# Safety Reviews on Airbnb: An Information Tale*

Aron Culotta[†] Ginger Zhe Jin[‡] Yidan Sun[§] Liad Wagman[¶]

Preliminary Draft, March 2022

**Abstract**

Many online platforms facilitate and publish user reviews in order to build trust between anonymous buyers and sellers. At the same time, platforms can also monitor, filter, and remove certain user reviews, even if they reflect the true experiences of buyers. Using Airbnb and official crime data in five major US cities, we study a December-2019 Airbnb policy that has a potential to remove and discourage reviews about the safety of a listing's vicinity. It is unclear how Airbnb implements this policy, but counterfactual simulation suggests that a complete removal of vicinity safety reviews would hurt guests and increase revenues from reservations on Airbnb, with positive sorting towards listings with such vicinity safety reviews. Conversely, highlighting vicinity safety reviews as if the guest had written a vicinity safety review out of her previous experience would generate opposite effects. Because vicinity safety reviews are more closely correlated with official crime statistics in low-income and minority neighborhoods, our findings suggest that suppressing or highlighting vicinity safety reviews would have different effects on different neighborhoods.

**Keywords:** Airbnb, safety review, crime, information design, online platform

*We are grateful to AirDNA for providing the data and to our home universities for financial support. None of us has a significant financial relationship with Airbnb or any competing short term rental platforms. Xiang Hui, Meng Liu, and participants at the Luohan Academy Webinar, Washington University at St. Louis, and Boston University have provided constructive comments. The content and analyses in this paper reflect the authors' own work and do not relate to any institution or organization with whom the authors are affiliated.  All errors are our own.

[†]Tulane University. Email: aculotta@tulane.edu.

[‡]University of Maryland & NBER. Email: ginger@umd.edu.

[§]Illinois Institute of Technology. Email: ysun100@hawk.iit.edu.

[¶]Illinois Institute of Technology. Email: lwagman@stuart.iit.edu.

# 1 Introduction

Information design is crucial for online platforms. Because online platforms derive revenue from the trade they intermediate, they often adopt information mechanisms that allow buyers to discern high- and low-quality sellers. For example, consumer feedback, product recommendations, quality certification, and advertising are widely used by online platforms. This is consistent with the vast economic literature about asymmetric information and market efficiency. Arguably, such mechanisms for conveying information can be even more effective online than offline, because online platforms face fewer physical constraints in space, have a lower barrier to entry, and allow online users to access, search, and utilize the vast pool of information collected by the platform (Einav et al., 2016).

However, there are times when online platforms explicitly choose to *limit* the flow of information, even if that information is likely to be authentic and reflects users' true experiences. For example, in a recent policy change effective December 11, 2019, Airbnb announced that, going forward, guest reviews about listings that include "content that refers to circumstances entirely outside of another's control" may be removed by the platform.[1] This policy change implies that reviews about the safety of a listing's vicinity (henceforth, "vicinity safety reviews") are discouraged and may be subject to deletion by the platform, while guest reviews regarding safety issues within the listed property are still permitted (henceforth, "listing safety review").

It is not obvious whether this limit on vicinity safety reviews is beneficial or detrimental to players on the platform. On the one hand, location is a fixed attribute of any specific listing, and the safety of the vicinity of a listing is usually out of the control of the host. Limiting guest feedback to within-listing safety may motivate hosts to focus on the dimensions they can control and improve. On the other hand, future guests may care about both listing and vicinity safety when they choose which listing to reserve, regardless of whether these safety issues are under the control of hosts or not.

To better understand the impact of this limit on information disclosure, we study detailed data of all Airbnb listings and their reviews in five major US cities (Atlanta, Chicago, Los Angeles, New Orleans, and New York City), from May 2015 to December 2019. This data

[1] See, for example, `https://www.airbnb.com/help/article/2673/airbnbs-review-policy` and `https://community.withairbnb.com/t5/Airbnb-Updates/Making-reviews-more-relevant-and-useful-for-our-community/td-p/1191576`.

is collected by AirDNA, a third party that tracks Airbnb listings and listing-specific feedback across the US. We use a Lexicon approach to identify vicinity safety reviews and listing safety reviews posted by Airbnb guests. Because our data ends in December 2019, our data largely precede the new Airbnb policy that discourages vicinity safety reviews. As detailed below, we find that 1.23% of guest reviews are related to safety, of which 98.27% are about vicinity safety rather than listing safety. This implies that Airbnb's limit on vicinity safety reviews, if strictly enforced during our sample period, would have eliminated almost all reviews that mention safety concerns on Airbnb.

Since guest feedback may reflect guests' subjective opinion of their stay experience, we also obtain (local) government-reported crime statistics for these five cities, by zip code and month during the same period. The data suggest that, as vicinity safety reviews accumulate slowly on Airbnb, the rank correlation between the normalized total count of vicinity safety reviews in a zip code up to a month $t$ and the normalized official crime statistics of that zip code-month is increasing over time. For low-income or minority neighborhoods, the rank correlation can be as high as 0.6 by the end of our sample period (December 2019).

One key question is how much impact these safety reviews have on consumer choice as far as which Airbnb listings to book. If prospective guests do not read or do not care about vicinity safety reviews, it does not matter whether the platform puts any limit on vicinity safety reviews. Our findings show that, within the same listing, having any vicinity safety review is associated with a 1.01% reduction in the listing's monthly occupancy rate and a 1.33% reduction in its average paid price per night. The association with listing safety review is even stronger: having any listing safety review is associated with a 2.59% drop in occupancy and 1.72% in price. These findings, all significant at 99% confidence, suggest that prospective guests are concerned about both listing and vicinity safety, and have different sensitivities to changes in these two types of safety reviews.

Another way to understand consumer sensitivity to vicinity safety is examining whether the guests that wrote a vicinity safety review on Airbnb have changed their subsequent Airbnb activities after having experienced the safety issue mentioned in their review. Arguably, the effect of self-experience is more direct and salient than reading vicinity safety reviews written by anonymous strangers. Indeed, our analysis supports this intuition: guests that wrote a vicinity safety review on Airbnb are less likely to book future stays on Airbnb after posting the safety review, and when they do book future stays on Airbnb, they tend to book in areas with fewer official crimes, fewer overall safety reviews, and fewer listings with safety reviews. All these findings are compared to guests that have used Airbnb with similar frequencies and booked similar listings in terms of crime and vicinity safety reviews but never write any safety review within our sample period.

Such guest sensitivity to vicinity safety reviews suggests that omitting them on Airbnb could make future guests worse off, as they may mistakenly book listings in potentially unsafe locales. To gauge the potential loss of guest welfare, we obtain a dataset of competing entire-home VRBO listings and use a discrete choice model to estimate consumer utility from Airbnb entire-home listings, while treating VRBO listings in the same zip code-month as the outside good. We then use the structural estimates to quantify consumer surplus under the status quo of our sample (i.e., vicinity safety reviews are largely permitted) versus two counterfactual scenarios: eliminating all vicinity safety reviews, or alerting all guests to the existing vicinity safety reviews and making them as sensitive as those that have written safety reviews themselves.

Compared to the scenario that eliminates all vicinity safety reviews on Airbnb, we find that the status quo increases consumer surplus by around 0.86% (without price change), because vicinity safety reviews help guests to substitute Airbnb listings with such reviews for Airbnb listings without them or located elsewhere, or for listings on a competing platform (VRBO). In comparison, making all guests as alert as those that have written vicinity safety reviews themselves can further increase consumer surplus relative to the status quo, by 0.21% (without price change).

These counterfactual scenarios have different implications for Airbnb hosts and Airbnb as a platform. Because vicinity safety reviews make guests more hesitant to book certain Airbnb listings, posting or emphasizing them have a sorting effect among Airbnb listings, and a demand shrinkage effect on Airbnb listings with vicinity safety reviews. In total, we find the status quo generates \$30.84M (0.46%) less general booking value (GBV), or revenue from reservations, for listings on Airbnb in our sample than when eliminating vicinity safety reviews from those Airbnb listings. Making all guests as alert as those that have written vicinity safety reviews could decrease Airbnb's GBV by 8.01% or \$538.87M. These calculations highlight the diverging interests between Airbnb guests, Airbnb hosts located in areas with different vicinity safety, and the entire Airbnb platform.

As we detail in the next section, our work contributes to the empirical literature of online feedback and seller reputation, and the rising literature of information design in online platforms. As an information intermediary, online platforms have more incentives than a traditional seller to alleviate information asymmetries between buyers and sellers. However, online platforms are still inherently different from a social planner, because they maximize their own profits rather than the total welfare of buyers and sellers on the platform, and they do not internalize the impact of their policies on competing platforms. Our empirical findings highlight these differences, and quantify the potential impact of a policy that limits the release of information for different economic stakeholders. We also document how the impact of the policy may vary for neighborhoods of different income or different minority populations, as being inclusive could be important for the platform or the social planner. These findings can help facilitate ongoing discussions as to what role and responsibility digital platforms should have as far as collecting and disseminating quality-related information online.

The remainder of the paper is organized as follows. Section 2 reviews the related literature. Section 3 provides some background regarding Airbnb's review system. Section 4 describes the dataset and provides summary statistics. Section 5 reports our empirical findings and implications. Section 6 provides back-of-envelope calculation on how listings' GBV and consumer surplus would change in two counterfactual scenarios. Section 7 offers a discussion, and Section 8 concludes.

## 2 Related Literature

Safety reviews are a type of buyer-to-seller feedback; thus, our study is directly related to the literature of online feedback and seller reputation.

Arguably, buyer-to-seller feedback is more important for marketplace efficiency than seller-to-buyer feedback, because the key information asymmetry is sellers' private information of their own product or service quality. By the law of large numbers, a sufficient volume of authentic buyer feedback would eventually reveal hidden information regarding seller quality. However, not all buyers are willing to provide feedback, partly because reviewers are not compensated for submitting reviews. For example, 64% of eBay transactions are rated by buyers in the sample studied by (Hui et al., 2021), and 73.5% of New Your City UberX trips are rated by passengers

(Liu et al., 2021). In comparison, 53.1% of Airbnb trips in our sample have received feedback from guests, which is in line with the guest review rate reported by Fradkin et al. (2021) based on earlier Airbnb data in 2014.

Since accurate feedback is a public good subject to under-provision, many platforms attempt to encourage buyer feedback by offering status, coupons, and merchandise discounts (Li and Xiao, 2014; Cabral and Li, 2015; Li et al., 2020; Fradkin et al., 2015). In contrast, the policy studied in this paper aims to discourage buyers from giving feedback on a particular dimension of quality, which could exacerbate the public good problem in online feedback.

The imperfect review rate is particularly problematic as far as negative feedback is concerned. Studies have shown that buyers tend to under-report bad experiences, with potential explanations that include fear of retaliation, preference to leave the platform after bad experience (Nosko and Tadelis, 2015), pressure to give above-average ratings (Barach et al., 2020), and social connection to the rated seller (Fradkin et al., 2015). For arguably rare, negative events such as safety, the probability of observing pertinent feedback from prior buyers could be further reduced, simply because the chance of experiencing a safety issue is small in absolute terms, even if a neighborhood has safety risks. A policy that discourages vicinity safety reviews could reinforce an existing bias against negative feedback. In fact, perhaps in part due to such a bias against negative feedback, Chakravarty et al. (2010) finds that consumers are more responsive to negative feedback than to positive feedback. This pattern is confirmed in our study: the marginal effect of having any vicinity safety review on a listing's occupancy rate is comparable to that of a 38.91% reduction in the listing's average guest rating.

Another consequence of the bias against negative feedback is that safety reviews on any Airbnb listing can only accumulate slowly over time. This could affect the overall informativeness of safety reviews. Between 2015 and 2019, we observe a growing rank correlation between a zip code's normalized cumulative safety review count and that zip code's normalized official crime statistics. This suggests that cumulative safety reviews do contain useful information regarding a zip code's actual safety status, with informativeness that may have increased over time. In comparison, a few studies argue that the online feedback systems may become less informative over time because of the feedback bias reasons cited above (Barach et al., 2020; Klein et al., 2009; Hui et al., 2021). Most of these studies infer feedback informativeness from the content of feedback or policy variations within the feedback system. Our approach is different, as we compare online feedback with a completely independent data source.

More broadly, our study contributes to the growing literature of information design in online platforms. Because feedback is under-provided and there is a selection against negative feedback, researchers have studied the design of feedback systems in terms of who is allowed to provide feedback (Klein et al., 2016; Mayzlin et al., 2014; Zervas et al., 2021), how to improve the authenticity of feedback (Wagman and Conitzer, 2008; Conitzer et al., 2010; Conitzer and Wagman, 2014), when the feedback is revealed to the public (Bolton et al., 2013; Fradkin et al., 2021), what kind of feedback is shown to the public, and how to aggregate historical feedback (Staats et al., 2017; Dai et al., 2018).

Interestingly, some platforms highlight negative consumer feedback, so that future consumers are aware of potential risks associated with the target seller or target product (Pan and Zhang, 2011). An economic reason to do so is that many consumers on online platforms tend to be more responsive to negative feedback than to positive feedback (Chakravarty et al., 2010). Highlighting negative feedback may hurt the sellers with negative feedback but divert buyers towards other sellers on the same platform with zero or not as much negative feedback. If this sorting effect generates more revenue for the platform or reinforces the platform's reputation as far as honesty and transparency, the platform would have an incentive to highlight negative feedback.

In our setting, we observe a counterexample where the platform discourages buyers from providing a specific type of negative feedback. This is similar to a platform hiding, obfuscating, or deleting negative feedback. To be clear, there are legitimate reasons to do so in some situations: for example, a platform may find certain feedback fake, abusive, or misleading ex post; omitting such feedback could make the information system more authentic and informative for both buyers and sellers (Luca and Zervas, 2016; Chevalier and Mayzlin, 2006).

At the same time, studies have shown that platforms may be strategically motivated to omit certain information, including negative feedback. For instance, Kovbasyuk and Spagnolo (2018) explain why sometimes platforms seek to erase some historical bad records of sellers, in order to increase matching rates. Romanyuk and Smolin (2019) show that platforms such as Uber may seek to hide some buyer information (say, destination) prior to completing a buyer-seller match, because doing so would avoid sellers waiting for a specific type of next buyer which would reduce the overall matching rate on the platform. These two papers differ in the direction of information withholding: the former withholds seller-relevant information from future buyers, while the latter withholds buyer-relevant information from future sellers. Both suggest that the party from whom the information is kept hidden may be worse off and the platform has an incentive to trade off their welfare loss against the welfare gain of the other side of the platform and the platform's overall matching rate. Also, Lewis (2011) shows that online disclosures are important price determinants, and the disclosure costs impact both the level of disclosure and prices.

Airbnb's new policy regarding vicinity safety reviews is an example of withholding or discouraging seller-related information from prospective buyers. As shown in our counterfactual analysis, Airbnb has economic incentives to downplay vicinity safety reviews, because the more guests are alerted about vicinity safety, the lower the matching rate for the whole platform. In theory, such incentives could be dominated by a sorting effect, if posting or highlighting vicinity safety reviews could direct buyers towards safer listings on the same platform and motivate the safer listings to increase their prices sufficiently high to compensate for the platform's loss from a lower matching rate. Our back-of-envelope calculation suggests that this is not the case.

One welfare aspect that is difficult to quantify but may be relevant for Airbnb is the long-run entry and exit of users. As shown in our back-of-envelope calculations, a policy that encourages and highlights vicinity safety reviews could disproportionately hurt Airbnb hosts in relatively unsafe neighborhoods. In the long run, this could lead to a smaller choice set for guests, drive away some types of hosts and guests, and affect the economic parity across different neighborhoods. How important these long-run considerations are for Airbnb and the social planner depends on how they weigh the welfare of different users. Unfortunately, such a policy was introduced by Airbnb at the end of our data period and just prior to the emergence of the COVID-19 pandemic, so aside from estimating counterfactuals, we cannot observe the de-facto changes in seller entry and exit because of this policy change.

Another aspect that is worth highlighting in our setting is the potential spillover effect of vicinity safety reviews across listings. As stated before, all buyer-to-seller feedback is a public good that provides little economic return to the reviewer but could benefit many future prospective buyers. In addition to this common feature, vicinity safety reviews could also generate spillovers among listings in nearby geographies, should guests infer the overall safety of the vicinity from multiple nearby listings. We find some suggestive but mixed evidence for such negative spillovers: for a focal listing, a higher percentage of other nearby listings with vicinity safety reviews within a 0.3-mile radius is negatively correlated with the focal listing's occupancy rate, but positively correlated with its price. Among listings within a 0.3-mile radius area, hosts may seek to minimize such a negative externality; but from a prospective guest's perspective, this is a positive information externality that could help guests make more informed choices ex ante. Hence the information design optimal to the platform can be different from the information design optimal to guests.

Our work in part hinges on guests' reactions to safety reviews, and thus relates to the literature on the role of information disclosure in online user behavior. Researchers have shown that product attributes, seller attributes, seller-buyer interactions, and the way in which an online platform aggregates and presents such information (e.g. search ranking, product recommendations, price, consumer ratings, and images of the property) are all important elements in consumers' decisions in e-commerce and the sharing economy (Tadelis, 2016; Ert et al., 2016; Tussyadiah and Park, 2018; Ursu, 2018; Reimers and Waldfogel, 2021; Xu et al., 2021; Zhang et al., 2021c,b). Furthermore, consumer decisions depend on the quality, quantity, resource, and accuracy of the disclosed information (De Pelsmacker and Janssens, 2007; Liu et al., 2017; Munzel, 2016), which in turn depend on the extent to which the users of a platform are willing to disclose such information (Hao and Tan, 2019; Liang et al., 2019; Morosan, 2018; Morosan and DeFranco, 2015; Moon, 2000). In this paper, we focus on guests' reactions to safety reviews, while taking the existence of each historical safety review as given. To ensure the response is specific to safety reviews, we control for listing and host attributes along with listing fixed effects.

We are not the first to study safety issues regarding online short-term rental platforms. Suess et al. (2020) find that non-hosting residents with a higher emotional solidarity with Airbnb visitors are more supportive of Airbnb hosts, and residents hold different views about safety ("stranger danger") and Airbnb depending on whether they have children in the household. Local planners pay attention to the impact of online short-term rentals on neighborhood noise, congestion, safety, and local housing markets (Gurran and Phibbs, 2017; Nieuwland and Van Melik, 2020; Kim et al., 2017). Zhang et al. (2021a) shows that regulations that negatively affect Uber/Lyft services may also negatively affect the demand for Airbnb. Han and Wang (2019) document a positive association between commercial house-sharing and the rise of crime rate in a city, while non-commercial house-sharing does not have this association. A number of studies find that an increase in Airbnb listings — but not reviews — relates to more neighborhood crimes in later years (Xu et al., 2019; Maldonado-Guzmán, 2020; Roth, 2021; Han et al., 2020; Filieri et al., 2021). More specifically, Airbnb clusters are found to correlate positively with property crimes such as robbery and motor vehicle theft, but negatively with violent crimes such as murder and rape. Also, Airbnb listings of the type in which guests may share a room with other unrelated guests are found to be more related to crimes (Xu et al., 2019; Maldonado-Guzmán, 2020) and to skirting local regulations (Jia and Wagman, 2020).

Our study complements this growing literature, by highlighting safety reviews, distinguishing vicinity and listing safety reviews, and documenting consumer responses to safety reviews or experiencing safety issues. Although we cannot identify the effect of Airbnb on local crime rates, our work helps quantify guest preferences regarding safety, and clarify how the interests of guests, different hosts and the platform diverge with respect to the disclosure of safety reviews. As shown in our counterfactuals, disclosing and highlighting vicinity safety reviews could encourage guests to shy away from potentially unsafe listings and disproportionately affect hosts in certain areas.

# 3 Background of Airbnb's Review System

Over the past decade, short-term vacation rental markets have quickly expanded worldwide. Airbnb, the leading home-sharing marketplace, now offers more than 7 million listings from approximately 3 million hosts across a wide range of geographic areas (Coles et al., 2017).[2] As with any lodging accommodation, the specific location of a listing can affect the experience of its guests. For instance, if a property is located in a relatively unsafe area, crimes such as carjacking or burglary may be more likely. In Los Angeles, the number of victims to crimes such as theft or burglary at short-term rental lodgings reportedly increased by 555% in 2017-2019.[3] As is common in the lodging industry, guests, who may be traveling outside their home towns and are therefore less familiar with local neighborhoods, are responsible for their own safety in the areas in which they choose to stay. In particular, as with hotels, guests receive little to no

[2]See, for example, https://www.stratosjets.com/blog/airbnb-statistics/ and https://ipropertymanagement.com/research/airbnb-statistics.

[3]See, e.g., https://xtown.la/2020/10/16/crime-short-term-rentals/.

protection from rental platforms as far as crimes they may experience in a listing's vicinity.[4]

However, prior to making a reservation, potential guests may refer to a number of sources to gauge the safety of a listing's area — these sources include local news, crime maps, websites that summarize neighborhoods[5], and perhaps most readily linked to each listing, the listing's reviews from prior guests.[6] Airbnb enables guests and hosts to blindly review each other after a guest's stay.[7] In an effort to appease hosts, and perhaps to encourage more listings across a larger number and variety of neighborhoods, a recent Airbnb policy effective December 11 2019 announced that, going forward, guest reviews about a listing that include "content that refers to circumstances entirely outside of another's control" may be irrelevant and subject to removal.[8] This policy change implies that reviews about the safety of a listing's vicinity ("vicinity safety reviews") may be deemed irrelevant and subject to removal, since such a safety aspect is outside the control of the host. Due to data limitations, we do not know how Airbnb enforces this policy as far as vicinity safety reviews, but anecdotes suggest that some reviews that touched on neighborhood safety have been removed.[9] The policy does not apply to "listing safety reviews," because these reviews are about the safety within the listed property, which presumably can be more readily controlled and improved by the listing's host.

