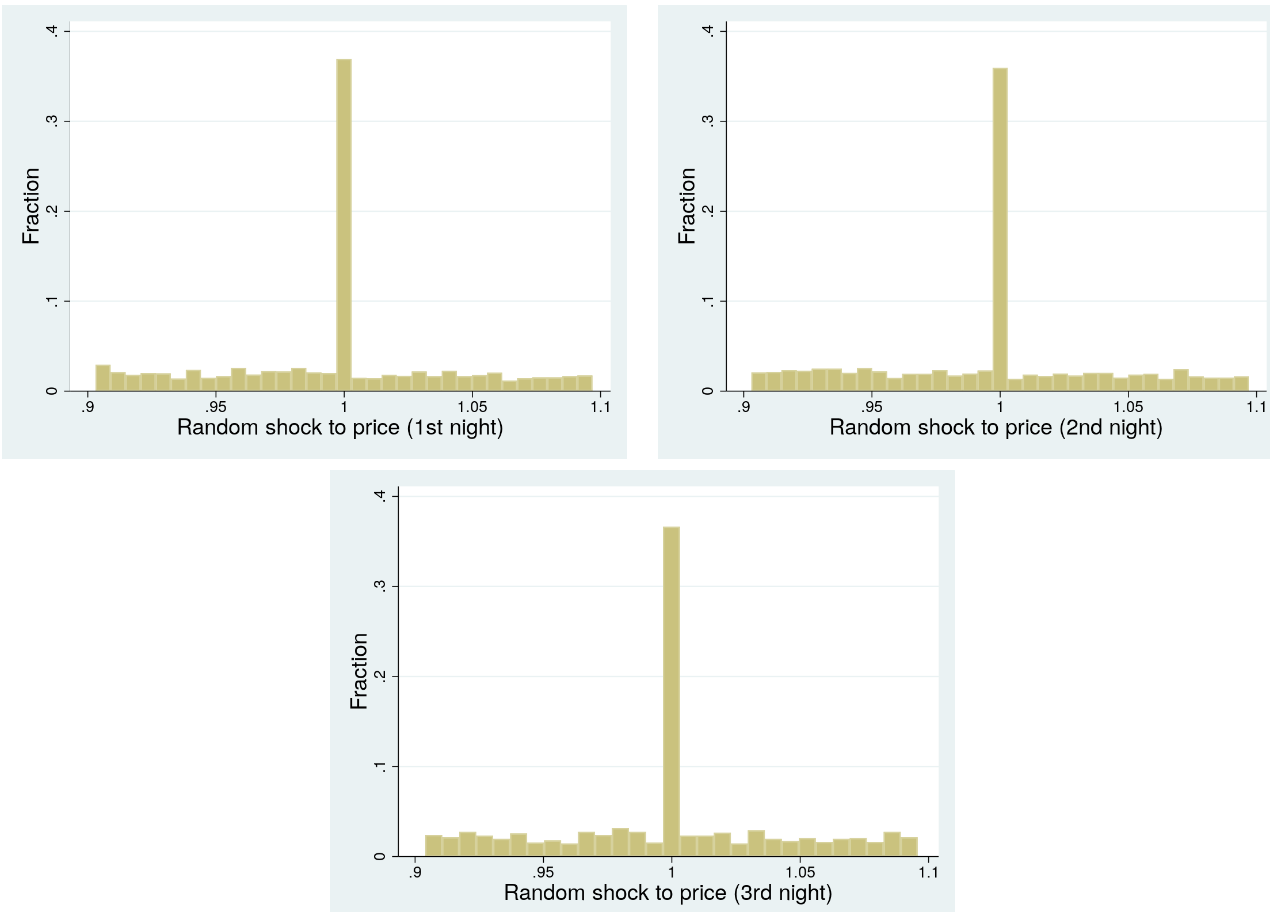


Figure 3: Distribution of the exploratory random price shocks. Second-day nightly fee shock graph excludes 1-day reservations. Similarly, the third-day nightly fee shock graph excludes 1-day and 2-day reservations. One-third of pricing events were randomly excluded from the experiment at random which resulted in a large mass at 1.

Figure 3 contains a histogram of pricing shocks. The histogram exhibits the mass at 1, reflecting that approximately one-third of pricing events are randomly excluded from the exploratory module. The pricing shocks are, as expected, uniformly distributed. To obtain

[33]Pricing shocks for adjacent dates may affect the reservation demand since guests may choose between stays of various lengths. We repeated the analysis using the actual pricing shocks for the missing dates instead of 1 and obtained the same results.

model-free evidence, we consider two pricing arms. We define a low (high) pricing arm as reservations that have 2 or more shocks that are negative (positive) and single-night reservations with negative (positive) pricing shocks. In our sample, 22% of reservations are in the low pricing arm, and 16% are in the high pricing arm. Low and high pricing manipulations are equally likely – the difference in the occurrence percentage results from the downward-sloping demand for reservations[34]. As discussed by our theoretical model in the previous section, downward-sloping demand results in customer selection across arms according to customers' willingness to pay. In particular, since arms are randomized, customers who bought in the high-pricing arm are willing to pay more than those who bought in the low-pricing arm. This difference would be essential to our story, supporting the *advantageous selection* hypothesis.

	(1) ζ_1	(2) ζ_2	(3) ζ_3
Bathrooms	-0.00175	0.000562	0.000619
	(0.00144)	(0.00137)	(0.000900)
Bedrooms	0.000400	0.000512	0.0000888
	(0.00110)	(0.00105)	(0.000687)
Reviewer count	-0.00000879	-0.0000383	0.00000220
	(0.0000181)	(0.0000373)	(0.0000114)
Guest median city income	9.29e-09	3.01e-08	2.73e-08
	(2.85e-08)	(2.71e-08)	(1.78e-08)
Distance	-0.00000115	-0.000000272	0.00000146**
	(0.00000103)	(0.000000980)	(0.000000644)
Non US	0.00329	0.00445	0.00188
	(0.00353)	(0.00337)	(0.00221)
N	2931	2931	2931

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Randomization checks.

Table 2 displays the results of randomization checks. We perform separate regression

[34]The difference in quantity between arms is significant at 1% level, which serves as a formal test for a downward-sloping demand curve. This result will prove useful in a subsequent analysis of the impact of reviews on demand.

for each ζ_t shock against the listing characteristics, including the number of bedrooms and bathrooms, and against guest characteristics, such as the number of reviews obtained by the guest, median income of the guest's home city, and the distance from the home city to the rental (measured in miles). Of the 18 coefficients, only one is statistically significant; however, its economic magnitude is negligible. One positive 5% test could likely be an outcome random chance due to multiple comparisons. Employing a simple Bonferroni correction nullifies its significance. We further conducted a t-test across different pricing arms and achieved congruent findings.

In the succeeding section, we present model-free and model-assisted results.

5 Results

This section provides evidence of the relationship between price and reviews. We examine impact across several review frames, focusing on the *overall rating* and *value rating* frames. The former is the only review frame immediately visible next to the listing name; thus, it should most significantly impact subsequent demand. The *value frame* is relevant to testing theory since it captures the effect of price hikes on reviews via consumer surplus. We conclude our argument by showing that reviews affect the hosts' bottom line, i.e., we show that the most recent review shifts demand curve for reservations. We start by presenting model-free evidence, which is a series of t-tests.