It is difficult to pin down exactly why Airbnb adopted this new review policy in December 2019. If Airbnb believes that the main role of online reviews is to motivate hosts to provide high-quality services to guests, review content regarding something outside the host's control may not

[4]See, for instance, `https://www.airbnb.com/help/article/2908/terms-of-service` and `https://www.vrbo.com/en-gb/legal/traveller-terms-and-conditions`.

[5]See, e.g., `https://www.neighborhoodscout.com/`

[6]Reviews have been well established as having a potential effect on buyer decisions and sellers' reputations, particularly in the tourism industry (Schuckert et al., 2015). The literature also suggests that negative information in reviews in particular can have an effect on guest decisions and be useful to platforms in distinguishing seller and product quality (Jia et al., 2021).

[7]If one side does not review the other, the other's review becomes visible after 14 days.

[8]See, for example, `https://www.airbnb.com/help/article/2673/airbnbs-review-policy` and `https://community.withairbnb.com/t5/Airbnb-Updates/Making-reviews-more-relevant-and-useful-for-our-community/td-p/1191576`.

[9]For example, on Jan. 27, 2020, a tweet from "PatrickR0820" wrote "I used @Airbnb when we went to Atlanta for the Panthers game. In my review I left numerous things that could be fixed as well as 'the area that it is located in, is pretty sketchy.' My review and 4 other similar recent reviews were deleted because it wasn't relevant." Another tweet by "AveryBrii" on May 18, 2021 stated: "@Airbnb is such a joke!!! we literally had a car stolen at the place we stayed at, didn't get refunded (which wahtever) & then i try to leave a review to inform others that it clearly was not a safe area (cops told us this & other info that i tried to include) & they didn't post." A journalist also describes his experience on Bloomberg Opinion: "Airbnb Took Down My Negative Review. Why?" (May 26, 2021 by Timothy L. O'Brien), accessed at `https://www.bloomberg.com/opinion/articles/2021-05-26/airbnb-took-down-my-negative-review-why`, on November 26, 2021.

help in that regard. Anecdotes suggest that hosts have complained about the harm they suffer from "irrelevant" reviews about the vicinity of their listings,[10] and this policy change could be a way to address these complaints. Another reason might be the concern of review accuracy: arguably, vicinity safety is a subjective feeling subject to the reviewer's prior and interpretation, and it is often difficult to prove correct or wrong. However, similar accuracy concerns could apply to other review content, though the degree of objectiveness may vary. A third reason may have something to do with the aspiration of being inclusive. Airbnb has advocated for inclusive design, which is defined as "consciously designing products, services, and environments that don't create barriers to belonging."[11] The same aspiration may have motivated Airbnb to adopt an anti-discrimination policy, establish a permanent anti-discrimination team, and encourage designs and services friendly to users with disabilities. To the extent that vicinity safety reviews are more present in low-income or minority neighborhoods, the new review policy could be another effort to make the platform more friendly to hosts in economically disadvantaged neighborhoods. The key question we address in this paper is how the new policy, if fully implemented as far as vicinity safety reviews, would redistribute the economic benefits and costs among hosts, guests, and the platform.

To be clear, Airbnb has adopted other methods to address neighborhood safety directly. For example, Airbnb introduced a neighborhood support hotline in December 2019[12], around the same time as Airbnb adopted the new review policy. This hotline is primarily intended to be a means for neighbors of Airbnb listings to contact the platform in certain situations (e.g., in the event of a party taking place at a listed property). In addition, since our dataset ends in December 2019 and we do not know how many guests that left vicinity safety reviews in our sample would have used the hotline should the hotline exist at the time of the review, we cannot predict how the hotline could counter some of the effects shown in our analysis. That being said, hotline usage is ex post and is not visible to future guests, hence its impact on guests can be fundamentally different from the impact of reviews visible under each listing on Airbnb.

[10] Nina Medvedeva, "Airbnb's Location Ratings as Anti-Black Spatial Disinvestment in Washington D.C." Platypus: The CASTAC Blog (March 16, 2021) accessed at `https://blog.castac.org/2021/03/airbnbs-location-ratings-as-anti-black-spatial-disinvestment-in-washington-d-c/`.

[11] https://airbnb.design/values-driven-design/.

[12] `https://news.airbnb.com/an-update-on-our-work-to-support-neighbors-around-the-world-2/`

Airbnb's review system also allows guests to leave a 1-5 star rating by specific categories (cleanliness, accuracy, check-in, communication, location, value, and amenities), in addition to leaving an overall rating and detailed review. According to Airbnb's response to a host's question, location rating is meant to "help future guests get a sense of the area and tends to reflect proximity to nearby destinations."[13] Hence, location rating could capture many location-specific aspects such as local transit, nearby stores, neighborhood walkability and noise, and may not be directly related to vicinity safety.

# 4 Data

The first dataset we use has information on the set of short-term rental listings that had been advertised on Airbnb from May 2015 to December 2019, and on VRBO from June 2017 to December 2019, in five US cities (Atlanta, Chicago, Los Angeles, New Orleans, and New York). The data was acquired from AirDNA, a company that specializes in collecting Airbnb and VRBO data. For Airbnb listings, this dataset includes the textual contents of all Airbnb listing reviews in those cities.

Each listing is identified by a unique property ID and comes with time-invariant characteristics such as the listing zip code, listing's property type (entire home, private room, shared room, or hotel room) as well as the host's unique identifier. Listings also have time-variant characteristics, including average daily rate,[14] the number of reservations, days that are reserved by guests, occupancy rate,[15] number of reviews, overall rating scores,[16] the listing's Superhost status,[17] the listing's guest-facing cancellation policy,[18] the average number of words in the list-

[13] `https://community.withairbnb.com/t5/Host-Voice/quot-Location-quot-As-A-Guest-Review-Point/idi-p/162137`.

[14] Average daily rate (ADR) is calculated by dividing the total revenue, including both nightly rates and cleaning fees, earned by the host from reservations over a given month by the total number of nights in that month's reservations.

[15] Occupancy rate is calculated by dividing the number of booked nights by the sum of the available nights and booked nights.

[16] Overall rating scores are normalized to 0-10 range. Our dataset also includes location star ratings. Adding it as an extra control variable does not change our main results, so we do not report it in this paper. Results are available upon request.

[17] Superhost refers to a status badge related to metrics concerning a listing's performance. Hosts who meet the following criteria, evaluated quarterly, receive a Superhost designation: (i) Completed at least 10 reservations in the past 12 months; (ii) maintained a high response rate and low response time; (iii) received primarily 5-star reviews; (iv) did not cancel guest reservations in the past 12 months.

[18] Cancellation policy could be strict, moderate, flexible. For simplicity, we use a dummy variable to indicate

ing's reviews, the number of listings in the same zip code, and whether the listing is cross-listed on VRBO.[19]

Our unit of observation is listing-month. We focus on "active listings" (listings whose calendars are not indicated as 'blocked' in the dataset for an entire month), and exclude observations with ADR over $1000, as some hosts may set their rates prohibitively high in lieu of blocking their calendars. We use regular monthly scrapes between May 2015 and December 2019 on Airbnb (July 2017 to December 2019 for VRBO). In total, the sample comprises 2,969,840 listing-months observations on Airbnb, and 164,034 listing-months observations on VRBO.

We define two different types of safety reviews — listing safety review and vicinity safety reviews. Listing safety reviews are those reviews that describe issues pertaining to safety within a listing (e.g., "the listing is unsafe because there are fire hazards", "the listing is unsafe because the floor is slippery", or "the listing is unsafe because the stairs are too small and there's no railing"). Vicinity safety reviews contain information pertaining to the safety of the nearby vicinity or neighborhood of the listing (e.g., "the neighborhood is unsafe", "dodgy neighborhood", or "unsafe area"). While there is considerable research regarding the use of machine learning for automated content analysis, these methods typically require a large number of hand-labeled examples for training. We instead use a lexicon approach due to its simplicity and transparency. Lexicons are also found to have high levels of precision as compared to machine learning approaches (Zhang et al., 2014; Hutto and Gilbert, 2014), and have been used extensively in the literature (Monroe et al., 2008; Dhaoui et al., 2017). To identify a suitable set of keywords, we use an iterative approach, starting with terms such as "unsafe," "dangerous," and "scary" and manually inspecting the matching reviews. Our aim is to have a broad definition of safety, so we include both reviews that are explicitly negative about safety (e.g., "The neighborhood is dodgy,"), as well as those that express only mild concerns (e.g., "Neighborhood was a little rough looking, but it wasn't a problem."). After several rounds, we expanded the list to 36 vicinity safety keywords and 5 listing safety keywords, as delineated in Table 1.

As far as vicinity safety reviews, to further improve the precision, and to ensure that the text is indeed describing issues pertaining to the safety of a listing's vicinity and not other

whether a listing's cancellation policy is strict or not.

[19] Only listings with entire home that could be both listed on Airbnb and VRBO.

aspects of a listing, we similarly identified a list of 21 location keywords that tend to indicate a statement about the surrounding area (e.g., "neighborhood", "area", "outside") in Table 1. We then categorized the matching reviews into those in which the vicinity safety keyword occurred within 20 words of a location keyword as vicinity safety reviews, and those in which the listing safety keyword occurred outside of the 20 word context as listing safety reviews.[20] This approach resulted in 70k matched vicinity safety reviews and 1.5k matched listing safety reviews across the 5 cities. As shown in Table 1 and Table 2, the top matching vicinity safety keywords are "unsafe" (11,648), "sketchy" (10,227), "dark" (10,208), and "safety" (7,000), and the top matching listing safety keywords are "unsafe" (588), "hazard" (384) and "slippery" (345). As an additional validation check, we sampled several hundred matches at random and manually labeled them as relevant or not, finding 86% accuracy.[21] The mislabeled data often used figurative language ("scary how perfect this neighborhood is") or used safety words in other contexts (e.g., "watched a scary movie on Netflix"). While any such method will be imperfect, we did not find any evidence suggesting that the error rates were systematically biased for some neighborhoods over others. However, we did restrict our keywords to English, so the method will be less effective in areas with many non-English reviews.

A second dataset we collect covers official crime records from databases tracking crimes in Chicago[22], New Orleans[23], New York City[24], Atlanta[25], and Los Angeles.[26] These databases cover different types of crimes, including property-related crimes and violent crimes. In terms of the geographical granularity of crimes, we consider crime events at the zip code level. We also obtain median income and other demographic information at the zip code level from 2014, one year before our Airbnb sample period begins, from the United States Census Bureau[27]. We

[20]While the 20 word window is arbitrary, a sensitivity analysis suggests no qualitative difference when using a slightly longer or shorter window. Moreover, the average review had roughly 50 words, so this seemed to restrict to the 1-2 sentences around the keyword match.

[21]This indicates a 14% false positive error rate. Since our lexicon approach aims to minimize the false positive rate while allowing false negatives, the safety reviews identified by this approach tends to make the estimated impact of safety reviews more conservative than the true effect.

[22]Official crime data in Chicago: `https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2`.

[23]Official crime data in New Orleans: `https://data.nola.gov/browse?q=crime`.

[24]Official crime data in New York City: `https://data.cityofnewyork.us/widgets/qgea-i56i`.

[25]Official crime data in Atlanta: `https://www.atlantapd.org/i-want-to/crime-data-downloads`.

[26]Official crime data in Los Angeles: `https://data.lacity.org/A-Safe-City/Crime-Data-from-2010-to-2019/63jg-8b9z`.

[27]See, e.g., `https://www.census.gov/data.html`.

make the assumption that the income and demographic information did not change significantly over our sample period. Throughout the paper, we refer to a zip code as high-income (H) or low-income (L) according to whether its average income is above or below the median of the city it locates in. Similarly, we refer to a zip code as minority (M) or white (W) according to whether its percentage of minorities in population is above or below the city median.

Table 2 summarizes the data at the listing-month level, where vicinity safety (VS) Airbnb listings are defined as observations that have at least one vicinity safety review (VSR) before the reporting month, while "normal" Airbnb listings do not have any VSR before the reporting month. As the table indicates, about 18% of the total observations are VS listings. On average, VS listings have higher occupancy rates, a higher number of reservations, a higher fraction of Superhosts, and a higher number of reviews than normal listings. In contrast, the nightly rates of VS listings are lower on average than normal listings.

The mean number of cumulative VSRs (aggregated up to the reporting month) is 0.33 across all Airbnb listings, while the mean number of cumulative listing safety reviews (LSR) is 0.005. The monthly trends for the percentages of VSRs and LSRs are depicted in Figure 3, where the percentage of VSRs is decreasing and the percentage of LSRs is increasing over time. Figure 4 and 5 demonstrate the distribution of VS keywords for four groups of zip codes (high-income, low-income, white, and minority). Comparing high-income with low-income groups, it appears that the low-income group dominates the volume of VSRs, though this trend is reversed in the minority group in comparison to the white group for some VS keywords (such as "steal," "safety," "poor," "drug," "dark," and "blight").

We also test the rank correlation between the official crime records and VSRs. Specifically, we use the percentile rank of normalized crime records in each zip code-month within each city — calculated as the number of reported crime cases in a month, divided by the size of the population in that zip code. For each month, we rank the normalized crime data within each city, and determine the percentile crime rank of the zip code for that month. For VSRs, we use the percentile rank of the number of cumulative VSRs in the zip code up to the reporting month.[28] We then test the percentile rank correlation index between the crime records and VSRs in each month, resulting in the time-series correlation trends depicted in Figure 6, which

[28]Due to data limitations, we assume that both records begin with clean slate (0 records) as of the beginning of our dataset.

illustrates the correlation trends for the four different groups of zip codes (high-income, low-income, white, and minority). Figure 6 indicates that the correlation in all four groups exhibits an increasing trend, suggesting that the cumulative number of VSRs in a zip code is more likely to reflect the actual crime reports in the zip code over time. Lower-income and minority zip codes tend to have a more significant increase in such correlation at the beginning of our sample period, compared with high-income and white zip codes, possibly due to their higher number of VSRs and crime records in the raw data.

# 5 Empirical Analysis

## 5.1 Effects of Safety Reviews

We begin by assessing the effects of VSRs and LSRs. Our hypotheses is that if potential guests view VSRs and LSRs as a proxy for safety around or within a listing, such reviews would reduce the guests' willingness to book the listing. Our base specification is given by:

$$y_{i,t} = \alpha_i + \alpha_{k,t} + \delta X_{i,t} + \beta_1 Crime_{i,t-1} + \\ \beta_2 LSR_{i,t-1} + \beta_3 VSR_{i,t-1} + \beta_4 VSRADIUS_{i,t-1} + \epsilon_{i,t}, \quad (1)$$

where $i$ denotes a listing $i$-month $t$ observation, $Crime_{i,t-1}$ is a log transformed variable that indicates the normalized number of cumulative official crime reports since the start of the sample period for the zip code where listing $i$ is located, $LSR_{i,t-1}$ and $VSR_{i,t-1}$ are two dummy variables that equal 1 if the listing has at least one LSR and VSR, respectively, before month $t$, $VSRADIUS_{i,t-1}$ is the percentage of listings that have at least one VSR within a 0.3-mile radius of listing $i$ prior to month $t$, $X_{i,t}$ are listing-level controls (logged except for dummy variables), including the number of reviews, overall ratings, cancellation policy, number of listing in the same zip code, cross-listing status (i.e., whether the listing is also listed on VRBO), and whether the listing is hosted by a Superhost. The dependent variable $y_{i,t}$ is either the log of listing $i$'s average daily rate (ADR) in month $t$, or the log of listing $i$'s monthly occupancy rate (calculated as log of 1 plus the occupancy rate).[29] Listing and City–year-month fixed effects are denoted

[29]Some listing-month observations have an occupancy rate of 0 and consequently are missing an average reserved daily rate in the dataset for those months, though the dataset does offer a separate "listing price" (i.e.,

by $\alpha_i$ and $\alpha_{k,t}$, respectively, where the city of listing $i$ is denoted by $k$. Standard errors are clustered by Airbnb property ID. The primary assumption is that, within a listing, the presence and timing of safety reviews are correlated with the true safety condition around or inside the listing and do not reflect selective reporting, fake reviews, or other strategic reasons once we control for other time-varying listing attributes.

Our main specifications (Columns 2 and 4 of Table 3) indicate that both VSRs and LSRs significantly decrease a listing's price (ADR) and occupancy. Specifically, for an average Airbnb listing in our sample, having any VSR is associated with a 1.01% reduction in the listing's monthly occupancy rate and a 1.33% reduction in its average price per reserved night; having an LSR is associated with a 2.59% drop in occupancy and 1.72% in price. LSRs thus have a larger effect on price and occupancy than VSRs, possibly because some prospective guests have a specific geographic area (e.g., neighborhood) in mind, regardless of safety issues concerning that area, whereas LSRs describe safety issues that pertain to the listing itself.

In contrast, normalized official crime records and the percentage of listings with VCRs within a 0.3-mile radius are both associated with lower prices but higher occupancy. A potential explanation is that hosts are aware of safety issues in the areas of their listings, and proactively lower their rates when their listings are located in relatively unsafe areas. These lower prices attract more guest bookings, perhaps either because guests tend not to seek information about crimes in the neighborhood or because they prioritize price. In particular, for the average Airbnb listing in our sample, given a 1% increase in the normalized official crime records, the daily rate is 0.065% lower whereas the occupancy rate is 0.067% higher.

## 5.2 Robustness

Column 1 of Table 4 considers as the dependent variable a dummy that equals 1 when a listing's occupancy rate is positive and 0 otherwise. It reports a positive coefficient on $Crime_{i,t-1}$, suggesting that the variable $Crime_{i,t-1}$ not only describes the relative crime status of a zip code, but may also capture the relative guest traffic to the area, where areas with relatively high guest traffic (e.g., downtown areas) tend to have a higher number of reported (normalized) crimes.

a base rate) for those listings. To extrapolate the ADR of these listings in the months in which they are missing, we calculate the mean ratio of their ADR to their listing price in the months in which they are available, and multiply this average by the listing price in the missing months (if available, or by using the listing price from the nearest month in which it is available).

Comparing the coefficients on VSRs and LSRs for the whole-sample specifications (Columns 2 and 4 of Table 3) to the conditional sample with positive occupancy rates (Columns 2 and 7 of Table 4), we find that the coefficients are similar but have somewhat higher magnitudes for the whole sample.

We conduct a number of additional checks. First, we split the sample by whether a listing has an above- or below-median number of reviews in a given month (median is 12), as a proxy for whether the listing is in its early or later "stage" of taking guest reservations, since only staying guests can post a review.[30] Another motivation for this partition is that prospective guests are more likely to notice safety reviews (both VSRs and LSRs) when listings have a lower number of reviews. Indeed, Columns 3 and 4 of Table 4 report that in the subsample of listings with 12 or fewer reviews, the negative effects of having any VSR and LSR on occupancy rate (0.98% for VSR and 4.39% for LSR) are higher than the corresponding negative effects for listings with more than 12 reviews (0.24% for VSR and 1.60% for LSR). However, Columns 8 and 9 indicate that as far as listings' daily rates are concerned, this comparison is reversed, possibly because hosts of newer listings may still be in the process of identifying their pricing for those listings.

Second, we add additional controls for the average word count of a listing's reviews, as well as for the number of properties hosted by the listing's host.[31] As Columns 5, 6, 10 and 11 of Table 4 indicate, the results do not qualitatively change from our main specifications when incorporating these additional controls.

## 5.3 Heterogeneous Effects

We next explore a number of heterogeneous effects. Table 5 provides summary statistics based on the type or area of a listing. In particular, the table reports different normalized zip code crime levels for listings in these categories. We proceed with a similar empirical methodology as in (1), but with different subsamples.

We begin by analyzing four groups of zip codes separately (high-income, low-income, white, and minority). Table 6 shows that VSRs have negative effects on occupancy rates across all four

[30]To be clear, the same listing may be in both subsamples over time, but belong to only one of the subsamples in any given month.

[31]Host responses to safety reviews are not observed in our data

subsamples. The negative effects of having any VSR have higher magnitudes in high-income and white zip codes (1.10% and 1.09%) than in low-income and minority zip codes (0.88% and 0.89%). A similar comparison holds for LSRs. One potential explanation is that guests may have different prior beliefs and different sensitivities to safety issues, and perhaps more so if their search targets a specific area that they believe is relatively safe. Hosts in different areas may also react differently to VSRs and LSRs, based on how they gauge guest perception and guest preferences.

We next consider subsamples comprising different listing types (entire home, private room, shared room, and hotel room). Additional heterogeneous effects may arise here because, for instance, for guests who seek partial spaces (private room, shared space) within a dwelling, safety issues may be more salient. The results in Table 7 indeed show that the magnitude of the negative effects from having any VSRs and LSRs on occupancy are larger for private rooms and shared spaces (1.03% and 1.31% for VSR and 3.29% and 3.52% for LSR, respectively) in comparison with entire-home listings (1.00% for VSR and 2.41% for LSR).

## 5.4 Safety Experience and Future Guest Activity on Airbnb

We conduct user-level analyses to test whether guests who leave any vicinity safety reviews (henceforth, VS guests) act differently before and after they post their first VSR in comparison to otherwise similar guests who did not leave any VSRs. To that end, we assume that the first VSR that a VS guest posts for one of the listings in our sample (i.e., covering Airbnb listings in the five cities we consider, with reviews beginning in May 2015) is the first VSR that this guest posted. To reiterate, any such guests who have ever posted VSRs in our sample are considered VS guests; otherwise, they are treated as 'normal' users. To ensure that the VS users have had some experience on Airbnb prior to leaving their first VSR, we focus on the subset of VS users that left at least three reviews in the five sample cities before leaving their first VSR.

In order to match VS users with normal users, we use a K-nearest neighbor (KNN) method to select the two most similar control (normal) users for each treatment (VS) user. The user characteristics used in the KNN method (as of the time of the treatment user's first VSR) are the user's number of prior reviews, the average normalized crime reports in the cities in which the user stayed (based on their prior reviews), the average number of VSRs for listings for which the user left reviews, and the average percentage of overall VSRs and VS listings in the same zip codes as listings for which the user had previously left reviews. The matching is done for each month (i.e., based on new treatment users in each month). The same "treatment month" is applied (hypothetically) to control users that are matched with a treatment (VS) user, based on the latter's timing of their first VSR.

In order to assess if the treatment and control users have the same tendency to post VSRs, we also calculate the propensity score for each user in our matched sample. In particular, we regress the dummy of a user being treated (i.e. being a VS user) on the number of reservations she had made on Airbnb before the treatment time, the average zip code-wide crime rate of these reservations at the time of reservation, the average number of VSRs in these reservations, and the percent of listings with any VSR in the zip code of these reservations. For a treated user, the treatment time is when she wrote her first VSR in our sample. For a control user, the treatment time is when the treatment user she is paired with wrote her first VSR in our sample. The result of this logistic regression is presented in Table 8. Panel A of Table 9 reports that the treatment and control users are similar as far as the characteristics considered in the KNN method (i.e., before the treatment users' first VS review — Panel B of Table 9 reports the characteristics for the two user groups considering all of their reviews); the two user groups also have similar propensity scores, as shown in Figure 7.

We first test whether VS users behave differently in terms of subsequent reservations on Airbnb after their first VSR (as exhibited by their subsequent listing reviews). We use a difference-in-differences methodology (DID) as follows:

$$y_{i,t} = \alpha_t + \alpha_p + \beta \cdot treat_i + \gamma \cdot treat_i \times post_t + \epsilon_{i,t}, \tag{2}$$

where the subscript $p$ denotes the treatment-control pair we have identified in the sample construction and $\alpha_p$ is pair fixed effects. We have constructed several measures for the dependent variable $y_{i,t}$: the first is the number of reviews that user $i$ wrote in month $t$. We use it as a proxy of user $i$'s Airbnb reservations in $t$, which can be zero. Because it is a count variable, we use a Poisson regression instead of ordinary least squares. The second measure is the normalized cumulative count of officially reported crimes in the zip codes of user $i$-reviewed listings in month $t$. The other measures are the number of VSRs in the reserved listings, the percentage of VS listings in the zip codes of the reserved listings, and whether the reserved listings have any VSR. The dummy variable $treat_i$ equals 1 for VS users and 0 otherwise, and the dummy variable $post_t$ equals 1 if $t$ is after the time of the first VSR of user $i$. Year-month fixed effects are denoted by $\alpha_t$, standard errors are robust and clustered by treatment-control pairs.

Column 1 of Table 10 reports results from an unbalanced monthly panel data, indicating that VS users tend to book fewer reservations (as evidenced by subsequent reviews) after posting their first VSR. In particular, the average monthly number of subsequent reviews is expected to be 31.06% lower for VS users in comparison with normal users.[32]

We also assess whether VS users are more sensitive to safety information when booking subsequent Airbnb listings after posting their first VSR. In order to test this hypothesis, we use the booked listings' characteristics as the dependent variables. Columns 2-6 of Table 10 suggest that the subsequent listings chosen by VS users exhibit the following characteristics: They tend to be located in zip codes that have fewer normalized crime reports, they are less likely to have VSRs, and they are less likely to be located in zip codes that have a higher overall percentage of VSRs or a higher percentage of other listings with VSRs. This suggests that VS users, relative to normal users, are more sensitive to safety information after posting their first VSR.

We further examine whether VS users subsequently act differently as a function of the area (high-income, low-income, minority or white) in which they posted their first VSR. To do so, we group VS users according to the zip code of the listing for which they posted their first VSR, and proceed to conduct the DID analysis separately for each of the four subsamples.

Table 11 reports that VS users tend to book subsequent stays (as proxied by their subsequent listing reviews) in areas that are the opposite of where they posted their first VS review. That is, VS users whose first VSRs are posted in high-income or white areas tend to book fewer subsequent stays in those areas but more in low-income or minority areas, and vice versa. One possible explanation for the former direction is that VS users expected a higher level of safety in high-income or white areas, and when they encountered the opposite, they preferred to pay the average lower rates for listings in low-income and minority areas in subsequent stays. A potential explanation for the latter direction is that VS users associate safety issues with

[32]This is not the coefficient of the treatment dummy (-0.372) because we use a Poisson model for this regression, i.e., the applicable percentage is $1 - e^{-.372}$.

low-income or minority areas, and tend to avoid such areas in future bookings.

From the interaction term in Table 11, it is apparent that VS users exhibit a positive effect on subsequent reservations in opposite zip codes (Columns 2, 3, 6, and 7) and a negative effect on in the same zip codes (Columns 1, 4, 5, and 8), and that the effect has a higher magnitude for reservations in high-income or white areas (Columns 1, 2, 5, and 6) compared with low-income or minority areas (Columns 3, 4, 7, and 8). However, since the number of VS users who post their first VSR in high-income or white areas is fewer than those in the low-income or minority areas (with a ratio around 4:6), those that shy away from high-income or white areas are fewer than those that switch to such areas because of their safety experience in low-income or minority areas. As a result, the overall effect on subsequent bookings (as proxied by the total number of subsequent reviews) is positive in high-income or white areas but negative in low-income or minority areas.

# 6 Back-of-the-Envelope Calculations

## 6.1 Airbnb's Gain and Loss

So far, our analysis has shown that (1) VSRs have negative effects on Airbnb listings' prices and occupancy, and (2) guests who have posted any VSRs appear to book fewer subsequent reservations and are more sensitive to safety information than other guests that never posted any VSRs.

These findings suggest that the status quo (in our data period, 2015-2019) has disclosed vicinity safety reviews but they are not as salient as they could be. For some listings, prospective guests can find vicinity safety information in prior consumer reviews, but the guests that incurred a vicinity safety issue during their own stay at Airbnb listings become more alert to safety information than other users, likely because self experience is more salient than safety reviews from other anonymous users. The status quo also implies diverging interests in the information value of vicinity safety reviews: while guests view VSRs as a negative but informative attribute of a listing, the host of VS listings may perceive VSRs as a harm to future business. In contrast, the hosts of normal listings may consider their lack of VSRs as a competitive advantage over VS listings.

To highlight these diverging interests, we run back-of-envelope comparisons with respect to the revenues of VS and normal listings, under three information regimes: (i) the status quo, (ii) no disclosure (where all VSRs are removed), and (iii) high alert (where all users react to VSRs as much as VS users react to their own reported VSR). To understand heterogeneity across areas, we run these back-of-envelope calculations for high-income, low-income, minority, and white areas.

For the no-disclosure counterfactual, we set vicinity safety reviews to zero while holding everything else equal. This implies that guests take zero reviews as literally zero and would not reinterpret the lack of vicinity safety reviews in the counterfactual. This assumption is reasonable because vicinity safety reviews are rare (only 1.23% of reviews are identified as safety reviews and only 18% of listings have ever had any safety reviews), many guests are casual users that are likely inattentive to Airbnb policy changes, and most guests may not know a listing could have no vicinity safety reviews because they are removed by the platform.[33] To run the counterfactual, we use the results in Table 6, which capture the effects of VSRs and the percentage of VS listings within a 0.3-mile radius on price and occupancy, in high-income, low-income, white, and minority areas. We next collect the number of Airbnb observations, average occupancy in days, average price (ADR), and average percentage of VS listings within a 0.3-mile radius area for both VS and normal listings in these four areas. The gain/loss of VS listings from the no-disclosure regime (from May 2015 to December 2019) is calculated using the change in occupancy rate and price, had their VSRs and the VSRs of other VS listings within a 0.3-mile radius area been removed. The gain/loss of normal listings from the no-disclosure regime is calculated as the change in occupancy and price had there been no VS listings in the 0.3-mile radius area.

For the high-alert counterfactual, where all prospective guests behave as VS users, we collect the number of reservations (as proxied by the number of reviews) and average days per reservation of all VS users for both VS and normal listings in each of the four demographic areas. The gains/loss are calculated as the decrease in revenue that results from the reduction in the number of reservations due to VSRs, should all normal users react in the same way as

[33] In theory, if most guests are fully aware of the no-disclosure policy and correctly predict the extent of vicinity safety reviews suppressed by the policy, they could readjust their safety perceptions of all listings. It is difficult to derive the new equilibrium because the adjustments in safety perceptions depend on whether and how guests may search for vicinity safety information elsewhere.

VS users, based on the results from Table 11.

The calculations are summarized in Table 12. The results suggest that the switch from the status quo to no disclosure leads to gains for VS listings in all four demographic areas. The revenues of VS listings in low-income and minority zip codes increase by 2.49% and 2.41%, respectively, which is higher than the corresponding increases of VS listings in high-income and white zip codes (2.17% and 2.11%, respectively). One possibility is that low-income and minority zip codes tend to have more VSRs, hence they stand to benefit more when such reviews are eliminated. Normal listings in high-income and white zip codes, on the other hand, are harmed when no VSRs are available, with percentage revenue decreasing 0.15% and 0.23%, respectively. However, normal listings in low-income and minority zip codes stand to gain slightly when switching to a regime of no disclosure, with percentage revenue increasing 0.14% and 0.15%, respectively. A potential explanation is that normal listings in low-income and minority zip codes have a higher percentage of VS listings within 0.3-mile radius area, and thus suffer from a higher negative spillover effect on daily rates under the status quo. When VSRs are unavailable, they stand to benefit from higher prices.

Table 12 also reports results for the high-alert counterfactual, where prospective guests behave as VS guests. Both VS and normal listings in low-income and minority zip codes stand to lose revenues in the high-alert regime as compared to the status quo, with 8.83% and 9.09% declines for VS listings and 7.48% and 7.92% declines for normal listings. And, their counterparts' VSL in high-income and white zip codes alsp lose revenues if the regime is changed from the status quo to high alert. This is because, after guests become more alert to VSRs, the guest switches from low-income and minority areas into high-income and white areas dominate those that switch away in the other direction, given the fact that VSRs are less likely to occur in high-income and white zip codes.

From the platform's perspective, the overall revenue or GBV sums up the revenue gains and losses across all areas. As shown in Panel C of Table 12, shifting from the status quo to the no-disclosure regime will increase Airbnb's GBV by 0.46%. In comparison, a shift from the status quo to the high-alert regime will reduce Airbnb's GBV by 8.01%.

## 6.2 Consumer Surplus

So far, the back-of-the-envelope calculations have focused on listing revenues under the three disclosure regimes. We now aim to do the same on the guest side.

To do so, we define the market as online short-term entire-home rentals in each zip code-month, where Airbnb and VRBO are assumed to be the only two platforms that supply this market. Each guest chooses among all Airbnb entire-home listings available in the target zip code-month, with the pool of VRBO-only listings in the same zip code-month as the outside good.[34] We focus on entire-home listings because only entire-home listings are available on VRBO. Since our VRBO data period is from June 2017 to December 2019, our analysis in this subsection considers Airbnb entire-home listings from June 2017 to December 2019 only.

Following Berry (1994), we assume that each prospective guest chooses an Airbnb entire-home listing or the outside good (VRBO) so as to maximize her utility from the listing, where the utility associated with an Airbnb listing $i$ in zip code $z$ of city $k$ and month $t$ can be written as:

$$\begin{aligned} U_{i,t} &= EU_{i,t} + \epsilon_{i,t} \\ &= \alpha_i + \alpha_{k,t} + \delta \cdot X_{i,t} + \beta_0 \cdot log(ADR_{i,t}) + \beta_1 \cdot Crime_{z,t-1} \\ &+ \beta_2 \cdot LSR_{i,t-1} + \beta_3 \cdot VSR_{i,t-1} + \beta_4 \cdot VSRADIUS_{i,t-1} + \epsilon_{i,t}. \end{aligned}$$

If $\epsilon_{i,t}$ conforms to the logistic distribution, we can express the market share of listing $i$ at time $t$ as $s_{i,t} = \frac{exp(EU_{i,t})}{1+\sum_j exp(EU_{j,t})}$. Thus:

$$ln(s_{i,t}) - ln(s_{0,t}) = EU_{i,t} \tag{3}$$

This is equivalent to regressing the difference of log market share between listing $i$ and the outside good ($ln(s_{i,t}) - ln(s_{0,t})$) on the attributes of listing $i$ in month $t$. The right-hand side of Equation 3 is similar to Equation 1 except for two changes: first, we exclude the number of Airbnb listings in the zip code-month because the discrete choice model already accounts for the size of the choice set; second, we include the log of the listing's ADR (i.e. price). To the extent that $log(ADR)$ might be endogenous, we instrument it by using the ADRs of private-

[34] Listings that co-list on Airbnb and VRBO are treated as Airbnb listings, thus inside goods.

room listings in the same zip code-month. The underlying assumption is that entire-home and private-room listings appeal to different sets of guests, but share common cost shocks in the same zip code-month. As shown in the first column of Table 13, the instrument is strongly correlated with $log(ADR)$, and the first stage F-statistics is high (327). The OLS and IV estimation results of the utility function are reported in the last two columns of Table 13. The results suggest that guest reservations are sensitive to price, and guests dislike listings with any VSRs or LSRs, everything else being equal.

Using the IV results in Column 4 of Table 13, we then calculate $EU_{i,t}$ for each Airbnb listing-month under the status quo, and normalize it into US dollars.[35] The sum of guest utility weighted by the simulated market shares give us the total consumer surplus under the status quo.

For the counterfactual of no-disclosure, we set all VSRs as zero in the utility function, recompute $EU_{i,t}$ for each Airbnb entire-home listing, and simulate its market share. This calculation assumes everything else remains the same when the platform removes all VSRs. It could be violated if listings adjust prices after the regime shift. Unfortunately, the vast majority of our data precede Airbnb's new review policy, so we cannot observe such price adjustments directly. The reduced-form regressions in Table 6 describe the relative price difference between VS and normal listings in the four demographic areas (under the status quo). In an alternative calculation, we assume the no-disclosure regime would erase the price discounts of VS listings while the pricing of normal listings remains unchanged. This gives us a comparison between no disclosure with price changes versus no disclosure without price changes.

To consider the high-alert counterfactual, we use the treatment effect estimated in Column 1 of Table 10 to adjust the coefficient on VSR in the utility equation. In particular, Column 1 of Table 10 estimates the coefficient of the interaction between treated and post as -0.372, suggesting that VS users would reduce their average monthly Airbnb reservations by 31.06% after they posted their first VSR on airbnb. Assuming this effect is completely driven by the coefficient on VSR in the utility equation, we calibrate how much this coefficient has to decline (i.e. become more negative) in order to generate the same decline as estimated in Table 10. We

[35]Normalized $EU_{i,t} = EU_{i,t} \cdot ADR_{i,t}/|\beta_0| + constant$ where $\beta_0$ is the estimated coefficient of $log(ADR)$ in Equation 3 and the constant is chosen such that the normalized $EU$ is always positive. Since we use the same constant when we compute utility in different scenarios, the value of the constant does not affect any comparison between scenarios.

then use the calibrated coefficient of VSR in the utility function (and all the other coefficients estimated in Table 13) to simulate market shares under the high-alert regime without price changes. Since we do not know how much prices would adjust as a result of the shift to the high-alert regime, in an alternative scenario we assume an ad-hoc price change (-1% for VS listings), to illustrate how price changes may alleviate the impact of making all users highly alert regarding VSRs. The resulting simulation is presented as high-alert with price changes.

Table 14 reports the consumer surplus results under the above four counterfactual scenarios (no disclosure with and without price changes, high alert with and without price changes), separating VS listings (on Airbnb), normal listings (on Airbnb), and VRBO-only listings. A particular element in the consumer surplus calculations is worth mentioning: because guest perception of safety can be different from guests' real experience of safety, the realized consumer surplus should use the utilities that represent guests' *realized* utilities and the simulated choice of market shares based on their *perceived* utilities as shown in Table 8. More specifically, we assume the utility function described above for each scenario represents the perceived utility, and a guest's realized utility is represented by her utility when the coefficient on VSR in her utility function is the same as the coefficient we have calibrated for VS users.

Table 14 indicates that, under the regime of high alert without price changes, consumer surplus from VS listings would decrease by 33.67% in comparison to the status quo, mostly because highly-alert guests would switch away from VS listings towards normal and VRBO listings. A hypothetical 1% price drop for VS listings (in the regime of high alert with price changes) may partially compensate the loss, leading to a smaller decline of consumer surplus from VS listings (26.92%) in comparison to the status quo. At the same time, consumer surplus from normal and VRBO listings (under high alert without price changes) increases by 6.37% and 3.90%, respectively, than the status quo, and by 5.15% and 3.16% if we incorporate the hypothetical 1% price drop of VS listings. Overall, consumer surplus under the high-alert counterfactual increases slightly relative to the status quo (0.213% without price changes and 0.210% with price changes), because the high-alert regime helps guests to reduce stays in relatively unsafe listings.

For the same reason, consumer surplus under the no-disclosure counterfactual declines slightly as compared to the status quo (by 0.087% with price change and 0.079% without price changes), because consumers cannot use VSRs as an information source to sort between

VS, normal and VRBO listings.

# 7 Discussion

Combined, the findings from our back-of-the-envelope and consumer surplus calculations suggest that the platform faces a tradeoff. On the one hand, the results suggest that all Airbnb listings, and thus the platform, stand to lose revenue under the high-alert regime. Moreover, listings in low-income and minority zip codes stand to lose a disproportionate share of their revenues than their counterparts in high-income and white zip codes. On the other hand, consumer surplus under the high-alert regime is higher than under the status quo and the no-disclosure regimes. The platform thus faces a tradeoff of generating higher revenues and attracting hosts in low-income and minority areas on the one hand, and providing additional value to its buyers on the other.

Such tradeoffs are becoming more common. YouTube, for instance, has recently adopted a policy of hiding dislike counts on shared videos,[36] and Instagram has considered giving users the option of hiding likes.[37] As digital platforms expand, tradeoffs regarding information disclosure are likely to attract more attention from researchers, user groups, and policymakers, and the consequences are likely to be more pronounced.

# 8 Conclusion

Examining the effects of vicinity safety reviews and listing safety reviews on listing performance, we find that they both negatively affect occupancy and price, and the effect from listing safety reviews is stronger. We also demonstrate that for guests that post about vicinity safety issues, concerns about vicinity safety appear to be more salient, such that they are less likely to book further stays on Airbnb, and when they do book, they tend to book in areas with fewer official crime reports and fewer vicinity safety reviews. Using back-of-the-envelope calculations, we show that expanding the disclosure of vicinity safety issues may disproportionately affect hosts in low income and minority areas, and that a GBV-centric platform may prefer to limit

[36]https://www.nbcnews.com/tech/tech-news/youtube-hiding-dislike-counts-effort-protect-content-creators-harassme-rcna5232

[37]https://about.instagram.com/blog/announcements/giving-people-more-control

the disclosure of safety information about the vicinity of listings altogether, even though the aggregate surplus of guests appears to increase when the related safety reviews are instead emphasized to alert prospective guests.

To the extent that being inclusive is one motivation behind Airbnb's new review policy, our findings suggest that the policy, if fully implemented, may have some unintended consequences on consumers and listings without safety reviews. How to balance the economic interests of all users is a challenge to platforms and policy makers that strive to maximize social welfare. One potential solution is that the platform may import external information about vicinity safety and present it as an alternative to vicinity safety reviews for each listing. Unfortunately, not all cities publish official crime statistics as the five cities in our sample do, and crime statistics may not fully capture all of the safety concerns a guest may have in mind at the time of booking. How to overcome these data difficulties and how to design an objective, universal, and user-friendly metric of vicinity safety certainly merits future research.