Model-free evidence

Figure 4 compares histograms of reviews across pricing arms. We note that the reservations in the high-price arm obtain a smaller fraction of 5-star reviews. The displaced 5-star reviews are distributed across 1- to 4-star reviews. The highest increase is present for 4-star reviews; however, we also observe a noticeable increase for 1- and 2-star reviews. The increase amongst the reviews with the lowest star rating is the most damaging to hosts because it prevents them from obtaining a *superhost* badge. The host obtains a superhost badge if their average rating is 4.8 and higher in the prior year. For example, obtaining merely a single 2-star

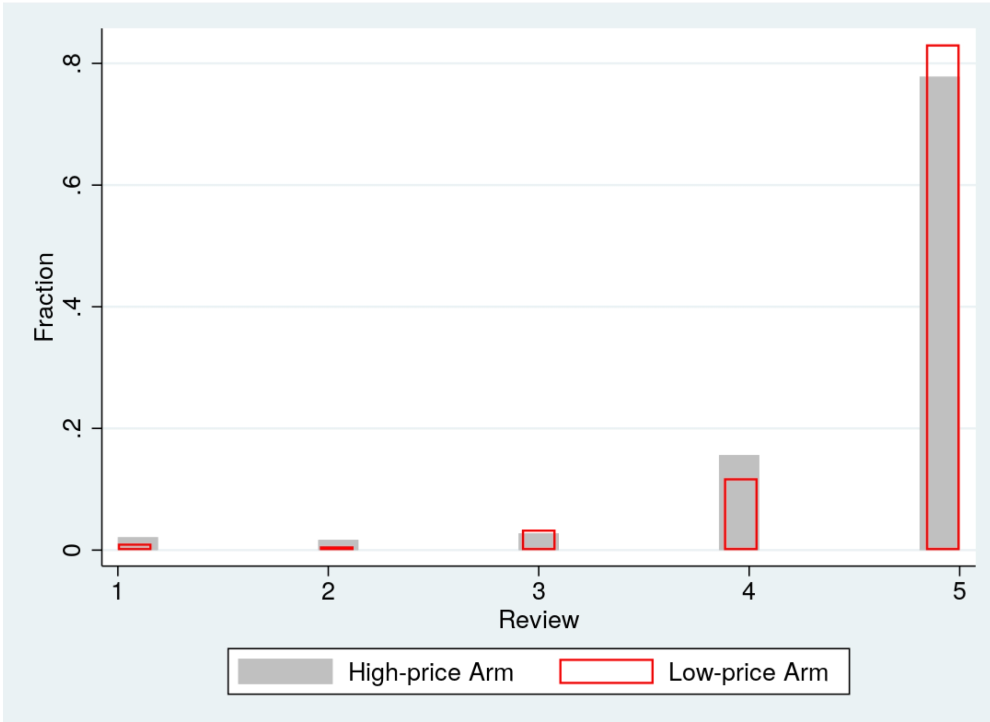


Figure 4: Histogram of reservation reviews across pricing arms.

review requires an offset of at least 14 subsequent 5-star reviews to maintain the badge.[35]

	(1) Guest review score (1-5)	(2) Guest Value score (1-5)	(3) Guest Check-in score (1-5)	(4) Guest Accuracy score (1-5)	(5) Guest Location score (1-5)	(6) Guest Communication score (1-5)	(7) Guest Cleanliness score (1-5)
High-price Arm	-0.0900**	-0.0998**	-0.0000454	-0.0862**	-0.0799**	-0.142***	-0.0680**
	(0.0418)	(0.0432)	(0.0260)	(0.0402)	(0.0343)	(0.0359)	(0.0340)
N	1132	1125	1125	1127	1125	1127	1127

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Model-free evidence. The table displays the average difference in the review following high-pricing vs. low-pricing arms transactions.

We conduct a t-test to compare average reviews across pricing arms. The results are presented in Table 3. The overall rating is 0.09 lower in the high-pricing arm. In addition,

[35]The Superhost badge on Airbnb significantly influences demand through several channels. Primarily, Airbnb's search ranking depends on the badge. Second, the status enhances search visibility because Airbnb allows users to filter for superhost listings, potentially sidelining hosts without the badge. Third, the badge is visibly placed next to the host's profile, serving as a symbol of reliability. According to Airbnb, Superhosts earn 64% more than other hosts, although this figure could partially reflect a correlational relationship. It is also worth noting that approximately 20% of Airbnb hosts possess Superhost status. While this indicates that achieving this status is a realistic goal for many hosts, it also means that absence of the badge could potentially be interpreted negatively by potential guests. Source: https://hospitable.com/airbnb-superhost-status-is-it-really-worth-it/

the value rating decreases by a comparable amount, indicating that diminished consumer surplus is a significant factor in the total rating decline. Interestingly, almost all other quality frames, except for *check-in* frame, reflect statistically comparable impact. These similarities suggest that transaction price affects surplus across many quality dimensions. One consistent mechanism that rationalizes these effects is formalized in our model with asymmetric quality information. For instance, consumers may infer the quality of, say, guest communication or cleanliness from the posted price, which leads to inflated beliefs of guests in the high-pricing arm and possible ex-post regret.

A handy frame to gauge asymmetric information is the *location* frame, as it measures description accuracy (this way of understanding the location rating is communicated to the guests during the review process). Since the listing description itself is unrelated to the pricing arm, it must be that guests update their priors about description accuracy using the posted price in addition to the actual listing description.

Analyzing the pricing arms provides a model-free method to demonstrate the causal influence of prices on reviews, requiring minimal assumptions about the inherent data generation mechanism. If the primary objective was merely to ascertain the existence of this effect, this examination would suffice. In the following subsection, we delve deeper by employing a regression model to evaluate the marginal effect of price on reviews.

Model-aided evidence

Consider the following linear model:

$$r_{it} = \alpha + \beta X_{it} + \gamma \log p_{it} + \epsilon_{it}, \tag{2}$$

r_{it} is the star rating, X_{it} denote time and listing covariates. These covariates account for characteristics of the period the booking was done, such as the number of days leading up to the reservation, represented as a quadratic function, and day of the week-, month-, and year-fixed effects for the check-in date and the booking date. Listing attributes are captured through dummy variables, reflecting the number of bedrooms, beds, and the associated zip code. In two specifications, we substitute listing covariates with listing fixed effects. The parameter γ encapsulates the causal impact of a price change on reviews.

	(1) Guest review score (1-5)	(2) Guest review score (1-5)	(3) Guest review score (1-5)	(4) Guest review score (1-5)	(5) Guest review score (1-5)	(6) Guest review score (1-5)
Log-price	0.122***	0.116***	-0.0273	-0.0859**	-1.602**	-0.851*
	(0.0116)	(0.0120)	(0.0339)	(0.0350)	(0.808)	(0.496)
Date Controls	no	yes	yes	yes	yes	yes
Listing Controls	no	no	yes	no	yes	no
Listing FE	no	no	no	yes	no	yes
Experiment	no	no	no	no	yes	yes
N	2929	2929	2924	2929	2924	2929

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Impact of pricing on subsequent reviews. Models (1)-(4) are OLS regressions that utilize organic and experimental variation in prices. Models (5)-(6) are 2SLS regressions that use only experimental variation in prices.

The demand shock for lodgings, denoted as ϵ_{it}, allows for cross-sectional and time series variation. Cross-sectional differences among properties encompass geographical location and available amenities, such as the caliber of furnishings or the presence of a fitness center. Dynamics specific to an individual listing, shifting over time, may range from demand seasonality to aspects like depreciation in quality, faulty amenities, and activities tied to refurbishments or renovations. After the stay, these demand shifters directly influence reviews, as in equation (2), because guests accommodated in higher-quality lodgings are predisposed to furnish more positive evaluations. However, most of these demand determinants are observable to the property host and the Keybee staff and are inherently mirrored in the listed prices. Consequently, transaction prices p_{it} are correlated with ϵ_{it} which results in an endogeneity. The magnitude of this challenge is contingent on the proportion of price variance that remains uncorrelated to the demand shifters.

Algorithmic pricing mechanisms can amplify the endogeneity problem because the algorithms are adept at recognizing and adjusting to demand changes, some of which might elude human observation. However, it is helpful to note that the endogeneity bias is not exclusive to markets employing algorithmic pricing. For instance, in traditional markets, it is acknowledged that leveraging cross-sectional price variation is often not the most robust approach for estimating demand elasticity. The limitation arises because prices can be influenced by quality aspects of products that are unobserved to the econometrician (as detailed in Berry, 1994). A common strategy to address this challenge is incorporating product- or brand-specific fixed effects. Nevertheless, even with the inclusion of these fixed effects, an econometrician requires residual price adjustments for a given product, either temporally or across markets, to obtain identification. It is often challenging to determine the exogeneity of this residual price variation in the absence of marginal cost data and without making strong assumptions about the supply side, like the exogeneity of product characteristics. Consequently, adopting experimental price variation, if available, emerges as a preferred approach to bypass these complications. This method fortifies the estimation of the parameter γ and is versatile, extending its applicability to many markets beyond short-term rentals.