There are a number of limitations to our analyses due to the limitations of the data. First, the listing reviews in our data do not include potential responses from hosts. On Airbnb, hosts can reply to guests' reviews, which may also play a role in prospective guests' decisions. Second, in the user-level analysis, we only observe a user's reservation provided that they post a review. It is unclear whether guests are more, less or equally likely to post subsequent reviews after posting their first VS review. More specifically, if VS users are more vocal and thus more likely to post subsequent reviews after their first VS review, then our findings underestimate the magnitude of the effects on their subsequent booking activity; if, however, VS users are less likely to post subsequent reviews, then our findings overestimate the effects. Third, the users in our user-level analysis are limited to those users who have ever made reservations in the five major US cities we consider. Fourth, we do not have listing reviews for VRBO listings nor did we consider hotels as an outside option in our utility estimation.[38] Fifth, our data analysis ends in December 2019, the same month when Airbnb announced its new review policy. Because we do not know exactly how Airbnb implements its new policy on "irrelevant" reviews, our simulations about no-disclosure and high-alert counterfactuals are hypothetical, and do not account for other changes in which Airbnb guests and hosts may engage should

[38]Hotels, in particular, may offer enhanced safety measures to their guests through security arrangements and by having door and security staff.

these counterfactual regimes happen in reality. In particular, we do not know how guests may readjust their beliefs regarding vicinity safety for all listings if they are fully aware that the lack of vicinity safety reviews is driven by a platform policy rather than user experience. In that case, they may seek safety information from alternative sources, and adjust their perspective regarding the vicinity safety of all listings on the platform.

These limitations suggest directions for future work. In particular, VRBO does not have a policy of discouraging reviews about the vicinity of listings, as Airbnb introduced in December 2019. This may facilitate an interesting comparison between VRBO and Airbnb listings in the same locales, given a sample period that encompasses Airbnb's implementation of its new review policy.

# References


Barach, M. A., J. M. Golden, and J. J. Horton (2020): "Steering in online markets: the role of platform incentives and credibility," *Management Science*, 66, 4047–4070.

Berry, S. (1994): "Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242–262.

Bolton, G., B. Greiner, and A. Ockenfels (2013): "Engineering trust: reciprocity in the production of reputation information," *Management science*, 59, 265–285.

Cabral, L. and L. Li (2015): "A dollar for your thoughts: Feedback-conditional rebates on eBay," *Management Science*, 61, 2052–2063.

Chakravarty, A., Y. Liu, and T. Mazumdar (2010): "The differential effects of online word-of-mouth and critics' reviews on pre-release movie evaluation," *Journal of Interactive Marketing*, 24, 185–197.

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

Coles, P. A., M. Egesdal, I. G. Ellen, X. Li, and A. Sundararajan (2017): "Airbnb usage across New York City neighborhoods: Geographic patterns and regulatory implications," *Forthcoming, Cambridge Handbook on the Law of the Sharing Economy*.

Conitzer, V., N. Immorlica, J. Letchford, K. Munagala, and L. Wagman (2010): "False-name-proofness in social networks," in *International Workshop on Internet and Network Economics*, Springer, 209–221.

Conitzer, V. and L. Wagman (2014): "False-name-proof voting over two alternatives," *International Journal of Game Theory*, 43, 599–618.

Dai, W. D., G. Jin, J. Lee, and M. Luca (2018): "Aggregation of consumer ratings: an application to Yelp. com," *Quantitative Marketing and Economics*, 16, 289–339.

De Pelsmacker, P. and W. Janssens (2007): "A model for fair trade buying behaviour: The role of perceived quantity and quality of information and of product-specific attitudes," *Journal of business ethics*, 75, 361–380.

Dhaoui, C., C. M. Webster, and L. P. Tan (2017): "Social media sentiment analysis: lexicon versus machine learning," *Journal of Consumer Marketing*.

Einav, L., C. Farronato, and J. Levin (2016): "Peer-to-peer markets," *Annual Review of Economics*, 8, 615–635.

Ert, E., A. Fleischer, and N. Magen (2016): "Trust and reputation in the sharing economy: The role of personal photos in Airbnb," *Tourism management*, 55, 62–73.

Filieri, R., E. Raguseo, and C. Vitari (2021): "Extremely negative ratings and online consumer review helpfulness: the moderating role of product quality signals," *Journal of Travel Research*, 60, 699–717.

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

Fradkin, A., E. Grewal, D. Holtz, and M. Pearson (2015): "Bias and Reciprocity in Online Reviews: Evidence From Field Experiments on Airbnb." *EC*, 15, 15–19.

Gurran, N. and P. Phibbs (2017): "When tourists move in: how should urban planners respond to Airbnb?" *Journal of the American planning association*, 83, 80–92.

Han, W. and X. Wang (2019): "Does Home Sharing Impact Crime Rate? A Tale of Two Cities." in *ICIS*.

Han, W., X. Wang, M. Ahsen, and S. Wattal (2020): "Does Home Sharing Impact Crime Rate? An Empirical Investigation," *An Empirical Investigation (January 16, 2020)*.

Hao, L. and Y. Tan (2019): "Who wants consumers to be informed? Facilitating information disclosure in a distribution channel," *Information Systems Research*, 30, 34–49.

Hui, X., T. J. Klein, K. Stahl, et al. (2021): "When and Why Do Buyers Rate in Online Markets?" Tech. rep., University of Bonn and University of Mannheim, Germany.

Hutto, C. and E. Gilbert (2014): "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8.

Jia, J., G. Z. Jin, and L. Wagman (2021): "Platform as a Rule Maker: Evidence from Airbnb's Cancellation Policies," Working Paper.

Jia, J. and L. Wagman (2020): "Platform, Anonymity, and Illegal Actors: Evidence of Whac-a-Mole Enforcement from Airbnb," *The Journal of Law and Economics*, 63, 729–761.

Kim, J.-H., T. C. Leung, and L. Wagman (2017): "Can restricting property use be value enhancing? Evidence from short-term rental regulation," *The Journal of Law and Economics*, 60, 309–334.

Klein, T. J., C. Lambertz, G. Spagnolo, and K. O. Stahl (2009): "The actual structure of eBay's feedback mechanism and early evidence on the effects of recent changes," *International Journal of Electronic Business*, 7, 301–320.

Klein, T. J., C. Lambertz, and K. O. Stahl (2016): "Market transparency, adverse selection, and moral hazard," *Journal of Political Economy*, 124, 1677–1713.

Kovbasyuk, S. and G. Spagnolo (2018): "Memory and markets," *Available at SSRN 2756540*.

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

Li, L., S. Tadelis, and X. Zhou (2020): "Buying reputation as a signal of quality: Evidence from an online marketplace," *The RAND Journal of Economics*, 51, 965–988.

Li, L. and E. Xiao (2014): "Money talks: Rebate mechanisms in reputation system design," *Management Science*, 60, 2054–2072.

Liang, S., H. Li, X. Liu, and M. Schuckert (2019): "Motivators behind information disclosure: Evidence from Airbnb hosts," *Annals of Tourism Research*, 76, 305–319.

Liu, M., E. Brynjolfsson, and J. Dowlatabadi (2021): "Do digital platforms reduce moral hazard? The case of Uber and taxis," *Management Science*.

Liu, Y., J. Feng, and X. Liao (2017): "When online reviews meet sales volume information: is more or accurate information always better?" *Information Systems Research*, 28, 723–743.

Luca, M. and G. Zervas (2016): "Fake it till you make it: Reputation, competition, and Yelp review fraud," *Management Science*, 62, 3412–3427.

Maldonado-Guzmán, D. J. (2020): "Airbnb and crime in Barcelona (Spain): testing the relationship using a geographically weighted regression," *Annals of GIS*, 1–14.

Mayzlin, D., Y. Dover, and J. Chevalier (2014): "Promotional reviews: An empirical investigation of online review manipulation," *American Economic Review*, 104, 2421–55.

Monroe, B. L., M. P. Colaresi, and K. M. Quinn (2008): "Fightin'words: Lexical feature selection and evaluation for identifying the content of political conflict," *Political Analysis*, 16, 372–403.

Moon, Y. (2000): "Intimate exchanges: Using computers to elicit self-disclosure from consumers," *Journal of consumer research*, 26, 323–339.

Morosan, C. (2018): "Information disclosure to biometric e-gates: The roles of perceived security, benefits, and emotions," *Journal of Travel Research*, 57, 644–657.

Morosan, C. and A. DeFranco (2015): "Disclosing personal information via hotel apps: A privacy calculus perspective," *International Journal of Hospitality Management*, 47, 120–130.

Munzel, A. (2016): "Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus," *Journal of Retailing and Consumer Services*, 32, 96–108.

Nieuwland, S. and R. Van Melik (2020): "Regulating Airbnb: how cities deal with perceived negative externalities of short-term rentals," *Current Issues in Tourism*, 23, 811–825.

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

Pan, Y. and J. Q. Zhang (2011): "Born unequal: a study of the helpfulness of user-generated product reviews," *Journal of retailing*, 87, 598–612.

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

Romanyuk, G. and A. Smolin (2019): "Cream skimming and information design in matching markets," *American Economic Journal: Microeconomics*, 11, 250–76.

Roth, J. J. (2021): "Home Sharing and Crime Across Neighborhoods: An Analysis of Austin, Texas," *Criminal justice review*, 46, 40–52.

Schuckert, M., X. Liu, and R. Law (2015): "Hospitality and tourism online reviews: Recent trends and future directions," *Journal of Travel & Tourism Marketing*, 32, 608–621.

Staats, B. R., H. Dai, D. Hofmann, and K. L. Milkman (2017): "Motivating process compliance through individual electronic monitoring: An empirical examination of hand hygiene in healthcare," *Management Science*, 63, 1563–1585.

Suess, C., K. M. Woosnam, and E. Erul (2020): "Stranger-danger? Understanding the moderating effects of children in the household on non-hosting residents' emotional solidarity with Airbnb visitors, feeling safe, and support for Airbnb," *Tourism Management*, 77, 103952.

Tadelis, S. (2016): "Reputation and feedback systems in online platform markets," *Annual Review of Economics*, 8, 321–340.

Tussyadiah, I. P. and S. Park (2018): "When guests trust hosts for their words: Host description and trust in sharing economy," *Tourism Management*, 67, 261–272.

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

Wagman, L. and V. Conitzer (2008): "Optimal False-Name-Proof Voting Rules with Costly Voting." in *AAAI*, 190–195.

Xu, X., S. Zeng, and Y. He (2021): "The impact of information disclosure on consumer purchase behavior on sharing economy platform Airbnb," *International Journal of Production Economics*, 231, 107846.

Xu, Y.-H., L. Pennington-Gray, and J. Kim (2019): "The sharing economy: A geographically weighted regression approach to examine crime and the shared lodging sector," *Journal of travel research*, 58, 1193–1208.

Zervas, G., D. Proserpio, and J. W. Byers (2021): "A first look at online reputation on Airbnb, where every stay is above average," *Marketing Letters*, 32, 1–16.

Zhang, H., W. Gan, and B. Jiang (2014): "Machine learning and lexicon based methods for sentiment classification: A survey," in *2014 11th web information system and application conference*, IEEE, 262–265.

Zhang, S., D. Lee, P. Singh, and T. Mukhopadhyay (2021a): "EXPRESS: Demand Interactions in Sharing Economies: Evidence from a Natural Experiment Involving Airbnb and Uber/Lyft," *Journal of Marketing Research*, 00222437211062172.

Zhang, S., D. Lee, P. V. Singh, and K. Srinivasan (2021b): "What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features," *Management Science*.

Zhang, S., N. Mehta, P. V. Singh, and K. Srinivasan (2021c): "Can an AI Algorithm Mitigate Racial Economic Inequality? An Analysis in the Context of Airbnb," *An Analysis in the Context of Airbnb (January 21, 2021)*.

# Appendix

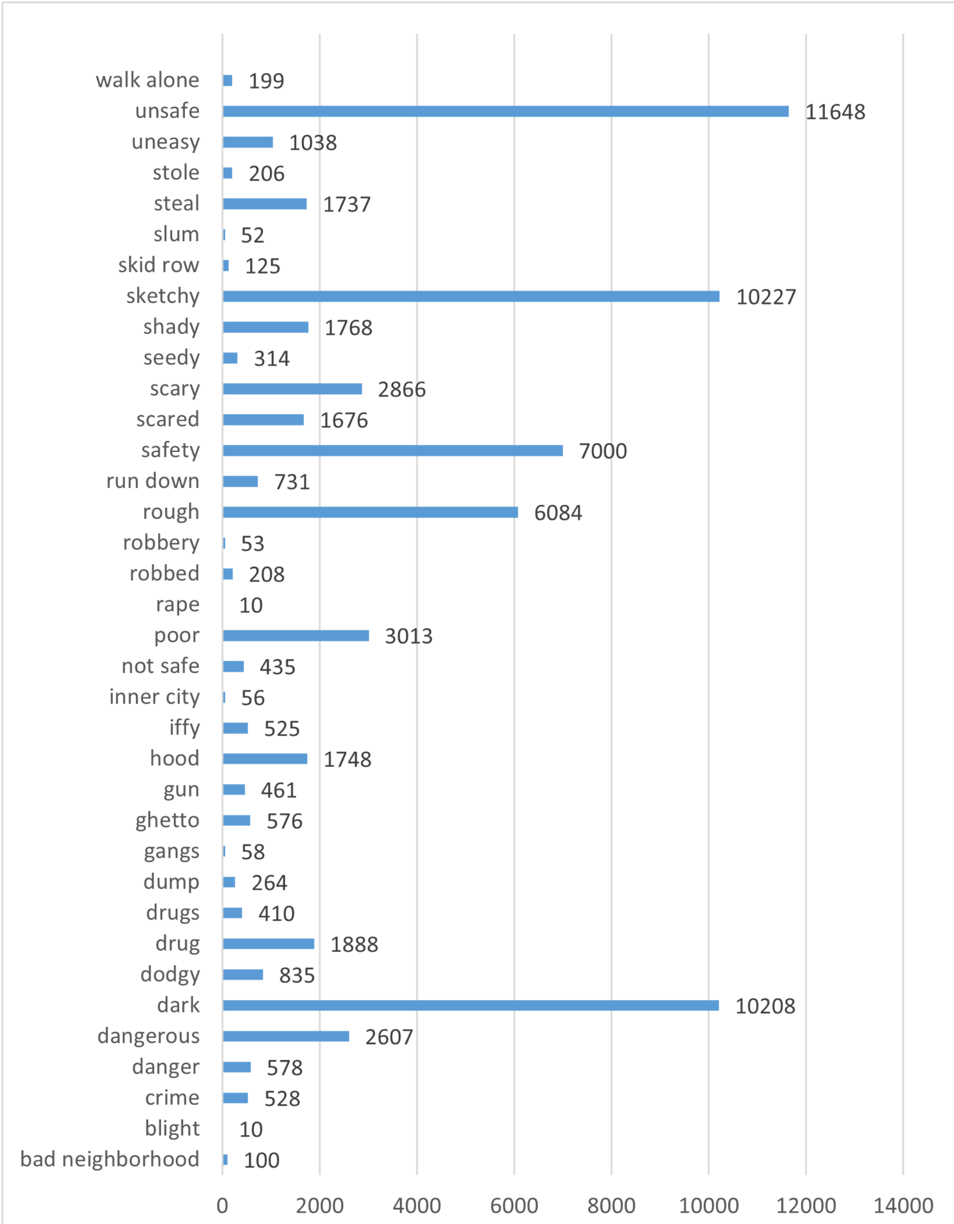


Figure 1: Distribution for keywords of vicinity safety review

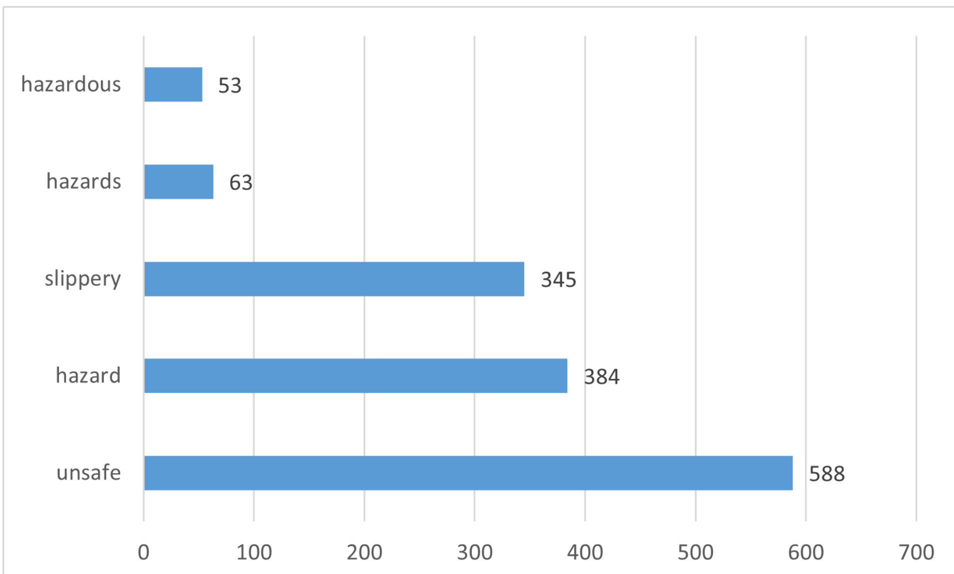


Figure 2: Distribution for keywords of listing safety review

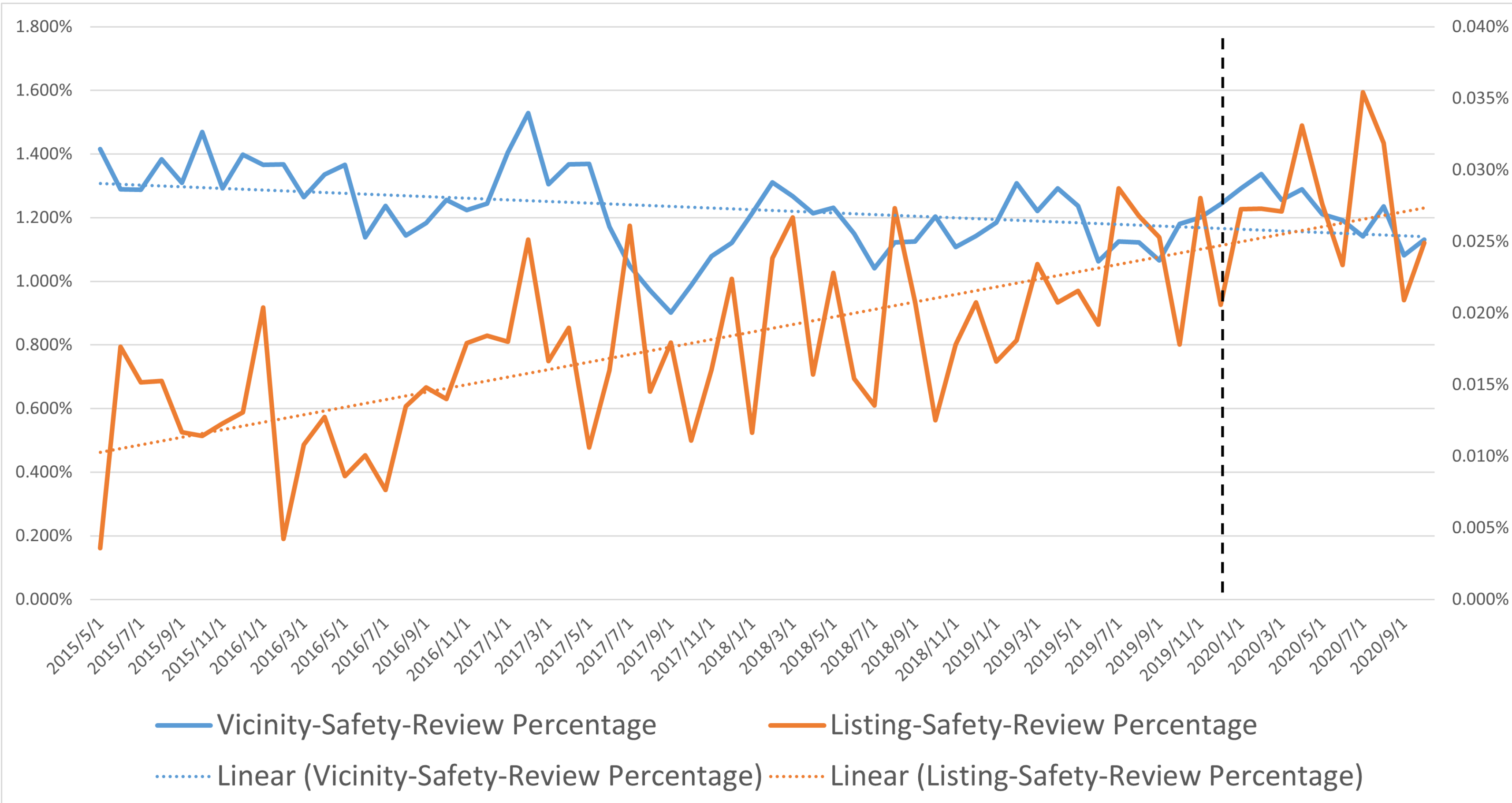


Figure 3: Percentage of safety review over time

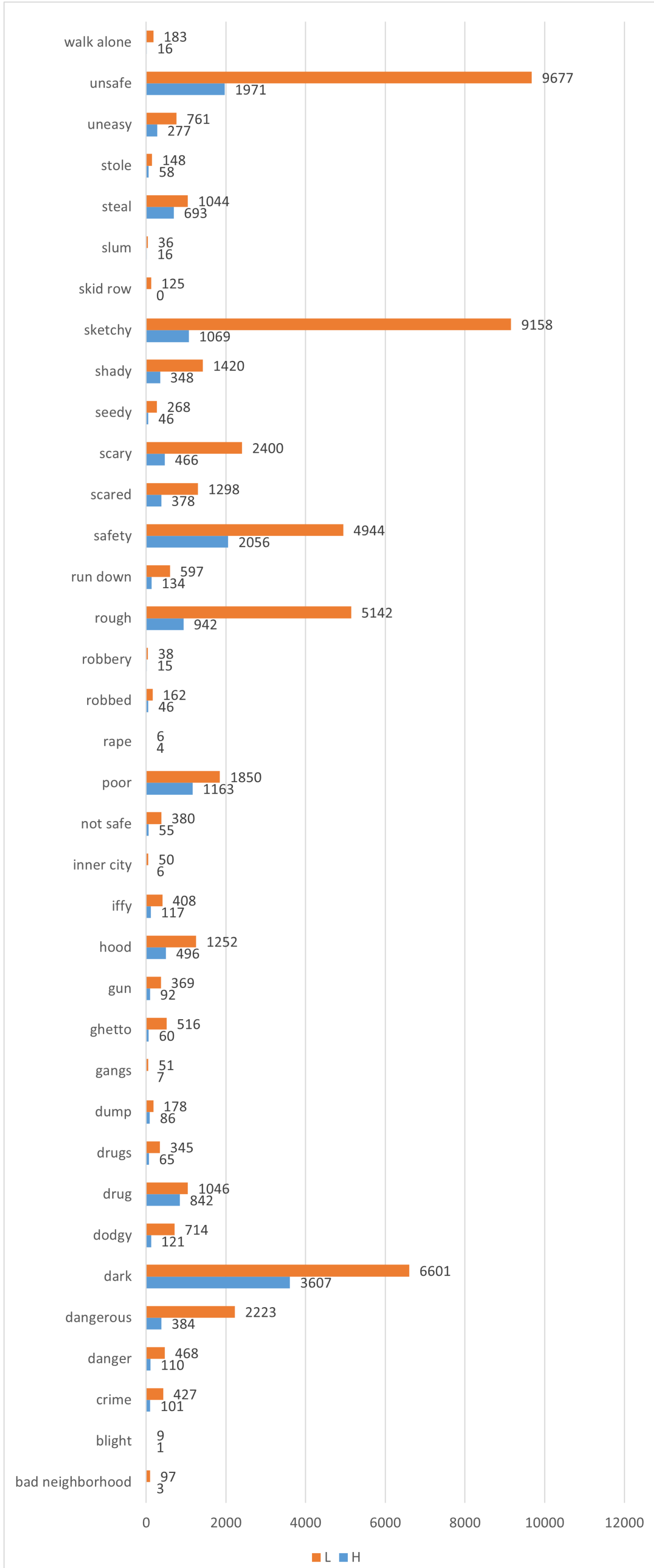


Figure 4: Distribution for keywords of vicinity safety review in H & L zip codes

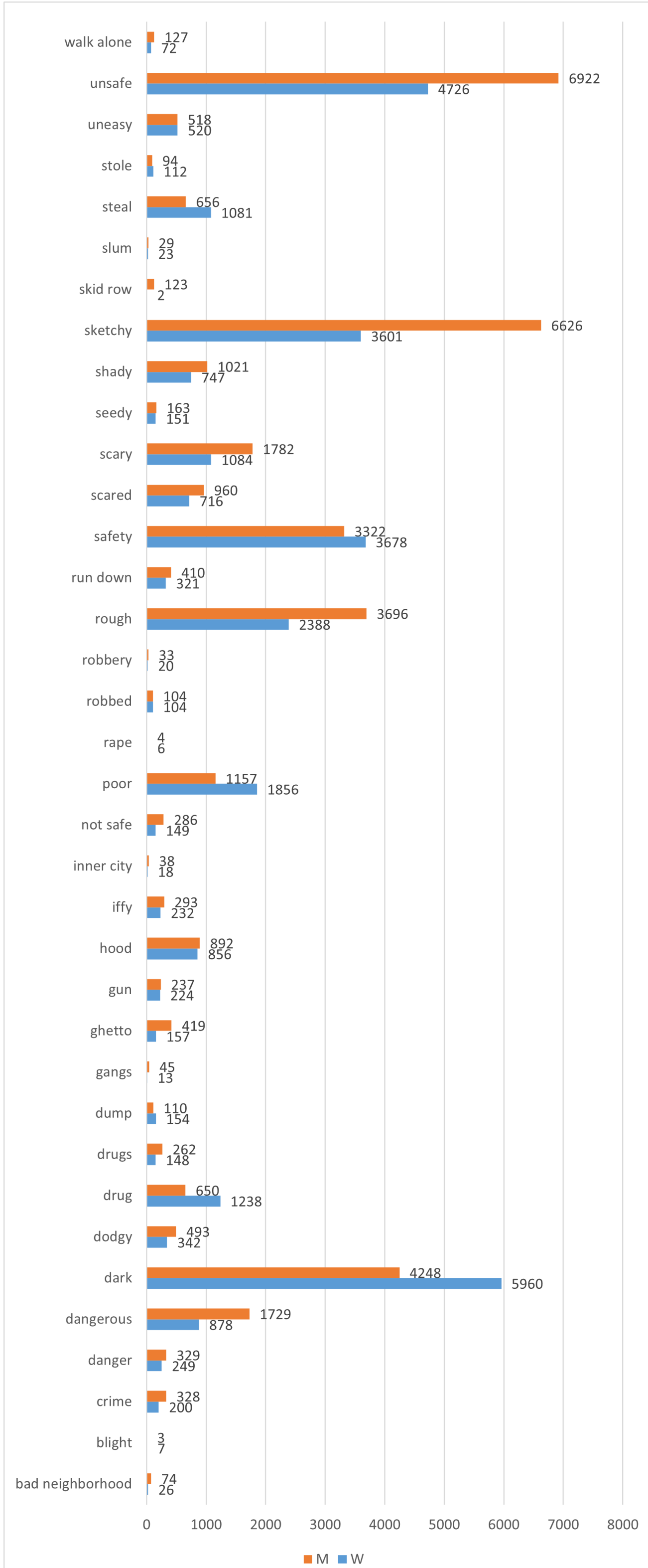


Figure 5: Distribution for keywords of vicinity safety review in W & M zip codes

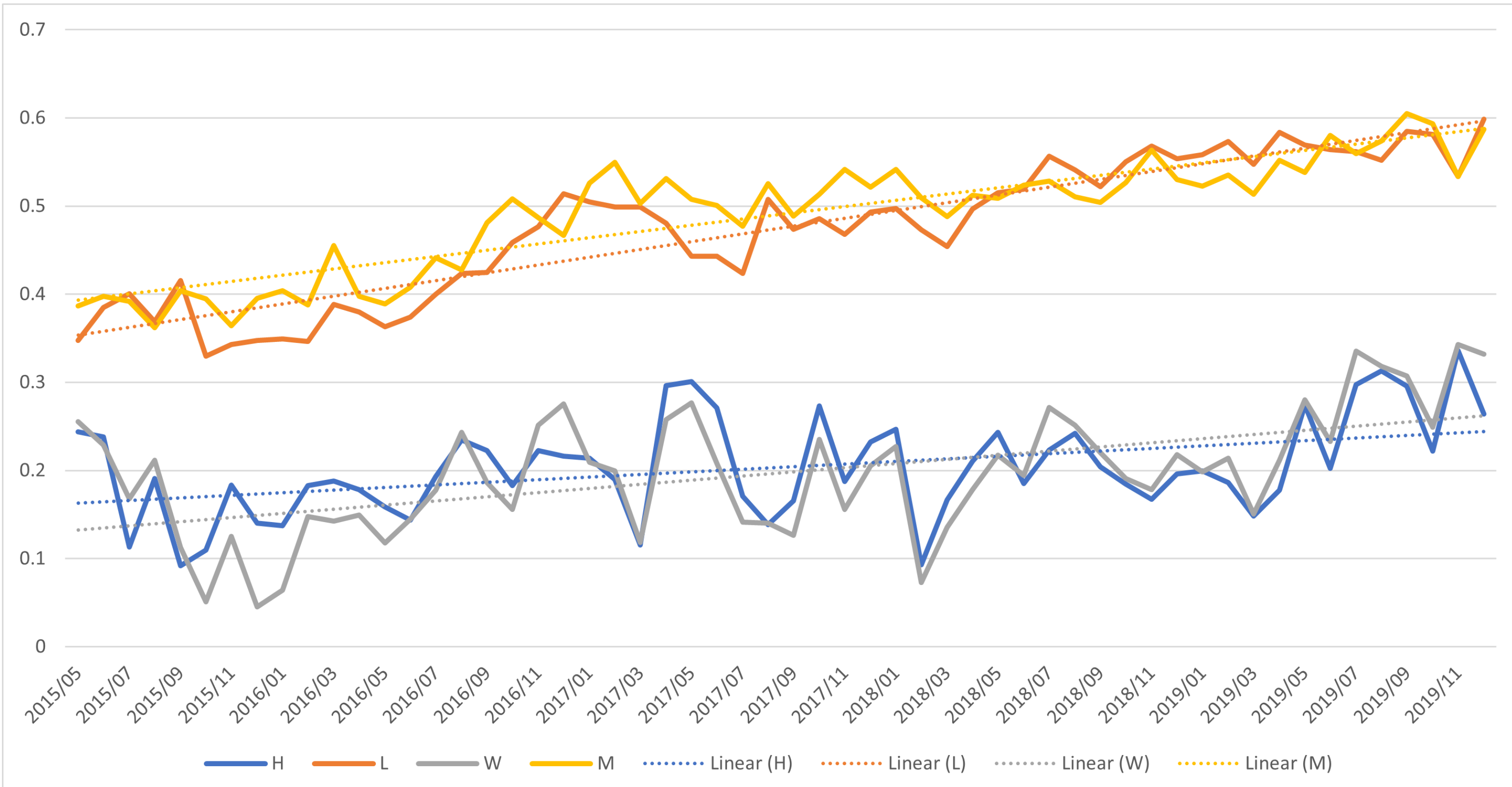


Figure 6: Correlation between the rank of normalized total crime size and the rank of normalized cumulative vicinity safety reviews across zip codes

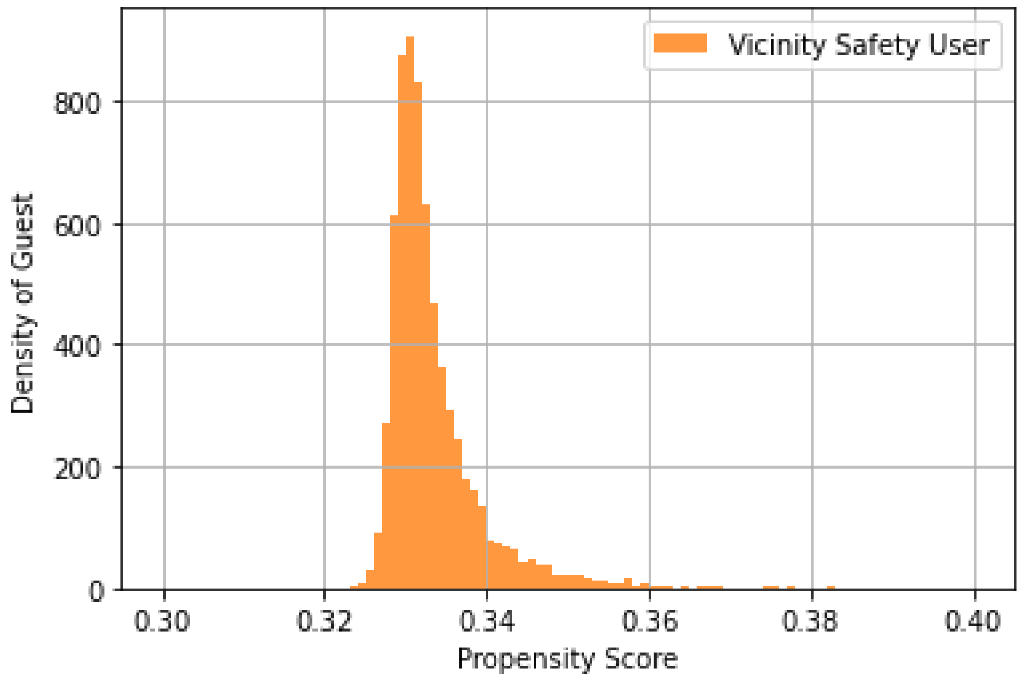


Figure 7: Distribution of Propensity Score for Control and Treatment Group

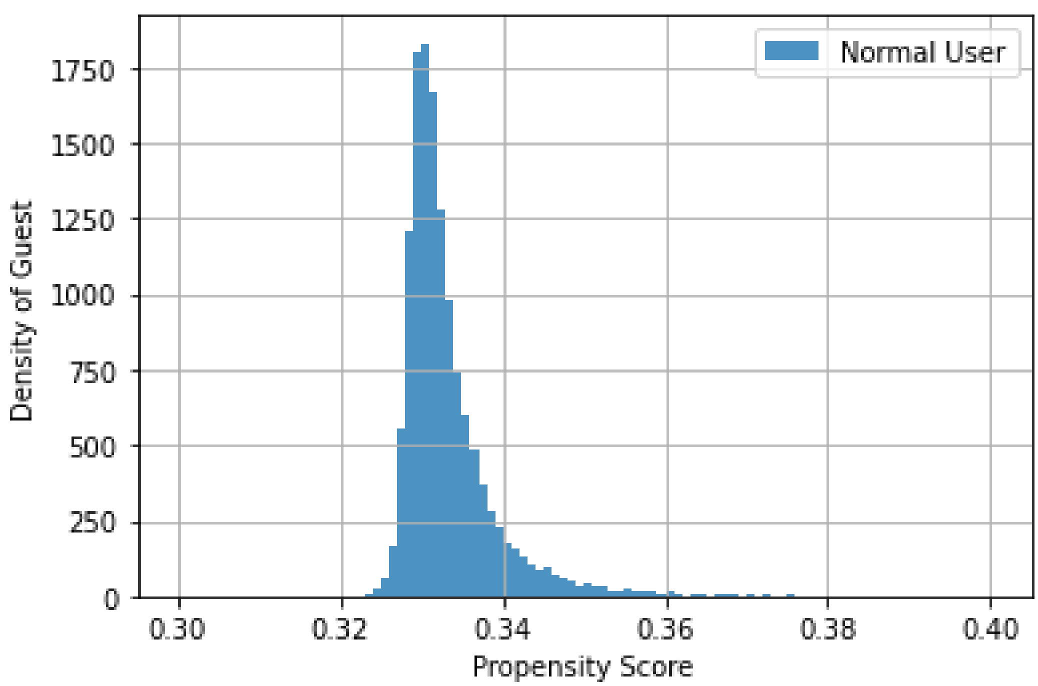


Figure 7: Distribution of Propensity Score for Control and Treatment Group (cont.)

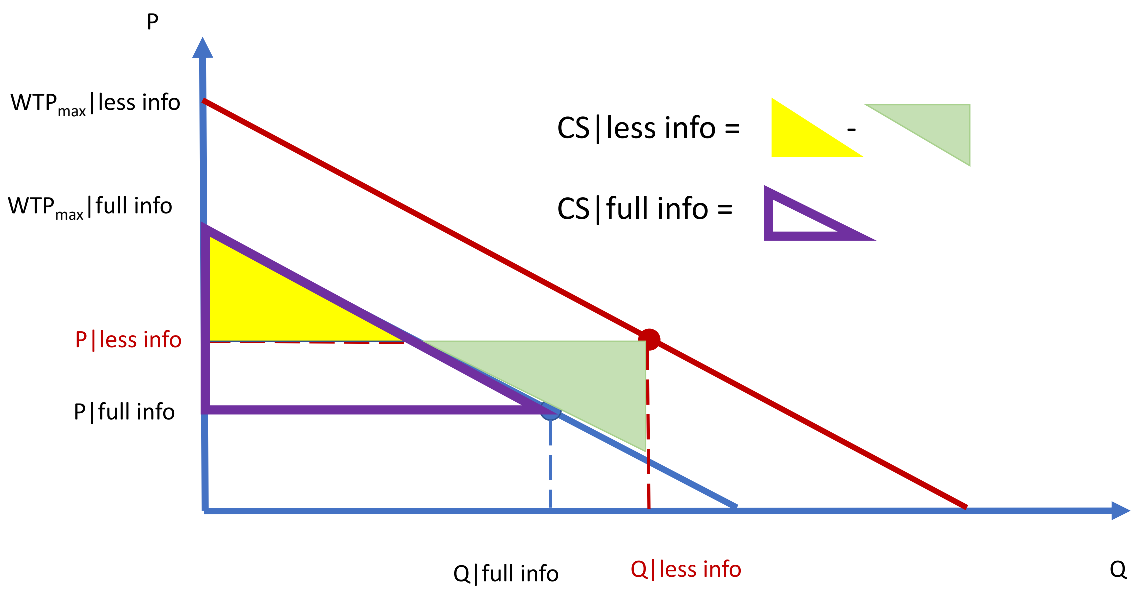


Figure 8: Consumer Surplus

| | |
|---|---|
| Vicinity safety keywords: | ‘crime’, ‘danger’, ‘dangerous’, ‘dodgy’, ‘gangs’, ‘ghetto’, ‘iffy’, ‘not safe’, ‘rough’, ‘safety’, ‘blight’, ‘scared’, ‘scary’, ‘seedy’, ‘shady’, ‘sketchy’, ‘uneasy’, ‘unsafe’, ‘walk alone’, ‘steal’, ‘stole’, ‘robbery’, ‘robbed’, ‘rape’, ‘gun’, ‘bad neighborhood’, ‘dark’, ‘drug’, ‘drugs’, ‘dump’, ‘inner city’, ‘hood’, ‘skid row’, ‘poor’, ‘slum’, ‘run down’ |
| Listing safety keywords: | ‘unsafe’, ‘hazard’, ‘hazards’, ‘hazardous’, ‘slippery’ |
| Vicinity location terms set: | ‘neighborhood’, ‘area’, ‘feel’, ‘felt’, ‘night’, ‘location’, ‘walking’, ‘people’, ‘seemed’, ‘outside’, ‘looked’, ‘looks’, ‘late’, ‘surrounding’, ‘located’, ‘neighbourhood’, ‘walked’, ‘areas’, ‘feeling’, ‘streets’, ‘street’ |

Table 1: Vicinity and listing safety review keywords

| ***Panel A*** | All Airbnb Listings | | | |
|---|---|---|---|---|
| VARIABLES | mean | p50 | sd | N |
| occupancyrate | 0.55 | 0.63 | 0.36 | 2,969,840 |
| occupancyrate_dummy | 0.85 | 1.00 | 0.36 | 2,969,840 |
| numberofreservations | 3.73 | 3.00 | 3.34 | 2,969,840 |
| reservationdays | 13.88 | 14.00 | 10.43 | 2,969,840 |
| adrusd | 163.13 | 124.67 | 171.44 | 2,969,840 |
| vicinity_sr_cumu | 0.3273 | 0.0000 | 1.0500 | 2,969,840 |
| listing_sr_cumu | 0.0054 | 0.0000 | 0.0755 | 2,969,840 |
| vicinity_sr_cumu_dummy | 0.1821 | 0.0000 | 0.3859 | 2,969,840 |
| listing_sr_cumu_dummy | 0.0052 | 0.0000 | 0.0722 | 2,969,840 |
| numberofreviews | 30.26 | 12.00 | 46.45 | 2,969,840 |
| ratingoverall | 9.26 | 9.60 | 1.29 | 2,969,840 |
| superhost_dummy | 0.23 | 0.00 | 0.42 | 2,969,840 |
| strict_cp | 0.44 | 0.00 | 0.50 | 2,969,840 |
| cross_listing_EH | 0.03 | 0.00 | 0.18 | 1,806,871 |
| cross_listing_PR | 0.00 | 0.00 | 0.00 | 1,052,322 |
| cross_listing_SR | 0.00 | 0.00 | 0.00 | 99,544 |
| cross_listing_HR | 0.00 | 0.00 | 0.00 | 11,103 |
| vicinity_sr_cumu_zip_pct | 0.02 | 0.01 | 0.02 | 2,969,840 |
| vicinity_listing_zip_pct | 0.16 | 0.14 | 0.09 | 2,969,840 |
| vicinity_listing_radius_pct | 0.16 | 0.14 | 0.13 | 2,969,840 |
| ave_review_word_count_cumu | 50.33 | 48.63 | 31.15 | 2,969,840 |
| listing_size_zipcode | 535.66 | 444.00 | 413.10 | 2,969,840 |
| crimewhole_cumu_zipcode | 19471.70 | 9741.00 | 34578.53 | 2,969,840 |
| median_income_zipcode | 57037.80 | 50943.00 | 26128.53 | 2,969,840 |
| population_zipcode | 48102.48 | 45284.00 | 24532.95 | 2,969,840 |
| white_percent_zipcode | 0.53 | 0.59 | 0.24 | 2,969,840 |