Notwithstanding the potential bias, estimating intermediate models with limited controls for endogeneity helps obtain a measure of the data variation. The estimated coefficient of the most straightforward model without any controls is presented in Column (1) of Table 4. We obtain a positive and statistically significant correlation between log price and reviews. The coefficient on the log price amounts to .122, meaning higher prices are associated with higher ratings. The correlation is possibly spurious for the reasons mentioned earlier.

To control for aggregate seasonality, we add date controls. In this specification, we detect an insignificant drop in the price coefficient compared to the model with no controls. Further, we include listing controls. After this addition, we observe a significant decline in γ, indicating that using cross-sectional variation leads to overestimating γ. Next, we include listing-fixed effects to control for a time-persistent unobserved heterogeneity. In this approach, the coefficient flips sign to negative, indicating that within-listing, higher prices are related to lower reviews. Nevertheless, the economic significance of this effect is only modest because a 10% price hike leads to only a 0.01-star rating decrease. However, as mentioned above, these estimates may still be biased downwards because price variation after partialling out cross-sectional differences may remain, to a large extent, endogenous.[36]

[36] It is helpful to highlight that the OLS regressions already make use of our experimental variation, even though the experiment has not been explicitly integrated into our econometric specification. The price variation utilized by the OLS models is a blend of the endogenous variation that remains after controlling for the covariates and the experimental variation. As a result, the bias on the OLS coefficient in our setting is likely to be smaller than in other settings, applying similar specifications using data not subject to an

Model (5) represents the IV specification with date and listing controls. Prior to analyzing the results, we perform a test for weak instruments by calculating a joint F-statistic for the three experimental shocks, ζ_t. The resulting p-value of 2.3% indicates the statistically significant explanatory power of the instruments even after partialling out the covariates.

The IV estimator results in a point estimate of -1.6. This model is our favored specification due to its exclusive reliance on the experimental variation in price while possessing necessary statistical power. The derived estimate implies a decrease of 0.16 in the star rating, or about 23% of its standard deviation, following a 10% increase in price. Economically, this effect is significant, particularly when considering the tendency for review compression at the top end in platforms like Airbnb. For example, a sustained 10% price elevation could likely lead to the loss of the superhost badge, which mandates an average rating of at least 4.8. A more pertinent price change can be discerned by examining the average absolute price deviation within each listing. Our observations indicate that prices fluctuate from the corresponding listing average by 23%, which translates to a change of 0.3 stars, or nearly half a standard deviation in ratings. In other words, a typical price fluctuation for a given listing accounts for nearly half of a standard deviation in ratings. We discuss further economic implications of these results later in this section.

Model (6) introduces an alternate IV specification with listing-fixed effects which may be useful for direct comparison with Model (4). Importantly, fixed effects are not necessary for removal of endogeneity because Model (5) already isolates experimental variation. Model (6) implies an 11% standard deviation drop in reviews after increasing the price by 10% (significant at a 10% level). Despite its usefulness for straightforward comparison with Model (4), Model (6) is not our preferred specification due to the considerable loss of precision, likely owing to the excessive inclusion of covariates (in this case, fixed effects).[37] An order of magnitude difference between estimates of Model (6) and Model (4) highlights the potential of underestimation of the effects when using fixed effects without IVs, even in the presence of substantial experimental price variation in data.[38]

experiment.

[37] Freedman (2008); Negi and Wooldridge (2021) discuss the cost of including covariates in the experimental regressions.

[38] In an earlier draft of this paper, we explored a third IV approach, mirroring Model (6), but it was applied to data that omitted 1-day reservations. The reasoning behind this exclusion was that 1-day reservations

	(1) Guest review score (1-5)	(2) Guest Value score (1-5)	(3) Guest Check-in score (1-5)	(4) Guest Accuracy score (1-5)	(5) Guest Location score (1-5)	(6) Guest Communication score (1-5)	(7) Guest Cleanliness score (1-5)
Log-price per night	-1.606**	-1.366*	0.221	-1.857**	-0.663	-1.560**	-0.713
	(0.810)	(0.760)	(0.381)	(0.857)	(0.508)	(0.731)	(0.566)
Date Controls	yes	yes	yes	yes	yes	yes	yes
Listing Controls	yes	yes	yes	yes	no	yes	yes
Listing FE	no	no	no	no	no	no	no
Experiment	yes	yes	yes	yes	yes	yes	yes
N	2926	2916	2916	2918	2916	2918	2918

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: The impact of pricing on subsequent reviews using only experimental variation across various review frames. The table presents IV regressions.

We also examine the impact of log price on the aided review frames using our preferred IV specification as in Model (5). Results are presented in Table 5. We find the impact of the price on the *value*, *accuracy*, and *guest communication* reviews; nevertheless, the impact on aided frames is smaller than the impact on the overall score. This difference suggests that guests do not disregard the aided frames and penalize expensive accommodations in the frames corresponding to perceived deficiencies. As before, the impact of the *value* score is of comparable magnitude to the *overall* score, which provides further evidence for the consumer-surplus-based scoring rule discussed in the theoretical section. Nevertheless, we also find a significant impact on the *accuracy* score, which supports the mechanism relying on asymmetric information.

5.1 Heterogeneous treatment effects

In Section 3, we lay out a theoretical model generating the identified relationship between prices and reviews. At the end of that section, we posit several empirical tests using moderators of the treatment effects. We collected several possible moderators of the pricing effect that we classify in the 3 groups presented in Table 6. We discuss each group below.

tend to have a more varied spread in reviews. Omitting these short-term reservations resulted in a similar point estimate with an improved statistical precision.

Listing covariates	Asymmetric information moderators	Demand elasticity moderators
Picture count	Customer ID	City median income
Word count	Review count	City unemployment rate
	Hosted listing count	*Distance*
	Superhost	
	Distance	

Table 6: Reservation covariates divided into groups.

Listing covariates

This group comprises variables at the property level that influence the information content of the Airbnb listing. Specifically, we gathered data on the number of photos included in the listing and the word count of the listing description. It is plausible that these moderators correlate with the degree of uncertainty surrounding the listing quality; more photos and comprehensive descriptions will likely convey more information. The primary limitation of listing covariates is that they may be correlated with unobserved listing characteristics. For example, the number of photos could correlate with the luxury status of the property. We discuss the implications of this potential correlation later in this section.

Asymmetric information moderators

We collected guest-level covariates that impact the extent of information a guest possesses prior to booking. We operate under the assumption that seasoned Airbnb users inherently possess more pre-booking information. For instance, those who have previously used Airbnb can set their expectations distinctively compared to traditional hotel services—anticipating factors such as the absence of housekeeping or no supplies replenishments during their stay. Over time, they also hone their skills at identifying inconsistencies or potential red flags in listings and become proficient at extracting crucial details that enhance their booking decisions.

For each guest, we collected data related to their duration of use on the platform, as indicated by their account ID. Historically, Airbnb allocated account IDs in a sequential manner, which means that smaller IDs correspond to longer tenure on the platform. During our sample period, Airbnb shifted to using hash-based IDs for newer accounts; this change impacted less than 10% of the data. We use the decile of the account ID as an indicator of a guest's tenure, placing all hash-based IDs in the decile that corresponds to the most recent accounts.[39]

We gathered information on the number of reviews each guest had received. As previously noted, between 70% to 85% of hosts receive reviews. However, we were unable to find comprehensive statistics on the rate of host-to-guest reviews. In our dataset, we note that the host-to-guest review rate is approximately 98%. But, because Keybee has a process in place aiming to review every guest, this review rate might not be truly representative. It is worth noting that other hosts might adopt similar review practices, particularly if they are associated with management companies. Thus, it stands to reason that hosts might exhibit greater consistency in reviewing guests compared to guests reviewing hosts. Keeping these caveats in mind, we use the number of host-to-guest reviews as a proxy for the number of prior reservations for each guest. Our goal here is not to pinpoint the exact number of past reservations; rather, we aim to rank guests based on their experience, distinguishing between seasoned and novice users. A moderate difference between the actual number of reservations and reviews is acceptable for our analysis.

Moreover, we assessed whether a guest is also an Airbnb host and documented the number of listings they manage. It is plausible to posit that Airbnb hosts, seasoned in navigating listings, would harbor more refined expectations about listing quality. We can link hosting and renting because these activities typically emanate from the same account.