Table 2: Data summary statistics – listing characteristics at listing-month level

| ***Panel B*** | Vicinity Safety Airbnb Listings | | | | Normal Airbnb Listings | | | |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | mean | p50 | sd | N | mean | p50 | sd | N |
| occupancyrate | 0.67 | 0.77 | 0.31 | 540,750 | 0.53 | 0.59 | 0.37 | 2,429,090 |
| occupancyrate_dummy | 0.95 | 1.00 | 0.22 | 540,750 | 0.82 | 1.00 | 0.38 | 2,429,090 |
| numberofreservations | 5.29 | 5.00 | 3.55 | 540,750 | 3.39 | 2.00 | 3.19 | 2,429,090 |
| reservationdays | 17.94 | 20.00 | 9.46 | 540,750 | 12.98 | 12.00 | 10.42 | 2,429,090 |
| adrusd | 142.88 | 112.64 | 135.38 | 540,750 | 167.64 | 127.50 | 178.17 | 2,429,090 |
| vicinity_sr_cumu | 1.7976 | 1.0000 | 1.8471 | 540,750 | 0.0000 | 0.0000 | 0.0000 | 2,429,090 |
| listing_sr_cumu | 0.0174 | 0.0000 | 0.1368 | 540,750 | 0.0027 | 0.0000 | 0.0526 | 2,429,090 |
| vicinity_sr_cumu_dummy | 1.0000 | 1.0000 | 0.0000 | 540,750 | 0.0000 | 0.0000 | 0.0000 | 2,429,090 |
| listing_sr_cumu_dummy | 0.0166 | 0.0000 | 0.1278 | 540,750 | 0.0027 | 0.0000 | 0.0520 | 2,429,090 |
| numberofreviews | 74.17 | 54.00 | 69.44 | 540,750 | 20.49 | 8.00 | 32.24 | 2,429,090 |
| ratingoverall | 9.24 | 9.40 | 0.76 | 540,750 | 9.26 | 9.60 | 1.38 | 2,429,090 |
| superhost_dummy | 0.33 | 0.00 | 0.47 | 540,750 | 0.21 | 0.00 | 0.41 | 2,429,090 |
| strict_cp | 0.50 | 0.00 | 0.50 | 540,750 | 0.42 | 0.00 | 0.49 | 2,429,090 |
| cross_listing_EH | 0.04 | 0.00 | 0.19 | 337,376 | 0.03 | 0.00 | 0.18 | 1,469,495 |
| cross_listing_PR | 0.00 | 0.00 | 0.00 | 187,599 | 0.00 | 0.00 | 0.00 | 864,723 |
| cross_listing_SR | 0.00 | 0.00 | 0.00 | 14,253 | 0.00 | 0.00 | 0.00 | 85,291 |
| cross_listing_HR | 0.00 | 0.00 | 0.00 | 1,522 | 0.00 | 0.00 | 0.00 | 9,581 |
| vicinity_sr_cumu_zip_pct | 0.02 | 0.02 | 0.03 | 540,750 | 0.02 | 0.01 | 0.01 | 2,429,090 |
| vicinity_listing_zip_pct | 0.21 | 0.18 | 0.11 | 540,750 | 0.15 | 0.14 | 0.08 | 2,429,090 |
| vicinity_listing_radius_pct | 0.24 | 0.21 | 0.16 | 540,750 | 0.14 | 0.13 | 0.11 | 2,429,090 |
| ave_review_word_count_cumu | 59.08 | 55.73 | 23.31 | 540,750 | 48.38 | 46.38 | 32.32 | 2,429,090 |
| listing_size_zipcode | 543.81 | 475.00 | 388.72 | 540,750 | 533.84 | 438.00 | 418.32 | 2,429,090 |
| crimewhole_cumu_zipcode | 28641.91 | 12668.00 | 45076.94 | 540,750 | 17430.28 | 9188.00 | 31410.60 | 2,429,090 |
| median_income_zipcode | 49174.28 | 40835.00 | 23775.20 | 540,750 | 58788.33 | 52399.00 | 26306.17 | 2,429,090 |
| population_zipcode | 46853.10 | 41689.00 | 25065.57 | 540,750 | 48380.61 | 46025.00 | 24404.10 | 2,429,090 |
| white_percent_zipcode | 0.46 | 0.44 | 0.24 | 540,750 | 0.54 | 0.60 | 0.24 | 2,429,090 |

Table 2: Data summary statistics – listing characteristics at listing–month level (cont.)

| VARIABLES | (1) log occupancy rate | (2) log occupancy rate | (3) log adr | (4) log adr |
|---|---|---|---|---|
| lag_log_crimewhole_cumu_norm | 0.0663*** | 0.0652*** | -0.0680*** | -0.0668*** |
| | (0.00870) | (0.00870) | (0.0137) | (0.0137) |
| lag_vicinity_sr_cumu_dummy | -0.00966*** | -0.0102*** | -0.0140*** | -0.0134*** |
| | (0.000862) | (0.000870) | (0.00123) | (0.00125) |
| lag_listing_sr_cumu_dummy | -0.0261*** | -0.0262*** | -0.0175*** | -0.0174*** |
| | (0.00398) | (0.00397) | (0.00594) | (0.00594) |
| lag_log_number_of_reviews | -0.00294*** | -0.00298*** | 0.0135*** | 0.0135*** |
| | (0.000291) | (0.000292) | (0.000447) | (0.000447) |
| log_listing_size_zip | -0.0195*** | -0.0196*** | 0.0136*** | 0.0136*** |
| | (0.00189) | (0.00189) | (0.00292) | (0.00291) |
| log_rating_overall | 0.0206*** | 0.0207*** | -0.00415* | -0.00417* |
| | (0.00136) | (0.00136) | (0.00216) | (0.00216) |
| superhost_dummy | 0.0196*** | 0.0197*** | 0.00736*** | 0.00735*** |
| | (0.000595) | (0.000595) | (0.000845) | (0.000845) |
| cross_listing | 0.0306*** | 0.0306*** | -0.00499 | -0.00499 |
| | (0.00284) | (0.00284) | (0.00393) | (0.00393) |
| strict_cp | 0.0402*** | 0.0402*** | 0.00460*** | 0.00463*** |
| | (0.000640) | (0.000640) | (0.000833) | (0.000833) |
| lag_vicinity_listing_radius_pct | | 0.0151*** | | -0.0165*** |
| | | (0.00346) | | (0.00546) |
| Constant | 0.485*** | 0.483*** | 4.759*** | 4.762*** |
| | (0.0124) | (0.0124) | (0.0192) | (0.0192) |
| Observations | 2,764,911 | 2,764,911 | 2,764,911 | 2,764,911 |
| R-squared | 0.554 | 0.554 | 0.928 | 0.928 |
| Time*City FE | Yes | Yes | Yes | Yes |
| PropertyID FE | Yes | Yes | Yes | Yes |
| Listing_Type FE | Yes | Yes | Yes | Yes |
| Standard Error Clustered by Propertyid | Yes | Yes | Yes | Yes |
| Sample | whole | whole | whole | whole |

Robust standard errors in parentheses

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

Note: The dependent variable is ln (occupancy rate + 1) in columns 1 & 3 and ln (ADR) in columns 2 & 4. We use OLS for all the regressions in this chart.

Table 3: Listing level: Regressions with all Airbnb listings, using the main specification.

| VARIABLES | (1) occupancy dummy | (2) log occupancy rate | (3) log occupancy rate | (4) log occupancy rate | (5) log occupancy rate | (6) log occupancy rate | (7) log adr | (8) log adr | (9) log adr | (10) log adr | (11) log adr |
|---|---|---|---|---|---|---|---|---|---|---|---|
| lag_log_crimewhole_cumu_norm | 0.155*** | -0.0116 | 0.246*** | 0.00387 | 0.0654*** | 0.0643*** | -0.0190 | -0.0720*** | -0.0370** | -0.0668*** | -0.0667*** |
| | (0.0128) | (0.00749) | (0.0166) | (0.0110) | (0.00867) | (0.00869) | (0.0124) | (0.0262) | (0.0176) | (0.0137) | (0.0137) |
| lag_vicinity_sr_cumu_dummy | -0.00889*** | -0.00698*** | -0.00986*** | -0.00243** | -0.00866*** | -0.0102*** | -0.00837*** | -0.00648** | -0.00812*** | -0.0132*** | -0.0134*** |
| | (0.00108) | (0.000710) | (0.00239) | (0.000977) | (0.000870) | (0.000870) | (0.00112) | (0.00314) | (0.00135) | (0.00124) | (0.00125) |
| lag_listing_sr_cumu_dummy | -0.0179*** | -0.0203*** | -0.0449*** | -0.0161*** | -0.0250*** | -0.0262*** | -0.0159*** | 0.0143 | -0.0167*** | -0.0172*** | -0.0174*** |
| | (0.00406) | (0.00342) | (0.0130) | (0.00418) | (0.00399) | (0.00397) | (0.00576) | (0.0171) | (0.00644) | (0.00593) | (0.00594) |
| lag_log_number_of_reviews | -0.0104*** | -0.00117*** | -0.00708*** | -0.0184*** | -0.00583*** | -0.00304*** | 0.0155*** | 0.0103*** | 0.0163*** | 0.0132*** | 0.0135*** |
| | (0.000410) | (0.000238) | (0.000456) | (0.000687) | (0.000302) | (0.000292) | (0.000376) | (0.000694) | (0.00116) | (0.000472) | (0.000447) |
| log_listing_size_zip | -0.0263*** | -0.00793*** | -0.0225*** | -0.0146*** | -0.0193*** | -0.0199*** | 0.00631** | 0.0280*** | 0.00824** | 0.0137*** | 0.0137*** |
| | (0.00283) | (0.00161) | (0.00341) | (0.00236) | (0.00188) | (0.00189) | (0.00255) | (0.00591) | (0.00339) | (0.00291) | (0.00292) |
| log_rating_overall | 0.0360*** | 0.0117*** | 0.0194*** | 0.00699*** | 0.0175*** | 0.0207*** | -0.00445*** | -0.0136*** | 0.00349* | -0.00456** | -0.00417* |
| | (0.00262) | (0.00109) | (0.00174) | (0.00165) | (0.00135) | (0.00136) | (0.00162) | (0.00292) | (0.00211) | (0.00215) | (0.00216) |
| superhost_dummy | 0.0184*** | 0.0135*** | 0.0141*** | 0.0163*** | 0.0202*** | 0.0197*** | 0.0135*** | 0.00418** | 0.00626*** | 0.00742*** | 0.00735*** |
| | (0.000822) | (0.000485) | (0.00140) | (0.000701) | (0.000595) | (0.000595) | (0.000760) | (0.00192) | (0.000997) | (0.000845) | (0.000845) |
| cross_listing | 0.0223*** | 0.0231*** | 0.0427*** | 0.0205*** | 0.0304*** | 0.0299*** | -0.00244 | -0.0160** | 0.000213 | -0.00502 | -0.00487 |
| | (0.00374) | (0.00228) | (0.00537) | (0.00342) | (0.00283) | (0.00283) | (0.00369) | (0.00750) | (0.00495) | (0.00393) | (0.00393) |
| strict_cp | -0.0148*** | 0.0509*** | 0.0320*** | 0.0444*** | 0.0399*** | 0.0402*** | -0.00381*** | 0.00685*** | 0.000765 | 0.00459*** | 0.00463*** |
| | (0.000829) | (0.000569) | (0.00103) | (0.000828) | (0.000639) | (0.000640) | (0.000712) | (0.00143) | (0.000999) | (0.000832) | (0.000833) |
| lag_vicinity_listing_radius_pct | 0.0229*** | 0.00530* | 0.0320*** | 0.0114*** | 0.0159*** | 0.0151*** | -0.0195*** | 0.00599 | -0.0144** | -0.0164*** | -0.0165*** |
| | (0.00499) | (0.00297) | (0.00739) | (0.00400) | (0.00345) | (0.00345) | (0.00475) | (0.0125) | (0.00574) | (0.00546) | (0.00546) |
| lag_log_ave_r_wordcount_cumu | | | | | 0.00623*** | | | | | 0.000786** | |
| | | | | | (0.000222) | | | | | (0.000310) | |
| host_property_sum | | | | | | 0.000159*** | | | | | -3.07e-05 |
| | | | | | | (2.64e-05) | | | | | (4.13e-05) |
| Constant | 0.949*** | 0.514*** | 0.408*** | 0.618*** | 0.474*** | 0.484*** | 4.772*** | 4.726*** | 4.714*** | 4.761*** | 4.762*** |
| | (0.0187) | (0.0106) | (0.0219) | (0.0158) | (0.0124) | (0.0124) | (0.0165) | (0.0378) | (0.0233) | (0.0192) | (0.0192) |
| Observations | 2,764,911 | 2,378,552 | 1,309,704 | 1,455,207 | 2,764,911 | 2,764,911 | 2,378,552 | 1,309,704 | 1,455,207 | 2,764,911 | 2,764,911 |
| R-squared | 0.407 | 0.498 | 0.567 | 0.524 | 0.554 | 0.554 | 0.943 | 0.932 | 0.939 | 0.928 | 0.928 |
| Time*City FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| PropertyID FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Listing_Type FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard Error Clustered by Propertyid | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample | whole | occ>0 | # reviews <=12 | # reviews >12 | whole | whole | occ>0 | # reviews <=12 | # reviews >12 | whole | whole |

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Note: The dependent variable is dummy variable of occupancy rate where it is 1 when occupancy rate is greater than 0 in columns 1. Dependent variable is ln (occupancy rate + 1) in columns 2 to 6 and ln (ADR) in columns 7 to 11. We use OLS for all the regressions in this chart.

Table 4: Listing Level: Regressions for robustness check

| ***Panel A*** | H | | | | L | | | | W | | | | M | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | mean | p50 | sd | N | mean | p50 | sd | N | mean | p50 | sd | N | mean | p50 | sd | N |
| occupancyrate | 0.55 | 0.63 | 0.36 | 1,531,920 | 0.56 | 0.64 | 0.36 | 1,437,920 | 0.55 | 0.63 | 0.36 | 1,772,330 | 0.56 | 0.64 | 0.36 | 1,197,510 |
| occupancyrate_dummy | 0.84 | 1.00 | 0.37 | 1,531,920 | 0.86 | 1.00 | 0.35 | 1,437,920 | 0.84 | 1.00 | 0.36 | 1,772,330 | 0.85 | 1.00 | 0.36 | 1,197,510 |
| numberofreservations | 3.62 | 3.00 | 3.30 | 1,531,920 | 3.86 | 3.00 | 3.38 | 1,437,920 | 3.67 | 3.00 | 3.29 | 1,772,330 | 3.83 | 3.00 | 3.41 | 1,197,510 |
| reservationdays | 13.60 | 13.00 | 10.41 | 1,531,920 | 14.18 | 14.00 | 10.44 | 1,437,920 | 13.72 | 13.00 | 10.38 | 1,772,330 | 14.12 | 14.00 | 10.49 | 1,197,510 |
| adrusd | 188.80 | 147.32 | 186.52 | 1,531,920 | 135.78 | 102.60 | 148.98 | 1,437,920 | 186.75 | 145.93 | 182.81 | 1,772,330 | 128.17 | 97.00 | 146.25 | 1,197,510 |
| vicinity_sr_cumu | 0.1933 | 0.0000 | 0.5909 | 1,531,920 | 0.4701 | 0.0000 | 1.3659 | 1,437,920 | 0.2512 | 0.0000 | 0.8667 | 1,772,330 | 0.4399 | 0.0000 | 1.2654 | 1,197,510 |
| listing_sr_cumu | 0.0049 | 0.0000 | 0.0716 | 1,531,920 | 0.0059 | 0.0000 | 0.0794 | 1,437,920 | 0.0055 | 0.0000 | 0.0767 | 1,772,330 | 0.0053 | 0.0000 | 0.0737 | 1,197,510 |
| vicinity_sr_cumu_dummy | 0.1380 | 0.0000 | 0.3449 | 1,531,920 | 0.2291 | 0.0000 | 0.4202 | 1,437,920 | 0.1534 | 0.0000 | 0.3604 | 1,772,330 | 0.2245 | 0.0000 | 0.4173 | 1,197,510 |
| listing_sr_cumu_dummy | 0.0048 | 0.0000 | 0.0691 | 1,531,920 | 0.0057 | 0.0000 | 0.0753 | 1,437,920 | 0.0053 | 0.0000 | 0.0725 | 1,772,330 | 0.0052 | 0.0000 | 0.0718 | 1,197,510 |
| numberofreviews | 29.69 | 12.00 | 45.99 | 1,531,920 | 30.87 | 13.00 | 46.93 | 1,437,920 | 30.17 | 12.00 | 46.39 | 1,772,330 | 30.41 | 12.00 | 46.53 | 1,197,510 |
| ratingoverall | 9.31 | 9.60 | 1.27 | 1,531,920 | 9.20 | 9.50 | 1.31 | 1,437,920 | 9.30 | 9.60 | 1.25 | 1,772,330 | 9.19 | 9.50 | 1.33 | 1,197,510 |
| superhost_dummy | 0.25 | 0.00 | 0.43 | 1,531,920 | 0.22 | 0.00 | 0.42 | 1,437,920 | 0.24 | 0.00 | 0.43 | 1,772,330 | 0.22 | 0.00 | 0.42 | 1,197,510 |
| strict_cp | 0.44 | 0.00 | 0.50 | 1,531,920 | 0.44 | 0.00 | 0.50 | 1,437,920 | 0.45 | 0.00 | 0.50 | 1,772,330 | 0.42 | 0.00 | 0.49 | 1,197,510 |
| cross_listing_EH | 0.03 | 0.00 | 0.18 | 1,054,672 | 0.03 | 0.00 | 0.18 | 752,199 | 0.04 | 0.00 | 0.18 | 1,215,902 | 0.03 | 0.00 | 0.18 | 590,969 |
| cross_listing_PR | 0.00 | 0.00 | 0.00 | 434,272 | 0.00 | 0.00 | 0.00 | 618,050 | 0.00 | 0.00 | 0.00 | 506,184 | 0.00 | 0.00 | 0.00 | 546,138 |
| cross_listing_SR | 0.00 | 0.00 | 0.00 | 36,315 | 0.00 | 0.00 | 0.00 | 63,229 | 0.00 | 0.00 | 0.00 | 41,987 | 0.00 | 0.00 | 0.00 | 57,557 |
| cross_listing_HR | 0.00 | 0.00 | 0.00 | 6,661 | 0.00 | 0.00 | 0.00 | 4,442 | 0.00 | 0.00 | 0.00 | 8,257 | 0.00 | 0.00 | 0.00 | 2,846 |
| vicinity_sr_cumu_zip_pct | 0.01 | 0.01 | 0.00 | 1,531,920 | 0.02 | 0.02 | 0.02 | 1,437,920 | 0.01 | 0.01 | 0.01 | 1,772,330 | 0.02 | 0.02 | 0.02 | 1,197,510 |
| vicinity_listing_zip_pct | 0.12 | 0.12 | 0.06 | 1,531,920 | 0.20 | 0.19 | 0.10 | 1,437,920 | 0.14 | 0.12 | 0.08 | 1,772,330 | 0.20 | 0.19 | 0.09 | 1,197,510 |
| vicinity_listing_radius_pct | 0.12 | 0.11 | 0.09 | 1,531,920 | 0.21 | 0.19 | 0.15 | 1,437,920 | 0.14 | 0.12 | 0.11 | 1,772,330 | 0.20 | 0.19 | 0.14 | 1,197,510 |
| ave_review_word_count_cumu | 50.29 | 48.85 | 31.15 | 1,531,920 | 50.37 | 48.39 | 31.15 | 1,437,920 | 50.81 | 49.39 | 31.09 | 1,772,330 | 49.61 | 47.43 | 31.22 | 1,197,510 |
| listing_size_zipcode | 498.03 | 461.00 | 303.87 | 1,531,920 | 575.75 | 420.00 | 500.97 | 1,437,920 | 604.52 | 508.00 | 454.23 | 1,772,330 | 433.74 | 369.00 | 316.96 | 1,197,510 |
| crimewhole_cumu_zipcode | 14796.26 | 7818.00 | 26515.26 | 1,531,920 | 24452.79 | 12257.00 | 40894.93 | 1,437,920 | 20987.80 | 8440.00 | 40901.05 | 1,772,330 | 17227.86 | 11664.00 | 21930.49 | 1,197,510 |
| median_income_zipcode | 75765.48 | 71278.00 | 22733.87 | 1,531,920 | 37085.85 | 35112.00 | 9363.81 | 1,437,920 | 69637.12 | 68346.00 | 25614.23 | 1,772,330 | 38390.65 | 37116.00 | 11807.96 | 1,197,510 |
| population_zipcode | 43482.99 | 41453.00 | 20828.46 | 1,531,920 | 53023.95 | 50162.00 | 27091.25 | 1,437,920 | 44643.85 | 38633.00 | 24135.43 | 1,772,330 | 53221.29 | 54440.00 | 24219.68 | 1,197,510 |
| white_percent_zipcode | 0.68 | 0.72 | 0.16 | 1,531,920 | 0.37 | 0.32 | 0.20 | 1,437,920 | 0.69 | 0.72 | 0.13 | 1,772,330 | 0.28 | 0.29 | 0.13 | 1,197,510 |