Furthermore, we acquired data on the proximity between a listing and a guest's home city, which is listed in their Airbnb profile. Many bookings originate from guests residing within the same state or city. Such bookings – including leisure "staycations," pragmatic stays during home repairs, or accommodations for visiting acquaintances – often result in reviews from account holders living nearby. Given their inherent knowledge of the locale, these guests likely possess a more precise prior about the stay quality. Sometimes, they might even inspect the property before confirming their booking.

[39] Airbnb displays the duration of a user's membership on the platform on their profile page, measured in years. However, we opted to utilize the ranking of Guest IDs instead of the number of membership years, as the latter does not differentiate between users who joined during the same year. Guest ID method aligns more closely with our principal aim – to rank users based on the length of their membership, rather than determining the exact length of time they have been members.

Guest wealth moderators

We obtained each guest's home city listed in the Airbnb profile. Further, we purchased demographic characteristics for each city from `simplemaps.com`. For our analysis, we need variables that affect demand elasticity, such as customer wealth. We select "median income" and "unemployment rate."

Considering its capacity to modulate prior expectations, we have also included the variable "distance to property" as a potential moderator of wealth. We posit that longer journeys are more expensive and are undertaken by more affluent clients. The wealth effect of distance is expected to manifest in a direction opposite to that of the information effect. Specifically, should distance predominantly influence wealth and the consumer surplus effect be substantial, we anticipate observing negative effects of prices on reviews for shorter distances. In contrast, if distance chiefly pertains to information and the signaling effect is predominant, we would expect negative effects for longer distances. Therefore, the distance may be handy in assessing the relative significance of consumer surplus versus signaling effects.

Data analysis

The objective of this Section is to delineate the distribution of treatment effects. We aim to utilize various methods that segment the data in complementary ways. The analysis begins with simple mean comparisons, facilitating an easily interpretable, model-free data description. The examination continues with the employment of causal forest (see Athey and Wager, 2019), enabling non-parametric estimation of the distribution of conditional average treatment effects (CATE). The utilized `econml` Python package, referenced by Battocchi, Dillon, Hei, Lewis, Oka, Oprescu, and Syrgkanis (2019), effectively integrates double machine learning as an inherent component in implementing causal forests. This methodology provides the necessary statistical power for a joint analysis of all moderators (see Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins, 2018, for discussion on double-machine learning). Ultimately, we discern which moderators contribute to more variation in treatment effects across diverse Airbnb transactions. As explained in the conclusion of our theoretical analysis in Section 3, the ranking of moderators is suggestive of the mechanisms underlying our effects.

Since the machine learning methodologies deployed for estimating heterogeneous treatment effects are optimized for binary treatment variables, we revert to examining raw data using binary experimental arms as defined in Section 4. Reiterating, the treatment variable equals 1 if the transaction occurs under the high pricing arm and 0 otherwise. The high price arm manifests as a random event when most daily price shocks for a given reservation are positive.[40]

Table 7 conducts an exploratory analysis of heterogeneous treatment effects using selected moderators. In this exercise, the moderators are scrutinized individually using mean splits. The first column of Table 7 contains the mean comparison of average reviews contingent on the pricing arm and the number of photos in the listing. We find that reviews of listings with fewer than average photos are negatively affected by high transaction prices. Conversely, listings with more than the average number of photos encounter no price impact on reviews, consistent with the information narrative. We also observe that listings with more photos consistently achieve superior ratings, hinting that such listings might be of superior quality. Considering the likelihood that high-quality listings might experience a less elastic demand, it is reasonable to infer that consumer surplus and signaling effects are concurrently at play. Conducting a more in-depth analysis of guest-level moderators would offer more precise insights, enabling the separation of these intertwined narratives.

Column (2) of Table 7 includes interactions with the Guest ID decile, where a higher ID decile indicates a newer Airbnb guest account. Our findings reveal that accounts with less tenure (IDs above the mean) encounter substantial negative treatment effects. This difference suggests that less experienced users leave lower reviews when paying higher prices. This result aligns with the expectation that less experienced guests have more diffused priors about listing quality. Another possible explanation could be that seasoned Airbnb users are more lenient reviewers, not penalizing listings for higher prices. However, this hypothesis is debunked by the data, as overall reviews do not depend on the tenure of the account.

[40]We can replicate our analysis, excluding machine learning, using the IV regression model. Nevertheless, we have decided to use a consistent methodology across this section. This decision is primarily due to the strong assumptions necessitated by using IVs to estimate heterogeneous treatment effects and the limited incremental insight it offers.

Column (3) examines an additional experience-related moderator: the total number of reviews a user has received on the platform. In this case, users with fewer reviews – and presumably, less experience – are subject to the negative impact of price on their ratings. In contrast, more experienced users, those with a substantial number of reviews, display a noticeable positive effect of price on their ratings, even though this effect is not statistically significant. The positive direction of these point estimates suggests that, under sufficiently tight priors, the selection effect might predominate over the consumer surplus effect. We plan to revisit this point when analyzing the moderators more granularly.

Columns (4) and (5) assess the impact of consumer wealth on the treatment effect. Our analysis demonstrates that guests from cities with higher incomes and lower unemployment rates do not experience the effects of prices on reviews. This null effect supports our hypothesis that diminished demand elasticity attenuates the effect of prices on reviews.

Thus far, our analysis supports both signaling and consumer surplus effects in creating a negative relationship between prices and reviews. We proceed to use causal forest to examine these narratives more comprehensively. The top panel of Figure 5 displays the distribution of CATEs, showing that 67% of the effects are negative. The distribution leans to the left, characterized by a long tail of negative effects and comparatively fewer substantial positive effects. The lower portion of the panel presents regressions of CATEs on two selected moderators, customer ID decile and city median income, mirroring our earlier OLS findings.

We calibrate a single decision tree using CATEs as the dependent variable to assess the importance of various moderators. This approach explains the variation in the estimated treatment effects by applying all the available moderators. The tree algorithm sequentially selects variable and their binary partitions that most enhance the explanatory power for predicting the treatment effect. The calibrated tree is illustrated in Figure 6, where each node corresponds to a subsample of our data, colored based on the value of CATE: red for a negative average effect of prices on reviews, green for a positive effect, and blue for a zero average effect. Standard errors are reported in brackets.

The tree's root node indicates a population CATE of -0.061, significant at the 1% level, yet again confirming the overall negative average effect of price on reviews. The root also contains the initial, most predictive data split: "Review count ≤ -0.107." The left child node represents data for users with a review count -0.1 standard deviation below the mean, and the right side of the tree represents the remaining, more experienced users. The importance of the experience reaffirms our regression analysis from Table 7, mainly that, on average, experienced users show almost no effect of price on reviews (CATE of -0.003), while inexperienced users exhibit significant effects (CATE of -0.116). Prior reviews, proxies for the number of past transactions, are determined to be the most crucial driver of effect heterogeneity, concluding that asymmetric information lies at the heart of our results.

Further analysis among less experienced users reveals that low experience nearly guarantees negative effects. The most inexperienced users (few reviews, newer accounts) renting properties with few photos generate a notably large treatment effect of -0.374. Notably, one can counteract the signaling effect but only in a niche 5% segment of users originating from areas with deficient unemployment. Additionally, we executed a regression analysis using Conditional Average Treatment Effects (CATEs) to examine substitution patterns between consumer surplus and signaling effects. In particular, we regressed CATEs on review count, guest ID, unemployment rate, and median income, as well as four interaction terms created between pairs of information and wealth moderators. The results from this analysis indicated that all interaction terms had point estimates suggesting a substitution effect. Moreover, two of the four interaction terms achieved statistical significance at the 1% level. When considered collectively, these results support our theoretical model, which anticipates the substitution between consumer surplus and signaling effects. In essence, the model suggests that if either of these effects is strong, the negative influence of prices on reviews is guaranteed.

Most null or positive effects are present on the right side of the tree. This observation suggests that experienced platform users generally do not experience a negative impact of prices on reviews. In most cases, experienced users see no effect, with two exceptions. Primarily, experienced users from extremely wealthy cities (more than 1.24 standard deviations above the mean income and less than 0.08 standard deviations below the mean unemployment) exhibit positive effects. In other words, higher prices result in better reviews for these users. This observation supports our theoretical model, which predicts that when asymmetric information is low and demand is inelastic, the selection effect will likely prevail. In such cases, removing marginal buyers impacts reviews more than the surplus loss of infra-marginal buyers (wealthy and experienced customers).