Table 5: Data summary statistics – listing characteristics at listing–month level for subgroups

| ***Panel B*** | EH | | | | PR | | | | SR | | | | HR | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | mean | p50 | sd | N | mean | p50 | sd | N | mean | p50 | sd | N | mean | p50 | sd | N |
| occupancyrate | 0.57 | 0.65 | 0.35 | 1,806,871 | 0.54 | 0.61 | 0.38 | 1,052,322 | 0.43 | 0.39 | 0.37 | 99,544 | 0.46 | 0.44 | 0.34 | 11,103 |
| occupancyrate_dummy | 0.86 | 1.00 | 0.34 | 1,806,871 | 0.83 | 1.00 | 0.38 | 1,052,322 | 0.77 | 1.00 | 0.42 | 99,544 | 0.85 | 1.00 | 0.35 | 11,103 |
| numberofreservations | 3.82 | 3.00 | 3.24 | 1,806,871 | 3.61 | 3.00 | 3.47 | 1,052,322 | 3.24 | 2.00 | 3.40 | 99,544 | 5.63 | 5.00 | 4.78 | 11,103 |
| reservationdays | 14.15 | 14.00 | 10.21 | 1,806,871 | 13.71 | 13.00 | 10.78 | 1,052,322 | 11.11 | 9.00 | 10.17 | 99,544 | 12.49 | 11.00 | 9.74 | 11,103 |
| adrusd | 211.51 | 169.55 | 175.22 | 1,806,871 | 90.05 | 75.14 | 134.25 | 1,052,322 | 55.27 | 38.09 | 106.58 | 99,544 | 182.48 | 146.36 | 252.32 | 11,103 |
| vicinity_sr_cumu | 0.3400 | 0.0000 | 1.0764 | 1,806,871 | 0.3162 | 0.0000 | 1.0321 | 1,052,322 | 0.2219 | 0.0000 | 0.7075 | 99,544 | 0.2576 | 0.0000 | 0.8549 | 11,103 |
| listing_sr_cumu | 0.0068 | 0.0000 | 0.0849 | 1,806,871 | 0.0033 | 0.0000 | 0.0583 | 1,052,322 | 0.0027 | 0.0000 | 0.0516 | 99,544 | 0.0040 | 0.0000 | 0.0628 | 11,103 |
| vicinity_sr_cumu_dummy | 0.1867 | 0.0000 | 0.3897 | 1,806,871 | 0.1783 | 0.0000 | 0.3827 | 1,052,322 | 0.1432 | 0.0000 | 0.3503 | 99,544 | 0.1371 | 0.0000 | 0.3439 | 11,103 |
| listing_sr_cumu_dummy | 0.0065 | 0.0000 | 0.0805 | 1,806,871 | 0.0033 | 0.0000 | 0.0572 | 1,052,322 | 0.0027 | 0.0000 | 0.0516 | 99,544 | 0.0040 | 0.0000 | 0.0628 | 11,103 |
| numberofreviews | 30.77 | 13.00 | 45.44 | 1,806,871 | 30.80 | 11.00 | 49.37 | 1,052,322 | 16.41 | 6.00 | 27.54 | 99,544 | 20.47 | 7.00 | 38.09 | 11,103 |
| ratingoverall | 9.32 | 9.60 | 1.17 | 1,806,871 | 9.19 | 9.60 | 1.41 | 1,052,322 | 8.85 | 9.30 | 1.71 | 99,544 | 9.06 | 9.40 | 1.23 | 11,103 |
| superhost_dummy | 0.25 | 0.00 | 0.43 | 1,806,871 | 0.22 | 0.00 | 0.41 | 1,052,322 | 0.11 | 0.00 | 0.32 | 99,544 | 0.13 | 0.00 | 0.33 | 11,103 |
| strict_cp | 0.47 | 0.00 | 0.50 | 1,806,871 | 0.38 | 0.00 | 0.49 | 1,052,322 | 0.46 | 0.00 | 0.50 | 99,544 | 0.38 | 0.00 | 0.49 | 11,103 |
| cross_listing | 0.03 | 0.00 | 0.18 | 1,806,871 | 0.00 | 0.00 | 0.00 | 1,052,322 | 0.00 | 0.00 | 0.00 | 99,544 | 0.00 | 0.00 | 0.00 | 11,103 |
| vicinity_sr_cumu_zip_pct | 0.02 | 0.01 | 0.01 | 1,806,871 | 0.02 | 0.01 | 0.02 | 1,052,322 | 0.02 | 0.02 | 0.05 | 99,544 | 0.02 | 0.02 | 0.01 | 11,103 |
| vicinity_listing_zip_pct | 0.16 | 0.14 | 0.10 | 1,806,871 | 0.16 | 0.15 | 0.09 | 1,052,322 | 0.18 | 0.16 | 0.10 | 99,544 | 0.19 | 0.16 | 0.11 | 11,103 |
| vicinity_listing_radius_pct | 0.20 | 0.18 | 0.14 | 1,806,871 | 0.19 | 0.19 | 0.14 | 1,052,322 | 0.20 | 0.18 | 0.16 | 99,544 | 0.22 | 0.21 | 0.12 | 11,103 |
| ave_review_word_count_cumu | 51.51 | 49.82 | 30.90 | 1,806,871 | 49.45 | 47.71 | 31.27 | 1,052,322 | 39.80 | 36.50 | 31.72 | 99,544 | 36.49 | 33.00 | 30.24 | 11,103 |
| listing_size_zipcode | 556.94 | 477.00 | 403.01 | 1,806,871 | 509.47 | 397.00 | 432.21 | 1,052,322 | 428.25 | 332.00 | 362.68 | 99,544 | 517.08 | 464.00 | 297.47 | 11,103 |
| crimewhole_cumu_zipcode | 22203.52 | 9640.00 | 39698.02 | 1,806,871 | 15274.91 | 9882.00 | 24197.73 | 1,052,322 | 12944.97 | 9260.00 | 17165.66 | 99,544 | 31183.66 | 14047.00 | 42631.04 | 11,103 |
| median_income_zipcode | 59512.44 | 54023.00 | 27300.25 | 1,806,871 | 53563.27 | 47050.00 | 23579.18 | 1,052,322 | 48431.39 | 40539.00 | 23559.74 | 99,544 | 60792.90 | 59125.00 | 28826.05 | 11,103 |
| population_zipcode | 44382.22 | 38752.00 | 23835.24 | 1,806,871 | 54247.03 | 54440.00 | 24496.82 | 1,052,322 | 52078.15 | 48852.00 | 23565.00 | 99,544 | 35513.54 | 30648.00 | 23414.28 | 11,103 |
| white_percent_zipcode | 0.56 | 0.61 | 0.23 | 1,806,871 | 0.48 | 0.46 | 0.24 | 1,052,322 | 0.45 | 0.44 | 0.23 | 99,544 | 0.55 | 0.61 | 0.21 | 11,103 |

Table 5: Data summary statistics – listing characteristics at listing–month level for subgroups(cont.)

| VARIABLES | (1) log occupancy rate | (2) log occupancy rate | (3) log occupancy rate | (4) log occupancy rate | (5) log adr | (6) log adr | (7) log adr | (8) log adr |
|---|---|---|---|---|---|---|---|---|
| lag_log_crimewhole_cumu_norm | 0.0452*** | 0.170*** | 0.0410*** | 0.158*** | -0.0609*** | -0.0746*** | -0.0636*** | -0.0793*** |
| | (0.0117) | (0.0144) | (0.0101) | (0.0174) | (0.0191) | (0.0221) | (0.0159) | (0.0270) |
| lag_vicinity_sr_cumu_dummy | -0.0111*** | -0.00883*** | -0.0110*** | -0.00889*** | -0.0127*** | -0.0135*** | -0.0138*** | -0.0125*** |
| | (0.00133) | (0.00116) | (0.00119) | (0.00128) | (0.00187) | (0.00168) | (0.00170) | (0.00184) |
| lag_listing_sr_cumu_dummy | -0.0253*** | -0.0262*** | -0.0280*** | -0.0220*** | -0.0203*** | -0.0144 | -0.0184** | -0.0156 |
| | (0.00603) | (0.00519) | (0.00530) | (0.00591) | (0.00769) | (0.00899) | (0.00749) | (0.00962) |
| lag_log_number_of_reviews | -0.00221*** | -0.00355*** | -0.00277*** | -0.00317*** | 0.0151*** | 0.0119*** | 0.0150*** | 0.0114*** |
| | (0.000411) | (0.000414) | (0.000382) | (0.000451) | (0.000615) | (0.000651) | (0.000575) | (0.000710) |
| log_listing_size_zip | -0.0242*** | -0.0129*** | -0.0187*** | -0.0123*** | 0.0236*** | 0.00447 | 0.0213*** | 0.00582 |
| | (0.00292) | (0.00270) | (0.00286) | (0.00274) | (0.00425) | (0.00426) | (0.00415) | (0.00440) |
| log_rating_overall | 0.0201*** | 0.0215*** | 0.0188*** | 0.0236*** | -0.00227 | -0.00586* | -0.00384 | -0.00447 |
| | (0.00190) | (0.00195) | (0.00173) | (0.00219) | (0.00286) | (0.00327) | (0.00274) | (0.00351) |
| superhost_dummy | 0.0195*** | 0.0194*** | 0.0194*** | 0.0199*** | 0.00540*** | 0.00950*** | 0.00609*** | 0.00915*** |
| | (0.000828) | (0.000854) | (0.000784) | (0.000911) | (0.00115) | (0.00124) | (0.00112) | (0.00128) |
| cross_listing | 0.0355*** | 0.0227*** | 0.0307*** | 0.0290*** | -0.00574 | -0.00339 | -0.00696 | -0.000729 |
| | (0.00369) | (0.00440) | (0.00348) | (0.00490) | (0.00487) | (0.00655) | (0.00459) | (0.00748) |
| strict_cp | 0.0390*** | 0.0412*** | 0.0403*** | 0.0398*** | 0.00468*** | 0.00485*** | 0.00341*** | 0.00683*** |
| | (0.000896) | (0.000912) | (0.000831) | (0.00100) | (0.00111) | (0.00124) | (0.00105) | (0.00136) |
| lag_vicinity_listing_radius_pct | 0.0341*** | 0.00587 | 0.0316*** | 0.00775* | -0.0197** | -0.0136** | -0.0120 | -0.0163** |
| | (0.00602) | (0.00424) | (0.00586) | (0.00430) | (0.00890) | (0.00687) | (0.00895) | (0.00690) |
| Constant | 0.516*** | 0.383*** | 0.489*** | 0.399*** | 4.845*** | 4.666*** | 4.860*** | 4.594*** |
| | (0.0187) | (0.0186) | (0.0186) | (0.0185) | (0.0271) | (0.0294) | (0.0269) | (0.0297) |
| Observations | 1,426,766 | 1,338,145 | 1,651,141 | 1,113,770 | 1,426,766 | 1,338,145 | 1,651,141 | 1,113,770 |
| R-squared | 0.546 | 0.564 | 0.545 | 0.568 | 0.922 | 0.925 | 0.920 | 0.926 |
| Time*City FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| PropertyID FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Listing_Type FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard Error Clustered by Propertyid | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample | H | L | W | M | H | L | W | M |

Robust standard errors in parentheses

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

Note: The dependent variable is ln (occupancy rate + 1) in columns 1 to 4 and ln (ADR) in columns 5 & 8. We use OLS for all the regressions in this chart.

Table 6: Listing level: Regression with all Airbnb listings (4 areas)

| VARIABLES | (1) log occupancy rate | (2) log occupancy rate | (3) log occupancy rate | (4) log occupancy rate | (5) log adr | (6) log adr | (7) log adr | (8) log adr |
|---|---|---|---|---|---|---|---|---|
| lag_log_crimewhole_cumu_norm | 0.0404*** | 0.103*** | 0.266*** | -0.656*** | -0.0910*** | -0.0115 | -0.388** | 0.731*** |
| | (0.0107) | (0.0155) | (0.0679) | (0.144) | (0.0166) | (0.0235) | (0.158) | (0.270) |
| lag_vicinity_sr_cumu_dummy | -0.0101*** | -0.0104*** | -0.0132** | -0.00895 | -0.0131*** | -0.0139*** | -0.0322*** | -0.00114 |
| | (0.00107) | (0.00155) | (0.00550) | (0.0143) | (0.00155) | (0.00213) | (0.00877) | (0.0227) |
| lag_listing_sr_cumu_dummy | -0.0244*** | -0.0335*** | -0.0358 | -0.00638 | -0.0201*** | -0.0196* | 0.0255 | 0.0312 |
| | (0.00442) | (0.00917) | (0.0354) | (0.0292) | (0.00700) | (0.0107) | (0.0599) | (0.0786) |
| lag_log_number_of_reviews | -0.00128*** | -0.00360*** | -0.0225*** | 0.0106** | 0.0157*** | 0.0106*** | 0.00429 | 0.0307*** |
| | (0.000365) | (0.000508) | (0.00168) | (0.00536) | (0.000565) | (0.000751) | (0.00289) | (0.00922) |
| log_listing_size_zip | -0.0140*** | -0.0255*** | -0.0141 | 0.0625** | 0.0175*** | 0.0205*** | -0.0313 | -0.0322 |
| | (0.00256) | (0.00295) | (0.00943) | (0.0304) | (0.00384) | (0.00471) | (0.0191) | (0.0515) |
| log_rating_overall | 0.0190*** | 0.0225*** | 0.0302*** | -0.0206 | -0.000197 | -0.00871*** | -0.0189 | -0.00710 |
| | (0.00183) | (0.00207) | (0.00918) | (0.0577) | (0.00272) | (0.00333) | (0.0204) | (0.0384) |
| superhost_dummy | 0.0166*** | 0.0236*** | 0.0349*** | 0.0195 | 0.00928*** | 0.00339** | 0.0122 | -0.0276 |
| | (0.000729) | (0.00103) | (0.00504) | (0.0156) | (0.00106) | (0.00138) | (0.00919) | (0.0260) |
| cross_listing | 0.0316*** | | | | -0.00807** | | | |
| | (0.00282) | | | | (0.00394) | | | |
| strict_cp | 0.0388*** | 0.0380*** | 0.0795*** | 0.00819 | 0.00348*** | 0.00828*** | 0.00965* | -0.0405** |
| | (0.000788) | (0.00114) | (0.00387) | (0.0105) | (0.00102) | (0.00149) | (0.00523) | (0.0186) |
| lag_vicinity_listing_radius_pct | 0.0220*** | 0.0136** | -0.0285 | -0.104 | -0.00982 | -0.0203** | 0.0389 | 0.270* |
| | (0.00462) | (0.00545) | (0.0175) | (0.0897) | (0.00726) | (0.00840) | (0.0289) | (0.144) |
| Constant | 0.461*** | 0.508*** | 0.315*** | 0.606*** | 5.075*** | 4.238*** | 4.037*** | 4.285*** |
| | (0.0169) | (0.0191) | (0.0650) | (0.232) | (0.0250) | (0.0312) | (0.138) | (0.362) |
| Observations | 1,685,717 | 978,240 | 90,552 | 10,402 | 1,685,717 | 978,240 | 90,552 | 10,402 |
| R-squared | 0.535 | 0.575 | 0.591 | 0.624 | 0.888 | 0.857 | 0.889 | 0.910 |
| Time*City FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| PropertyID FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Listing_Type FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard Error Clustered by Propertyid | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample | EH | PR | SR | HR | EH | PR | SR | HR |

Robust standard errors in parentheses

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

Note: The dependent variable is ln (occupancy rate + 1) in columns 1 to 4 and ln (ADR) in columns 5 & 8. We use OLS for all the regressions in this chart.

Table 7: Listing Level: Regression with the all Airbnb listings (4 areas)

| | (1) |
|---|---|
| VARIABLES | VS_user |
| | |
| reservation_bef | 0.00775 |
| | (0.00491) |
| ave_crimewhole_cumu_norm_bef | 0.00704 |
| | (0.00758) |
| ave_vicinity_sr_cumu_bef | 0.00785 |
| | (0.0113) |
| ave_vicinity_sr_zip_cumu_pct_bef | 1.003 |
| | (1.846) |
| ave_vicinity_listing_zip_pct_bef | -0.162 |
| | (0.283) |
| Constant | -0.733*** |
| | (0.0421) |
| | |
| Observations | 20,316 |
| Standard Error | Robust |
| Sample | Normal & VS User |
| Log pseudolikelihood: | -12929 |
| Chi-squared | 4.005 |
| Prob Wald: | 0.549 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The dependent variable is dummy variable of vicinity safety user where it is 1 when the user had ever posted vicinity safe review(s). We use Logistic Regression in this chart.

Table 8: Propensity Score estimation

| ***Panel A*** | Treatment (Vicinity Safety) Users | | | | Control (Normal) Users | | | |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | mean | p50 | sd | N | mean | p50 | sd | N |
| reservation_before | 4.37 | 3.00 | 3.12 | 6,772 | 4.30 | 3.00 | 2.94 | 13,544 |
| ave_crimewhole_cumu_norm_before | 0.98 | 0.25 | 2.18 | 6,772 | 0.95 | 0.23 | 2.13 | 13,544 |
| ave_vicinity_sr_cumu_before | 1.07 | 0.67 | 1.48 | 6,772 | 1.06 | 0.67 | 1.46 | 13,544 |
| ave_vicinity_sr_zip_cumu_pct_before | 0.02 | 0.02 | 0.01 | 6,772 | 0.02 | 0.02 | 0.01 | 13,544 |
| ave_vicinity_listing_zip_pct_before | 0.17 | 0.16 | 0.08 | 6,772 | 0.17 | 0.16 | 0.08 | 13,544 |
| propensity_score | 0.33 | 0.33 | 0.01 | 6,772 | 0.33 | 0.33 | 0.01 | 13,544 |

| ***Panel B*** | Vicinity Safety Users' All Booking | | | | Normal Users' All Booking | | | |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | mean | p50 | sd | N | mean | p50 | sd | N |
| reservation_total | 4.66 | 3.00 | 4.73 | 6,772 | 4.93 | 4.00 | 7.85 | 13,544 |
| crimewhole_cumu_norm | 0.79 | 0.20 | 2.48 | 31,540 | 0.82 | 0.20 | 2.57 | 66,764 |
| vicinity_sr_cumu | 0.86 | 0.00 | 1.89 | 31,540 | 0.85 | 0.00 | 1.87 | 66,764 |
| vicinity_sr_cumu_zip_pct | 0.02 | 0.01 | 0.02 | 31,540 | 0.02 | 0.01 | 0.02 | 66,764 |
| vicinity_listing_zip_pct | 0.17 | 0.15 | 0.10 | 31,540 | 0.17 | 0.15 | 0.11 | 66,764 |

Table 9: Data summary statistics – users characteristics

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| VARIABLES | reservation_monthly | crimewhole_cumu_norm | vicinity_sr_cumu | vicinity_sr_cumu_zip_pct | vicinity_listing_zip_pct | vicinity_sr_dummy |
| treat | -0.100*** | 0.0254*** | 0.0227*** | 5.04e-05 | 0.000112 | 0.123*** |
| | (0.0114) | (0.00404) | (0.00335) | (0.000104) | (0.000418) | (0.0143) |
| post | -0.721*** | 0.634*** | -0.00550 | 0.00211*** | 0.0451*** | 0.192*** |
| | (0.0379) | (0.0385) | (0.0323) | (0.000528) | (0.00197) | (0.0261) |
| interaction | -0.372*** | -0.318*** | -0.245*** | -0.00156* | -0.0134*** | -0.214*** |
| | (0.0299) | (0.0381) | (0.0505) | (0.000836) | (0.00433) | (0.0388) |
| Constant | | 0.662*** | | 0.0185*** | 0.161*** | |
| | | (0.00905) | | (8.36e-05) | (0.000319) | |
| Observations | 1,066,559 | 98,304 | 98,304 | 98,304 | 98,304 | 98,304 |
| R-squared | | 0.430 | | 0.360 | 0.444 | |
| Pair ID FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard Error Clustered by Pair ID | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample | Normal & VS User | Normal & VS User | Normal & VS User | Normal & VS User | Normal & VS User | Normal & VS User |
| Log pseudolikelihood: | -392954 | | -107550 | | | -64361 |
| Chi-squared | 1348 | | 79.33 | | | 111.5 |
| Prob Wald: | 0 | | 0 | | | 0 |

Robust standard errors in parentheses

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

Note: The dependent variable is monthly reservation numbers in columns 1, normalized total crime in column 2, number of total vicinity safety reviews in column 3, percentage of vicinity safety reviews in the same-zip listings, and percentage of vicinity listings in the same-zip. We use Poisson regression for column 1 & 3, and OLS for others in this chart.