Secondly, we observe a null effect within the group of more seasoned yet less affluent users, except for instances of exceedingly lengthy trips. The decision tree algorithm has organized these instances using the squared distance criterion. However, it is essential to emphasize that no short-distance stays or staycations are categorized as "$\text{Distance}^2 \leq 4.52$," given that all these trips exceed the median distance. This pattern could be attributed to the idea that longer travel distances give rise to more uncertain initial expectations.

In summary, our data support prices' positive and negative effects on reviews, aligning consistently with our theoretical model. Even though the overall effect is negative on average, Airbnb hosts might consider employing a more nuanced and personalized pricing strategy if they can ascertain the probable direction of the price effect on reviews.

The data highlight that asymmetric information is the primary driver of this effect. Users with diffused priors predominantly experience a negative price effect on reviews, aside from a few outliers. This trend suggests that Airbnb hosts may benefit from offering proactive discounts to new users, mitigating the risk that the property may not align with their expectations. It is crucial to note that a potential mismatch is attributed to the guests' diffused prior beliefs rather than necessarily to the low quality of the listing. Offering a discount is not justified in the case of high-end listings with inelastic demand.

Conversely, listings catering to users with extensive experience and high affluence could display a positive correlation between prices and reviews. In these instances, hosts might contemplate increasing their prices as a strategic move to capitalize on advantageous selection. Implementing such a pricing strategy aligns with our theoretical framework, which posits that elevating prices would filter out marginal users, thereby retaining affluent, infra-marginal users who prioritize accommodation quality.

	(1) Guest review score (1-5)	(2) Guest review score (1-5)	(3) Guest review score (1-5)	(4) Guest review score (1-5)	(5) Guest review score (1-5)
High-price Arm	-0.218***	-0.145**	-0.123***	-0.156***	-0.107**
	(0.0631)	(0.0563)	(0.0427)	(0.0533)	(0.0431)
Picture count above mean	0.137***				
	(0.0309)				
High price arm× Picture count above mean	0.229***				
	(0.0790)				
Guest ID (decile) below mean		0.00873			
		(0.0303)			
High price arm× Guest ID (decile) below mean		0.140*			
		(0.0768)			
Reviewee count above mean			-0.00523		
			(0.0402)		
High price arm× Reviewee count above mean			0.272***		
			(0.0961)		
City median income above mean				0.0385	
				(0.0302)	
High price arm× City median income above mean				0.182**	
				(0.0765)	
City unemployment rate below mean					0.0240
					(0.0348)
High price arm× City unemployment rate below mean					0.194**
					(0.0941)
N	2460	2460	2460	2460	2460

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Exploratory analysis of heterogeneous treatment effects. Each column is an OLS regression. High-price arm is a dummy variable indicating a high combination of shocks as defined in Section 4. Interaction terms are dummy-with-dummy interactions.

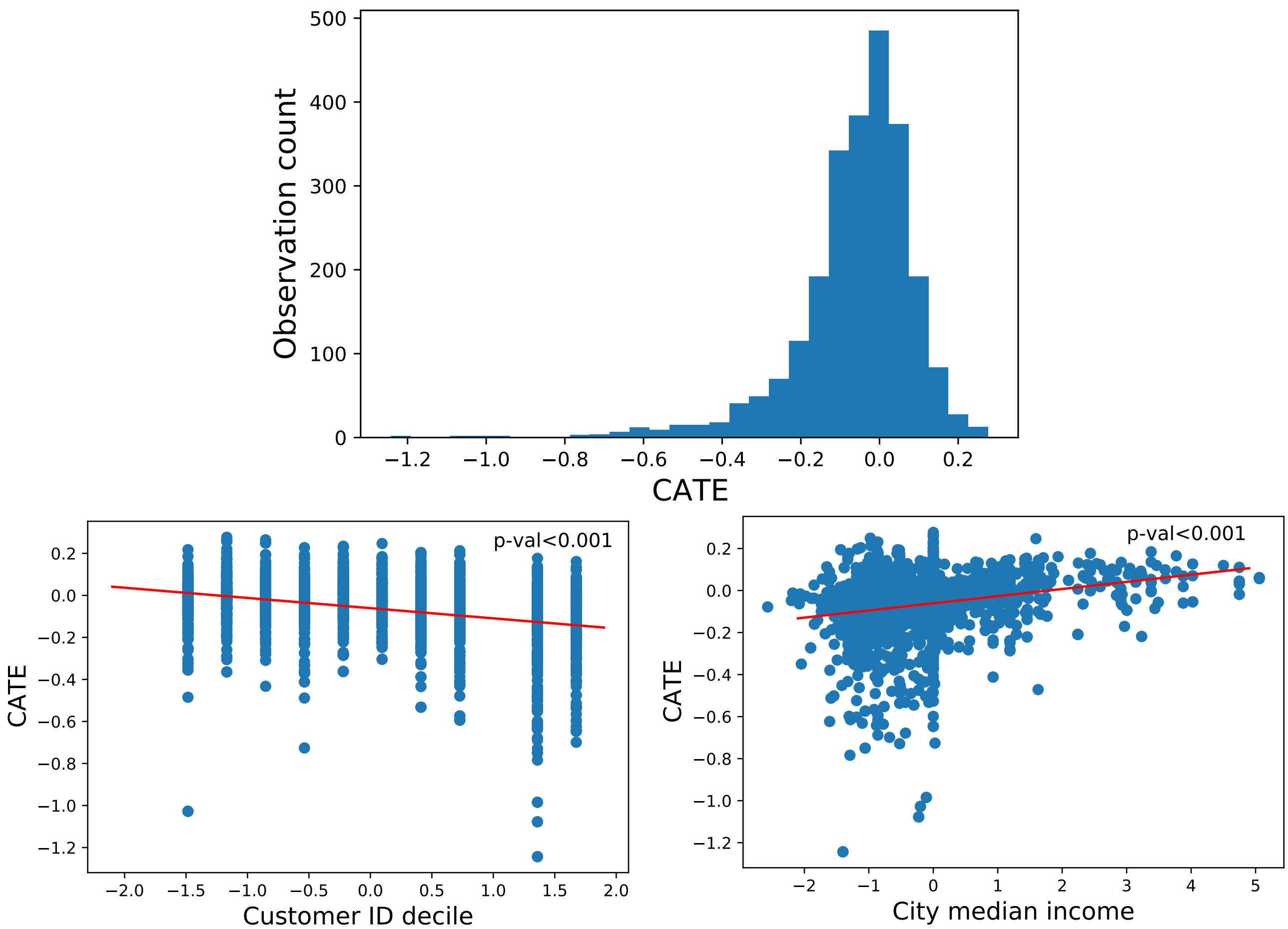


Figure 5: Distribution of Conditional Average Treatment Effects estimated using causal forest. The bottom two panels display scatter plots of CATE against the selected moderators: information precision and income. Alongside the raw data, OLS regressions and p-values are shown. Each regressor is standardized with a mean of zero and a standard deviation of one. Therefore, a value of -1 means one standard deviation below the mean, while a value of 1 indicates one standard deviation above the mean.

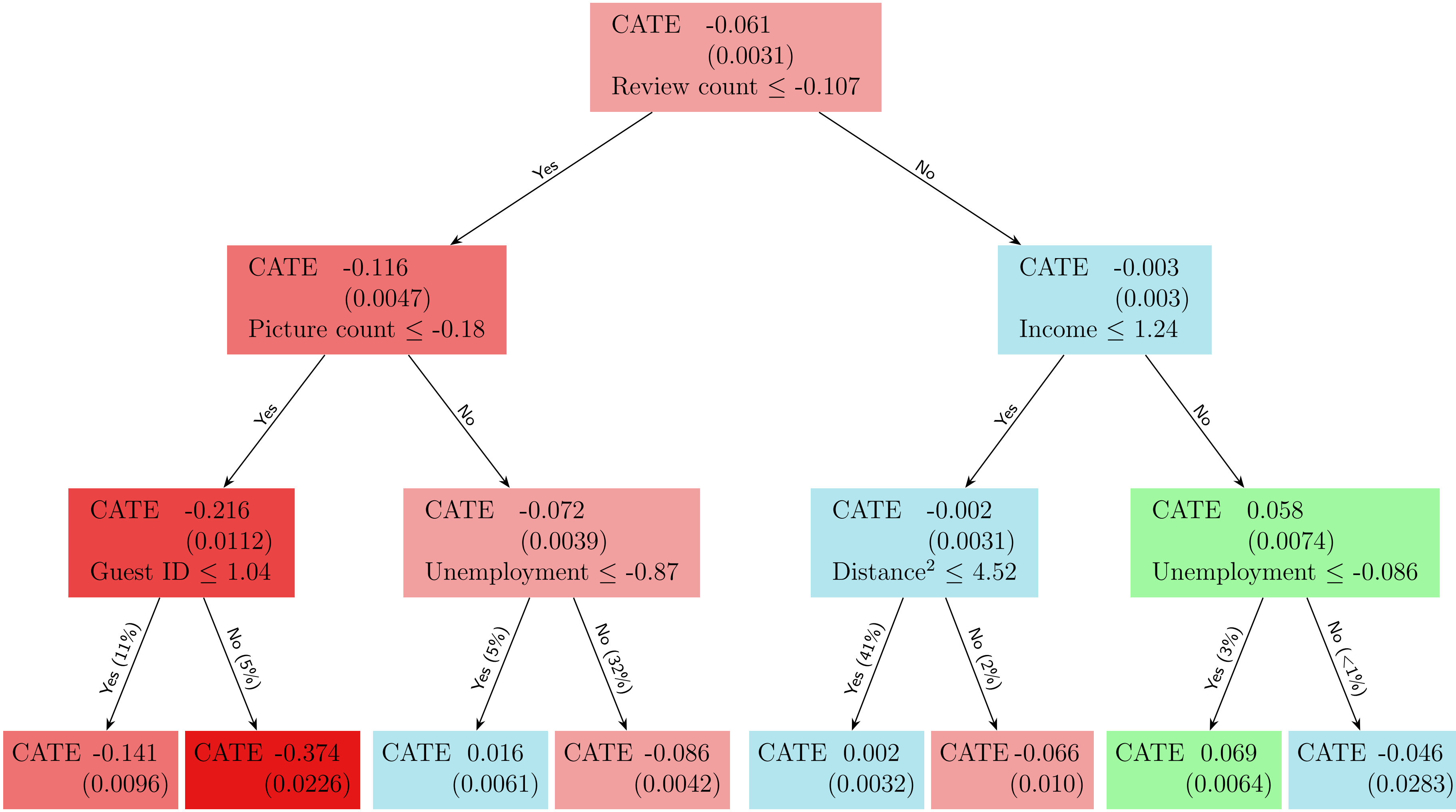


Figure 6: Decision tree depicting decomposition of Conditional Average Treatment Effects. Each regressor is standardized with a mean of zero and a standard deviation of one. Therefore, a value of -1 means one standard deviation below the mean, while a value of 1 indicates one standard deviation above the mean. The bottom set of arrows is accompanied by percentages indicating the size of the sub-sample for each leaf.

Guest selection

Changing the posted price leads to a different selection of guests ultimately making the booking. In particular, guests who book at higher rates must have a greater willingness to pay (WTP) than those who book at a lower cost. WTP may vary among guests for various reasons, including differences in the marginal utility for quality, horizontal preferences, outside options, and search costs. Our empirical analysis cannot discern these pathways, as we would require additional data to explain why WTP varies across customers. Our primary findings about the role of consumer surplus and asymmetric information merely require that WTP is greater in the high-price arm.

WTP may be connected to a resulting review through channels not involving consumer surplus or product quality. Suppose, for instance, guests from certain countries are more stringent reviewers and have higher WTP, with both quantities related only spuriously via location. Within our theoretical framework, this would mean that the review production function, R, and marginal utility α may be correlated via a common unobservable (such as location).

Since we never observe the same guest paying two different prices for the same booking, it is necessary to contemplate the potential influence of common unobservable factors on R and α. While direct testing for sorting on common unobservables is not possible, Table 2 demonstrates that observable guest characteristics remain constant across different experimental arms. Moreover, common unobservable factors cannot justify the varying impacts of price within the aided frames concurrent with its substantial influence in the value frame. In addition, these factors fall short of explaining the variation in heterogeneous treatment effect, depending on the level of experience and the wealth of the customer. As a result, while we cannot entirely dismiss the existence of common unobservables, we assert that the impact of prices on reviews, driven by signaling and consumer surplus, is considerable.

Expanding on this perspective, one could posit that variations in wealth and experience are not random phenomena. Consequently, the interactive effects could also be prone to guest selection. This line of criticism is a familiar terrain for most heterogeneous treatment effect methodologies. Short of randomizing the moderators, the typical countermeasure involves considering a gamut of variables tied to a specific underlying factor. We adopt this strategy, highlighting several interactions driven by assorted proxy variables collectively reinforcing our primary assertions.

In the following part, we finish our research with a concise examination of the impact of reviews on pricing from the supply side.

5.2 Impact of reviews on demand

	(1) Log-price per night	(2) Log-price per night	(3) Log-price per night	(4) Days to the next reservation	(5) Days to the next reservation	(6) Days to the next reservation
Previous score	0.235***	0.221***	0.622*	-2.967***	-2.875***	-25.55**
	(0.0310)	(0.0297)	(0.348)	(0.965)	(0.967)	(12.14)
Date Controls	no	yes	yes	no	yes	yes
IV	no	no	yes	no	no	yes
N	2330	2330	2330	2330	2330	2330

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Impact of the last review on the subsequent transaction price. Columns (1)-(4) are OLS regressions. Column (5) is the IV regression.

In this subsection, we demonstrate the impact of reviews on the demand curve for reservations. The goal is to contextualize the previous results within the supply-side relationship between reviews and sales.

Airbnb employs various methods to convey listing reviews to potential guests. Next to the listing's name, the STR platform displays the average review. Additionally, after clicking on the listing name, Airbnb provides a chronological list of reviews, displaying star ratings and review content. This chronological display highlights the most recent reviews as they more accurately reflect the current listing quality. Our investigation examines the influence of the most recent review by artificially modulating that review through our experimental design. Specifically, we manipulate past random price shocks as an external instrument to influence the most recent review. This methodology augments the study by Vana and Lambrecht (2021), which assesses the effect of the most recent review on purchase likelihood by isolating externally induced variations in past reviews.

Customers do not directly observe past pricing shocks. As such, these can influence current demand exclusively via reviews. Therefore, if we identify a relationship between demand and reviews, leveraging the variations introduced by past pricing shifts, it implies that reviews exert a causal influence on demand.

Directly estimating dynamic demand is complex, given that Airbnb lists hundreds of prices on any given day, each pertinent to every subsequent day over the next two years. The listing is subject to a plethora of daily demand curves that vary based on the specific calendar days of the reservation and the number of days until the check-in date. Thus, dealing directly with demand would likely necessitate a complicated structural model (see Williams, 2022; Chen and Jeziorski, 2022).

Instead, our strategy is to demonstrate the simultaneous positive influence of previous reviews on equilibrium prices and quantities. While a positive impact on price is indicative, it does not conclusively signify a rise in demand. Such an influence might stem from a supply shift, leading to the movement of transaction prices along the demand curve. For instance, after a positive review, a firm might choose to raise its listed prices without any real change in demand, driven either by a learning process or sub-optimal pricing. Moreover, even if consumers are not cognizant of historical price shocks, firms might be. Such knowledge can trigger additional shifts in supply unrelated to reviews, perhaps due to changes in dynamic pricing algorithms. Yet, in such circumstances, a supply-driven price increase should correspondingly reduce quantity because, as demonstrated earlier, demand for reservations is downward-sloping (see Footnote 34). Conversely, a concurrent rise in both price and quantity suggests a shift in the demand curve (in addition to a possible shift in supply). As mentioned earlier, qualitatively, this shift in demand should be interpreted causally because of the nature of the utilized review variation.

To articulate the reasoning above, we propose a reduced-form model. In this model, we utilize the waiting time between bookings as a proxy for quantity. A more direct metric might be occupancy rates; however, we observe little variation in occupancy as the listings are frequently at capacity due to price skimming. Furthermore, occupancy rates are generally computed over extended periods, making them less directly connected to the most recent review.

Suppose the host i obtains review r_{it} after concluding reservation t. We examine the following system of equations:

$$\log p_{it+1} = \gamma^{(1)} r_{it} + \epsilon_{it}^{(1)},$$
$$\log \Delta_{it+1} = \gamma^{(2)} r_{it} + \epsilon_{it}^{(2)},$$

where $\log p_{it+1}$ is the logarithm of the transaction daily rate for the new reservation $t+1$, and Δ_{it+1} represents the number of days between obtaining review r_{it} and the arrival of booking $t+1$.

The results are outlined in Table 8. Columns (1)-(3) present the impact of the most recent review on a subsequent transaction price. Column (1) offers the most straightforward regression, showing that an increase in the recent review by 1 star augments the transaction price by approximately 25%; in other words, an increase of a star rating by 1 standard deviation leads to a 17% rise in the next transaction price. Column (2) adds controls for seasonality, including month and year-fixed effects. Column (3) employs experimental pricing shocks for reservation t as instruments for review r_t. As outlined earlier, they serve as good demand instruments since past prices and pricing shocks are unobserved to the guests and uncorrelated with contemporaneous demand shocks. This analysis reveals that one standard deviation increase amplifies the next price by over 50%. The estimates are considerably noisy but still significant at 10%; thus, the point estimate should be interpreted cautiously.

As discussed earlier, hosts may increase prices based on past pricing shocks. Columns (4)-(6) tackle this scenario, demonstrating that an enhancement in reviews hastens the arrival rate of the next reservation. The impact is notable, spanning a reduction of waiting time from nearly 1.8 days to over 17 days for each standard deviation increase in reviews. The Instrumental Variable (IV) estimates are more substantial, albeit accompanied by increased variability. Synthesizing these findings, the simultaneous increase in quantity and transaction prices after a positive review suggests a shift in the demand curve. Additionally, considering that price estimates are noisier than quantity estimates, it is worth noting that even a neutral impact of reviews on prices, combined with a positive effect on quantity, would be enough to infer a shift in the demand curve. This scenario is conceivable if the short-term supply of STRs is flat.

The favorable effect of a prior review on demand encourages an investment in earning positive reviews, potentially through a temporary price reduction. Specifically, firms might find it advantageous to drop their prices following a negative review to cultivate better future reviews. On the flip side, they are also motivated to hike prices and maximize profits when reviews are stellar. The data in Table 8 aligns with these pricing cycles but can also be explained by optimal static pricing based on the current review score. Reduced form modeling cannot distinctly separate static pricing from dynamic pricing incentives, and we leave this question for further research.

The Keybee pricing algorithm leverages advanced, forward-looking machine learning routines to optimize hosts' revenue. This approach ensures automatic estimation and adaptation to any abrupt shifts in demand and, if well calibrated, would account for the endogeneity of the reviews. Casual hosts or firms in markets with more considerable menu costs could implement rudimentary pricing that discounts the property when the reviews are low. A more automated approach could be a simple price skimming, available as "early-bird" and "last-minute" pricing options on the Airbnb Hosting Dashboard. Such skimming may start from a higher and finish at a lower price compared to a case in which reviews do not depend on the price. This pricing model leverages a greater likelihood of being sold out quicker when reviews are stellar (early-bird) and the potential for more prolonged vacancy when reviews are less than ideal (last-minute).

The ability to "buy" positive reviews has long-term competition consequences. In markets like short-term rentals (STR), where businesses frequently come and go, those with lower reviews are more likely to shut down. Meanwhile, incumbent businesses with high reviews might preempt entry. Thus, incumbents may use predatory pricing and review systems to push out competitors or keep new ones from entering the market. This strategy will likely be more common in thinner markets, possibly leading to market dominance.

6 Conclusion

The paper uses a field experiment to establish the causal impact of transaction price on subsequent product reviews. Increasing the price has a negative impact on subsequent reviews.

Conversely, present high reviews have a positive impact on future transaction prices. We also show that consumer surplus and asymmetric information about the quality are essential considerations in the consumer post-purchase deliberation and contribute to our results.

The identified effects have several consequences for market efficiency. First, if reviews depend on past pricing, the current score must be evaluated in the context of past prices; otherwise, it would not provide an apples-to-apples comparison of quality across products. On the supply side, because past reviews influence future company profits, firms face downward pricing pressure to keep the ratings high. The pressure occurs because consumers can punish the firm if the transaction price seems excessive compared to the offered value. The effect is present even for a monopolist, demonstrating that the rating systems can be an essential competitive force even in markets with concentrated ownership. Finally, reviews may be a source of market power for the incumbent because the upcoming competitors must invest in good reviews, which may be prohibitively costly.

References

Akerlof, G. A. (1978): "The market for "lemons": Quality uncertainty and the market mechanism," in *Uncertainty in economics*, pp. 235–251. Elsevier.

Ananthakrishnan, U. M., D. Proserpio, and S. Sharma (2019): "I Hear You: Does Quality Improve with Customer Voice?," *Available at SSRN 3467236*.

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

Arellano, M., and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The review of economic studies*, 58(2), 277–297.

Askalidis, G., S. J. Kim, and E. C. Malthouse (2017): "Understanding and overcoming biases in online review systems," *Decision Support Systems*, 97, 23–30.

Athey, S., and S. Wager (2019): "Estimating Treatment Effects with Causal Forests: An Application," .

Babić Rosario, A., F. Sotgiu, K. De Valck, and T. H. Bijmolt (2016): "The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors," *Journal of marketing research*, 53(3), 297–318.

Battocchi, K., E. Dillon, M. Hei, G. Lewis, P. Oka, M. Oprescu, and V. Syrgkanis (2019): "EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation," https://github.com/py-why/EconML, Version 0.x.

Berry, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, pp. 242–262.

Bondi, T., and R. Stevens (2019): "The Good, The Bad and The Picky: Consumer Heterogeneity and The Reversal of Movie Ratings," .

Borkovsky, R. N., A. Goldfarb, A. M. Haviv, and S. Moorthy (2017): "Measuring and understanding brand value in a dynamic model of brand management," *Marketing Science*, 36(4), 471–499.

Brandes, L., D. Godes, and D. Mayzlin (2022): "Extremity bias in online reviews: The role of attrition," *Journal of Marketing Research*, 59(4), 675–695.

Burtch, G., Y. Hong, R. Bapna, and V. Griskevicius (2018): "Stimulating online reviews by combining financial incentives and social norms," *Management Science*, 64(5), 2065–2082.

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

Cabral, L. M., and M. H. Riordan (1994): "The learning curve, market dominance, and predatory pricing," *Econometrica: Journal of the Econometric Society*, pp. 1115–1140.

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

Chen, N., and P. Jeziorski (2022): “Consequences of dynamic pricing in competitive airline markets,” *Available at SSRN 4285718*.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): “Double/debiased machine learning for treatment and structural parameters,” .

Chevalier, J. A., and D. Mayzlin (2006): “The effect of word of mouth on sales: Online book reviews,” *Journal of marketing research*, 43(3), 345–354.

Clemons, E. K., G. G. Gao, and L. M. Hitt (2006): “When online reviews meet hyper-differentiation: A study of the craft beer industry,” *Journal of management information systems*, 23(2), 149–171.

Cooper, R., and T. W. Ross (1984): “Prices, product qualities and asymmetric information: The competitive case,” *The Review of Economic Studies*, 51(2), 197–207.

Danescu-Niculescu-Mizil, C., G. Kossinets, J. Kleinberg, and L. Lee (2009): “How opinions are received by online communities: a case study on amazon. com helpfulness votes,” in *Proceedings of the 18th international conference on World wide web*, pp. 141–150.

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

——— (2006): “Strategic manipulation of internet opinion forums: Implications for consumers and firms,” *Management science*, 52(10), 1577–1593.

Dellarocas, C., and R. Narayan (2006): “What motivates consumers to review a product online? A study of the product-specific antecedents of online movie reviews,” *Statistical Science, C*, 21, 277–285.

Dellarocas, C., and C. 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.

Dubé, J.-P., G. J. Hitsch, and P. Manchanda (2005): "An empirical model of advertising dynamics," *Quantitative marketing and economics*, 3(2), 107–144.

Erdem, T., M. P. Keane, and B. Sun (2008): "A dynamic model of brand choice when price and advertising signal product quality," *Marketing Science*, 27(6), 1111–1125.

Everitt, B. (2010): "The Cambridge dictionary of statistics," in *The Cambridge dictionary of statistics*. Cambridge University Press.

Filippas, A., J. J. Horton, and J. M. Golden (2022): "Reputation Inflation," *Marketing Science*, 41(4), 733–745.

Fradkin, A., E. Grewal, and D. Holtz (2021): "Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on airbnb," *Marketing Science*, 40(6), 1013–1029.

Fradkin, A., and D. Holtz (2022): "Do incentives to review help the market? Evidence from a field experiment on Airbnb," *American Economic Review: Insights*.

Freedman, D. A. (2008): "On regression adjustments to experimental data," *Advances in Applied Mathematics*, 40(2), 180–193.

Gao, G., B. N. Greenwood, R. Agarwal, and J. S. McCullough (2015): "Vocal minority and silent majority," *MIS quarterly*, 39(3), 565–590.

Godes, D., and D. Mayzlin (2004): "Using online conversations to study word-of-mouth communication," *Marketing science*, 23(4), 545–560.

Gu, B., and M. Lin (2006): "The dynamics of online consumer reviews," in *Workshop on Information Systems and Economics (WISE)*. Citeseer.

Guo, X., and B. Jiang (2016): "Signaling through price and quality to consumers with fairness concerns," *Journal of Marketing Research*, 53(6), 988–1000.

Hu, N., P. A. Pavlou, and J. Zhang (2006): "Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication," in *Proceedings of the 7th ACM conference on Electronic commerce*, pp. 324–330.

Hu, N., P. A. Pavlou, and J. J. Zhang (2017): "On self-selection biases in online product reviews.," *MIS Q.*, 41(2), 449–471.

Hui, X., T. J. Klein, and K. O. Stahl (2022): "When and why do buyers rate in online markets?," *CEPR Discussion Paper No. DP17006.*

Hunter, M. (2020): "Chasing Stars: Firms' Strategic Responses to Online Consumer Ratings," *Available at SSRN 3554390.*

Karaman, H. (2021): "Online review solicitations reduce extremity bias in online review distributions and increase their representativeness," *Management Science*, 67(7), 4420–4445.

Laouénan, M., and R. Rathelot (2022): "Can information reduce ethnic discrimination? evidence from airbnb," *American Economic Journal: Applied Economics*, 14(1), 107–32.

Lewis, G. (2011): "Asymmetric information, adverse selection and online disclosure: The case of eBay motors," *American Economic Review*, 101(4), 1535–46.

Lewis, G., and G. Zervas (2019): "The supply and demand effects of review platforms," *Available at SSRN 3468278.*

Li, X., and L. M. Hitt (2008): "Self-selection and information role of online product reviews," *Information Systems Research*, 19(4), 456–474.

Luca, M., and O. Reshef (2021): "The effect of price on firm reputation," *Management Science*, 67(7), 4408–4419.

Moe, W. W., and D. A. Schweidel (2012): "Online product opinions: Incidence, evaluation, and evolution," *Marketing Science*, 31(3), 372–386.

Negi, A., and J. M. Wooldridge (2021): "Revisiting regression adjustment in experiments with heterogeneous treatment effects," *Econometric Reviews*, 40(5), 504–534.

Nosko, C., and S. Tadelis (2015): "The limits of reputation in platform markets: An empirical analysis and field experiment," Discussion paper, National Bureau of Economic Research.

Oliver, R. L. (1980): "A cognitive model of the antecedents and consequences of satisfaction decisions," *Journal of marketing research*, 17(4), 460–469.

Olson, J. C., and P. A. Dover (1979): "Disconfirmation of consumer expectations through product trial.," *Journal of Applied psychology*, 64(2), 179.

Peterson, R. A., and W. R. Wilson (1992): "Measuring customer satisfaction: fact and artifact," *Journal of the academy of marketing science*, 20(1), 61–71.

Resnick, P., and R. Zeckhauser (2002): "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system," in *The Economics of the Internet and E-commerce*. Emerald Group Publishing Limited.

Stenzel, A., C. Wolf, and P. Schmidt (2020): "Pricing for the stars: Dynamic pricing in the presence of rating systems," in *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 273–274.

Swan, J. E., and I. F. Trawick (1981): "Disconfirmation of expectations and satisfaction with a retail service.," *Journal of retailing*.

Vana, P., and A. Lambrecht (2021): "The effect of individual online reviews on purchase likelihood," *Marketing Science*, 40(4), 708–730.

Wang, S. S., and R. Van Der Lans (2018): "Modeling gift choice: The effect of uncertainty on price sensitivity," *Journal of Marketing Research*, 55(4), 524–540.

Williams, K. R. (2022): "The welfare effects of dynamic pricing: Evidence from airline markets," *Econometrica*, 90(2), 831–858.

Wolinsky, A. (1983): "Prices as signals of product quality," *The review of economic studies*, 50(4), 647–658.

Yao, J., D. Wang, and G. Mortimer (2022): "All new smaller size! Why getting less with shrinkflation is preferable to paying more," *The Conversation*.

Zervas, G., D. Proserpio, and J. 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, X., D. Chris, and A. Neveen (2004): "Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning," in *Proceedings of the 25th International Conference on Information Systems (ICIS), Washington DC.*

Zhong, Z. (2021): "Chasing diamonds and crowns: Consumer limited attention and seller response," *Management Science*.

Disclosure, Author A

Author A holds more than 10% of equity in and sits on the Board of Directors of Keybee Inc. They received more than $10,000 compensation from Keybee Inc. that was unrelated to the work on this manuscript. They received no compensation related to work on this manuscript. Keybee Inc. has no financial interest in the publication of the results.

Disclosure, Author B

Author B holds more than 10% of equity in, sits on the Board of Directors, and is the CEO of Keybee Inc. They received more than $10,000 compensation from Keybee Inc. that was unrelated to the work on this manuscript. They received no compensation related to work on this manuscript. Keybee Inc. has no material financial in the publication of the results.

Online Appendix

A Placebo tests

	(1) Guest review score (1-5)	(2) Guest Value score (1-5)	(3) Guest Check-in score (1-5)	(4) Guest Accuracy score (1-5)	(5) Guest Location score (1-5)	(6) Guest Communication score (1-5)	(7) Guest Cleanliness score (1-5)
Previous reservation High-price Arm	-0.0531	-0.00108	-0.00439	0.00350	0.0259	0.00544	-0.0266
	(0.0428)	(0.0426)	(0.0306)	(0.0389)	(0.0312)	(0.0300)	(0.0374)
N	1387	1379	1379	1381	1379	1381	1381

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Placebo tests. The table presents regression of current reviews on past price shocks.

The placebo tests are designed to rule out any spurious correlation due to an error in the pricing and randomization algorithms or selection of observations due to downtimes in the data collection scripts. To conduct the test, we obtained the pricing arm of the previous observation for a given listing and used it as a covariate in the review t-tests instead of using the correct pricing arm. Since the current guest is unaware of the previous reservation price, it should have no relationship with the current review. Indeed, we find no significant relationship between the previous pricing arm and the current review score across all review frames.

B Additional graphs in figures

	(1)	(2)	(3)	(4)
	Guest review score (1-5)	Guest Value score (1-5)	Guest review score (1-5)	Guest Value score (1-5)
Log-price per night	-0.230***	-0.143*	-0.277***	-0.218**
	(0.0835)	(0.0870)	(0.0912)	(0.0990)
Log-price× above median description length	0.178*	-0.00196		
	(0.0912)	(0.107)		
Log-price× above median picture count			0.241**	0.119
			(0.0993)	(0.116)
Date Controls	yes	yes	yes	yes
Listing Controls	no	no	no	no
Listing FE	yes	yes	yes	yes
Experiment	no	no	no	no
N	2082	2072	2082	2072

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: The impact of pricing on subsequent reviews as a function of the listing description length and number of photos. The table presents OLS regressions.