Table 10: User level: DID for all users

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| VARIABLES | h_zip | l_zip | h_zip | l_zip | w_zip | m_zip | w_zip | m_zip |
| treat | 0.910*** | -0.910*** | -0.183*** | 0.183*** | 0.771*** | -0.771*** | -0.444*** | 0.444*** |
| | (0.0570) | (0.0570) | (0.0231) | (0.0231) | (0.0428) | (0.0428) | (0.0315) | (0.0315) |
| post | -0.0972* | 0.0972* | 0.451*** | -0.451*** | -0.143*** | 0.143*** | 0.370*** | -0.370*** |
| | (0.0567) | (0.0567) | (0.0263) | (0.0263) | (0.0464) | (0.0464) | (0.0411) | (0.0411) |
| interaction | -0.303*** | 0.303*** | 0.0783* | -0.0783* | -0.315*** | 0.315*** | 0.197** | -0.197** |
| | (0.114) | (0.114) | (0.0430) | (0.0430) | (0.114) | (0.114) | (0.0852) | (0.0852) |
| Observations | 42,670 | 42,670 | 55,634 | 55,634 | 42,817 | 42,817 | 55,487 | 55,487 |
| Pair ID FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Time FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Standard Error Clustered by Pair ID | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Sample | first_vsr_h_zip | first_vsr_h_zip | first_vsr_l_zip | first_vsr_l_zip | first_vsr_w_zip | first_vsr_w_zip | first_vsr_m_zip | first_vsr_m_zip |
| Log pseudolikelihood: | -14807 | -14807 | -35713 | -35713 | -17946 | -17946 | -21873 | -21873 |
| Pseudo R2: | 0.0316 | 0.0316 | 0.0103 | 0.0103 | 0.0228 | 0.0228 | 0.0134 | 0.0134 |

Robust standard errors in parentheses

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

Table 11: User level: DID for 4 groups of users based on their first safety review booking's zip code

Note: The dependent variable is a dummy variable to identify whether the booked listing is in a certain area. We use Logistic regression in this chart.

| Area | # of Listings | # of Reservations | Avg. Occupancy of all Listings (in Days) | Avg. ADR of all Listings ($) | Avg. days per Reservation | Avg % of VSL in the 0.3 radius area | GBV changes if no VSRs are available ($) | GBV changes if all Guests are VS Guests ($) |
|---|---|---|---|---|---|---|---|---|
| | | | | ***Panel A: Vicinity Safety Listings*** | | | | |
| H | 211,337 | 1,112,423 | 17.90 | 165.93 | 3.40 | 0.17 | 13,646,515.78 | -56,484,594.66 |
| | | | | | | | 2.17% | -9.00% |
| L | 329,413 | 1,747,654 | 17.97 | 128.09 | 3.39 | 0.29 | 18,867,399.91 | -66,958,823.46 |
| | | | | | | | 2.49% | -8.83% |
| W | 271,889 | 1,446,951 | 17.82 | 166.85 | 3.35 | 0.21 | 17,044,568.72 | -70,428,060.43 |
| | | | | | | | 2.11% | -8.71% |
| M | 268,861 | 1,413,126 | 18.06 | 118.63 | 3.44 | 0.28 | 13,867,450.17 | -52,351,167.85 |
| | | | | | | | 2.41% | -9.09% |
| | | | | ***Panel B: Normal Listings*** | | | | |
| H | 1,320,583 | 4,425,449 | 12.92 | 192.46 | 3.85 | 0.11 | -4,912,395.49 | -259,094,918.66 |
| | | | | | | | -0.15% | -7.89% |
| L | 1,108,507 | 3,803,658 | 13.06 | 138.06 | 3.80 | 0.18 | 2,812,600.04 | -149,554,127.29 |
| | | | | | | | 0.14% | -7.48% |
| W | 1,500,441 | 5,060,079 | 12.98 | 190.35 | 3.85 | 0.12 | -8,647,281.53 | -280,593,250.17 |
| | | | | | | | -0.23% | -7.57% |
| M | 928,649 | 3,169,028 | 12.98 | 130.93 | 3.80 | 0.17 | 2,365,911.72 | -124,997,390.54 |
| | | | | | | | 0.15% | -7.92% |
| | | | | ***Panel C: All Listings*** | | | | |
| | 2,969,840 | 11,089,184 | 13.88 | 163.13 | 3.72 | 0.16 | 30,838,382.31 | -538,866,569.50 |
| | | | | | | | 0.46% | -8.01% |

Table 12: GBV of the platform using back of envelope calculation

| VARIABLES | (1)<br>log_adr | (2)<br>utility | (3)<br>utility | (4)<br>utility |
|---|---|---|---|---|
| log_ave_adr_pr_zip | 0.0137*** | | | |
| | (0.00330) | | | |
| lag_log_crimewhole_cumu_norm | -0.0747*** | 0.494*** | 0.419*** | -0.238** |
| | (0.0214) | (0.0855) | (0.0854) | (0.102) |
| lag_vicinity_sr_cumu_dummy | -0.0102*** | -0.0586*** | -0.0699*** | -0.169*** |
| | (0.00159) | (0.00663) | (0.00650) | (0.0108) |
| lag_listing_sr_cumu_dummy | -0.0260*** | -0.0513* | -0.0802*** | -0.333*** |
| | (0.00753) | (0.0267) | (0.0268) | (0.0347) |
| lag_log_number_of_reviews | 0.0179*** | -0.0198*** | 0.000133 | 0.174*** |
| | (0.000586) | (0.00222) | (0.00220) | (0.0152) |
| log_rating_overall | 0.0193** | 0.313*** | 0.335*** | 0.523*** |
| | (0.00818) | (0.0336) | (0.0340) | (0.0373) |
| superhost_dummy | 0.0158*** | 0.0567*** | 0.0742*** | 0.227*** |
| | (0.00106) | (0.00405) | (0.00397) | (0.0138) |
| cross_listing | -0.00999** | 0.0614*** | 0.0506*** | -0.0431*** |
| | (0.00402) | (0.0145) | (0.0141) | (0.0166) |
| strict_cp | -0.00818*** | 0.211*** | 0.202*** | 0.122*** |
| | (0.00100) | (0.00409) | (0.00398) | (0.00795) |
| lag_vicinity_listing_radius_pct | -0.0113 | 0.0616* | 0.0478 | -0.0727** |
| | (0.00704) | (0.0341) | (0.0336) | (0.0356) |
| log_adr | | | -1.113*** | |
| | | | (0.00874) | |
| log_adr_iv_fitted | | | | -10.84*** |
| | | | | (0.840) |
| Constant | 5.053*** | -3.996*** | 1.695*** | 51.40*** |
| | (0.0271) | (0.0939) | (0.105) | (4.296) |
| | | | | |
| Observations | 1,008,480 | 1,008,480 | 1,008,480 | 1,008,480 |
| R-squared | 0.912 | 0.768 | 0.782 | 0.768 |
| Time*City FE | Yes | Yes | Yes | Yes |
| PropertyID FE | Yes | Yes | Yes | Yes |
| Listing_Type FE | Yes | Yes | Yes | Yes |
| Standard Error Clustered by Propertyid | Yes | Yes | Yes | Yes |
| Sample | post2017 EH | post2017 EH | post2017 EH | post2017 EH |
| F statistic | 327 | | | |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: We use OLS regression in this chart.

Table 13: Utility estimation

| | Airbnb VS Listing | Airbnb Normal Listing | VRBO Listing | All Listing |
|---|---|---|---|---|
| CS (Status quo) | 475,092.49 | 1,677,042.27 | 1,564,977.07 | 3,717,111.83 |
| CS (High Alert without P change) | 315,121.54 | 1,783,911.73 | 1,625,980.37 | 3,725,013.64 |
| CS (High Alert with P change) | 347,192.59 | 1,763,336.93 | 1,614,370.95 | 3,724,900.46 |
| CS (No Disclosure without P change) | 538,972.69 | 1,634,747.08 | 1,540,180.28 | 3,713,900.05 |
| CS (No Disclosure with P change) | 493,344.16 | 1,663,631.88 | 1,557,196.68 | 3,714,172.71 |
| | | | | |
| CS (High Alert without P change) - CS (Status quo) | -159,970.95 | 106,869.46 | 61,003.30 | 7,901.81 |
| | -33.6715% | 6.3725% | 3.8980% | 0.2126% |
| CS (High Alert with P change) - CS (Status quo) | -127,899.90 | 86,294.66 | 49,393.88 | 7,788.63 |
| | -26.9211% | 5.1456% | 3.1562% | 0.2095% |
| CS (No Disclosure without P change) - CS (Status quo) | 63,880.19 | -42,295.19 | -24,796.78 | -3,211.78 |
| | 13.4458% | -2.5220% | -1.5845% | -0.0864% |
| CS (No Disclosure with P change) - CS (Status quo) | 18,251.67 | -13,410.40 | -7,780.39 | -2,939.12 |
| | 3.8417% | -0.7996% | -0.4972% | -0.0791% |
| | | | | |
| Market share (Status quo) | 0.1410 | 0.4763 | 0.3901 | 1.0000 |
| Market share (High Alert without P change) | 0.0934 | 0.5064 | 0.4053 | 1.0000 |
| Market share (High Alert with P change) | 0.1025 | 0.5006 | 0.4024 | 1.0000 |
| Market share (No Disclosure without P change) | 0.1600 | 0.4643 | 0.3839 | 1.0000 |
| Market share (No Disclosure with P change) | 0.1470 | 0.4725 | 0.3882 | 1.0000 |

Table 14: Consumer Surplus estimation

# A Mathematics Proof

## A.1 Proof on how to do reduced-form back of envelope calculation for the "high-alert" counterfactual

Denote:

$X^H_{vsl}$ = # bookings of vicinity safety listings in high-income neighborhoods, under the status quo

$X^L_{vsl}$ = # bookings of vicinity safety listings in low-income neighborhoods, under the status quo

$X^H_{nl}$ = # bookings of normal listings in high-income neighborhoods, under the status quo

$X^L_{nl}$ = # bookings of normal listings in low-income neighborhoods, under the status quo

$Y^H_{vsl}$ = # bookings of vicinity safety listings in high-income neighborhoods, under "high-alert"

$Y^L_{vsl}$ = # bookings of vicinity safety listings in low-income neighborhoods, under "high-alert"

$Y^H_{nl}$ = # bookings of normal listings in high-income neighborhoods, under "high-alert"

$Y^L_{nl}$ = # bookings of normal listings in low-income neighborhoods, under "high-alert"

We observe $\{X^H_{vsl}, X^L_{vsl}, X^H_{nl}, X^L_{nl}\}$ and reduced-form impacts of a user submitting a safety review out of self-experience. Our goal is to solve for $\{Y^H_{vsl}, Y^L_{vsl}, Y^H_{nl}, Y^L_{nl}\}$.

Self-experience of vicinity safety makes a guest $\beta_1$ (percent) more likely to make any bookings on Airbnb according to the Column 1 of Table 10. The coefficient on $safetyuser \times post = -0.372$ in a Poisson regression, implying that safety review reduces monthly reservations by $\beta_1 = exp(-0.372) - 1 = -0.3106$. In our notation, we have function 4:

$$\begin{aligned} & (Y^H_{vsl} + Y^L_{vsl} + Y^H_{nl} + Y^L_{nl}) \\ = \quad & \underbrace{(X^H_{vsl} + X^L_{vsl}) \cdot (1 + \beta_1) + X^H_{nl} + X^L_{nl}}_{\text{item A}} \end{aligned} \tag{4}$$

Self-experience of vicinity safety makes a guest $\beta_2$ (percent) more likely to make a booking in an $H$ neighborhood, conditional on she makes any booking on Airbnb. Note that the safety experience, if it occurs, has a 40% chance to occur in an $H$ neighborhood and a 60% chance to occur in an $L$ neighborhood.

According to Table 11 column 1, the coefficient of $safetyuser \times post = -0.303$ for the probit of booking in $H$ if $VS$ in $H$, which implies that having a safety review in $H$ will change the probability of the booking in H by $\beta_{2|VS\ in\ H} = -0.0679$. According to Table 11 Column 3, the coefficient of $safetyuser \times post = +0.0783$ for the probit of booking in $H$ if $VS$ in $L$, which implies that having a safety review in $L$ will change the probability of the booking in $L$ by $\beta_{2|VS\ in\ L} = +0.0191$. And denote the probability of vicinity safety listings in H(L) area as $Pr_{V|H} = 0.1380$ ($Pr_{V|L} = 0.2291$ ) In our notation, this means:

$$\begin{aligned} & prob(booking\ in\ H|anybooking) \\ = & \frac{Y^H_{vsl}+Y^H_{nl}}{Y^H_{vsl}+Y^L_{vsl}+Y^H_{nl}+Y^L_{nl}} \\ = & \underbrace{\frac{X^H_{vsl}+X^H_{nl}}{X^H_{vsl}+X^L_{vsl}+X^H_{nl}+X^L_{nl}}}_{\text{item B}} \cdot (0.4 \cdot (1 + Pr_{V|H} \cdot \beta_{2|VS\ in\ H}) + 0.6 \cdot (1 + Pr_{V|L} \cdot \beta_{2|VS\ in\ L})) \end{aligned} \tag{5}$$

Self-experience of vicinity safety makes a guest $\beta_2$ (percent) more likely to make a booking in a VSL, conditional on she makes any booking on Airbnb. According to Table 10 column 6, the coefficient of $safetyuser \times post = -0.214$ for the probit of booking in any $VSL$, which implies that having a safety review will change the probability of the booking in any $VSL$ by $\beta_3 = -0.0529$ and the probability of vicinity safety listings is $Pr_V = 0.1821$.

In our notation, this means:

$$\begin{aligned} & prob(booking\ in\ VS|anybooking) \\ = & \frac{Y^H_{vsl}+Y^L_{vsl}}{Y^H_{vsl}+Y^L_{vsl}+Y^H_{nl}+Y^L_{nl}} \\ = & \underbrace{\frac{X^H_{vsl}+X^L_{vsl}}{X^H_{vsl}+X^L_{vsl}+X^H_{nl}+X^L_{nl}}}_{\text{item C}} \cdot (1 + Pr_V \cdot \beta_3) \end{aligned} \tag{6}$$

So far, we have four unknowns and three equations, so we need an extra equation, which implies that the fraction of $VSL$ in each neighborhood does not change. Denote this fraction as $\alpha^H_{vsl}$, this amounts to:

$$\frac{Y^H_{vsl}}{Y^H_{vsl}+Y^H_{nl}} = \frac{X^H_{vsl}}{X^H_{vsl}+X^H_{nl}} \cdot (1+Pr_V \cdot \beta_3) = \alpha^H_{vsl} \cdot (1+Pr_V \cdot \beta_3) \tag{7}$$

Combining Equations 4, 5, 6 and 7 and solving for $\{Y^H_{vsl}, Y^L_{vsl}, Y^H_{nl}, Y^L_{nl}\}$, we have:

$$\frac{Y^H_{vsl}+Y^H_{nl}}{Y^H_{vsl}+Y^L_{vsl}+Y^H_{nl}+Y^L_{nl}} = \frac{Y^H_{vsl}+Y^H_{nl}}{A} = B\cdot(0.4\cdot(1+Pr_{V|H}\cdot\beta_{2|VS\ in\ H})+0.6\cdot(1+Pr_{V|L}\cdot\beta_{2|VS\ in\ L})) \tag{8}$$

$$Y^H_{vsl}+Y^H_{nl} = A \cdot B \cdot (0.4 \cdot (1+Pr_{V|H} \cdot \beta_{2|VS\ in\ H}) + 0.6 \cdot (1+Pr_{V|L} \cdot \beta_{2|VS\ in\ L})) \tag{9}$$

$$\frac{Y^H_{vsl}+Y^L_{vsl}}{Y^H_{vsl}+Y^L_{vsl}+Y^H_{nl}+Y^L_{nl}} = \frac{Y^H_{vsl}+Y^L_{vsl}}{A} = C \cdot (1+Pr_V \cdot \beta_3) \tag{10}$$

$$Y^H_{vsl}+Y^L_{vsl} = A \cdot C \cdot (1+Pr_V \cdot \beta_3) \tag{11}$$

$$\frac{Y^H_{vsl}}{Y^H_{vsl}+Y^H_{nl}} = \frac{Y^H_{vsl}}{A \cdot B \cdot (0.4 \cdot (1+Pr_{V|H} \cdot \beta_{2|VS\ in\ H}) + 0.6 \cdot (1+Pr_{V|L} \cdot \beta_{2|VS\ in\ L}))} = \alpha^H_{vsl}\cdot(1+Pr_V\cdot\beta_3) \tag{12}$$

So the solutions are:

$$\begin{aligned}
Y^H_{vsl} &= \alpha^H_{vsl} \cdot (1+Pr_V \cdot \beta_3) \cdot A \cdot B \cdot (0.4 \cdot (1+Pr_{V|H} \cdot \beta_{2|VS\ in\ H}) + 0.6 \cdot (1+Pr_{V|L} \cdot \beta_{2|VS\ in\ L})) \\
Y^H_{nl} &= A \cdot B \cdot (0.4 \cdot (1+Pr_{V|H} \cdot \beta_{2|VS\ in\ H}) + 0.6 \cdot (1+Pr_{V|L} \cdot \beta_{2|VS\ in\ L})) - Y^H_{vsl} \\
Y^L_{vsl} &= A \cdot C \cdot (1+Pr_V \cdot \beta_3) - Y^H_{vsl} \\
Y^L_{nl} &= A - Y^H_{vsl} - Y^H_{nl} - Y^L_{vsl}
\end{aligned} \tag{13}$$

Similar calculations are performed for W/M areas.

## A.2 Extend the DID results of VS users to the consumer surplus counterfactual based on utility

The coefficient on $safetyreview \times post = -0.372$ in a Poisson regression according to the Table 10 Column 1, which is a 31.065% decrease. Given the average number of reservations per month for a single VS user in our sample is 0.1086 and review rate is 53%, the reservation that a VS user book in Airbnb is 0.03374/0.53 less than a normal user after she has reported a VS issue in her first VSR. A $VS$ user is less likely to book Airbnb reservations than a normal user after she has reported a $VS$ issue in her first $VSR$ is described by:

$$\begin{aligned} &[\#Airbnbbooking_{VS\ user,aft} - \#Airbnbbooking_{VS\ user,bef}] \\ &-[\#Airbnbbooking_{NM\ user,aft} - \#Airbnbbooking_{NM\ user,bef}] = -0.03374/0.53 \end{aligned} \tag{14}$$

Assuming $VS$ and normal users have the same tendency to book short-term rental (i.e. # of total short-term rentals are the same), the above equation can be rewritten as user $i$'s market share for all Airbnb choices $\sum_{j \in Airbnb} s_{ij}$:

$$\left(\frac{\partial \sum_{j \in Airbnb} s_{ij}}{\partial 1_{VSR}}\right)_{i=VS\ user} - \left(\frac{\partial \sum_{j \in Airbnb} s_{ij}}{\partial 1_{VSR}}\right)_{i=NM\ user} = -0.03374/0.53 \tag{15}$$

Assume utility function is:

$$U_{ij} = \beta X_j + \gamma_{NM} + \Delta\gamma \cdot 1_{VSR,j} + \varepsilon_{ij} \tag{16}$$

Where $\gamma_{NM}$ indicates normal users' sensitivity to observing any $VSR$ in a listing, $\gamma_{NM} + \Delta\gamma$ indicates $VS$ users' sensitivity to $VSR$. The market share of all Airbnb reservations is:

$$\sum_{j \in Airbnb} s_{ij} = 1 - s_{i,VRBO} = 1 - \frac{1}{1 + \sum_{j \in Airbnb} exp(U_{ij})} \tag{17}$$

Then:

$$\left(\frac{\partial \sum_{j\in Airbnb} s_{ij}}{\partial 1_{VSR}}\right)_{i=NM\ user} = +\gamma_{NM}\cdot s_{NM\ user,VRBO}\cdot \sum_{j\in Airbnb\ \&\ 1_{VSR}} s_{NM\ user,j} \tag{18}$$

$$\left(\frac{\partial \sum_{j\in Airbnb} s_{ij}}{\partial 1_{VSR}}\right)_{i=VS\ user} = +(\gamma_{NM}+\Delta\gamma)\cdot s_{VS\ user,VRBO}\cdot \sum_{j\in Airbnb\ \&\ 1_{VSR}} s_{VS\ user,j} \tag{19}$$

Denote a user's total probability of choosing any Airbnb listing with $VSR > 0$ as:

$$s_{NM\ user,Airbnb\ \&\ 1_{VSR}} = \sum_{j\in Airbnb\ \&\ 1_{VSR}} s_{NM\ user,j} \tag{20}$$

$$s_{VS\ user,Airbnb\ \&\ 1_{VSR}} = \sum_{j\in Airbnb\ \&\ 1_{VSR}} s_{VS\ user,j} \tag{21}$$

The DID results can be written as:

$$\begin{aligned}&+(\gamma_{NM}+\Delta\gamma)\cdot s_{VS\ user,VRBO}\cdot s_{VS\ user,Airbnb\ \&\ 1_{VSR}}\\ &-\gamma_{NM}\cdot s_{NM\ user,VRBO}\cdot s_{NM\ user,Airbnb\ \&\ 1_{VSR}} = -0.03374/0.53\end{aligned} \tag{22}$$

Note that we observe normal users' market shares in the data because almost all users are normal users, but we do not observe $VS$ users' market shares because we cannot track $VS$ users in all Airbnb and $VRBO$ bookings. However, the utility framework spells out how these two types of users differ. More specifically, the model implies:

$$\begin{aligned}\frac{s_{NM\ user,VRBO}}{s_{VS\ user,VRBO}} &= \frac{(1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR}))^{-1}}{(1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR}+\Delta\gamma\cdot 1_{VSR}))^{-1}}\\ &= \frac{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR}+\Delta\gamma\cdot 1_{VSR})}{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR})}\\ &= \frac{1+\sum_{j\in Airbnb} exp(\beta X)+exp(\Delta\gamma)\cdot\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM})}{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR})}\\ &= s_{NM\ user,VRBO} + s_{NM\ user,Airbnb\ \&\ VSR=0} + exp(\Delta\gamma)\cdot s_{NM\ user,Airbnb\ \&\ 1_{VSR}}\end{aligned} \tag{23}$$

This implies:

$$s_{VS\ user,VRBO} = \frac{s_{NM\ user,VRBO}}{s_{NM\ user,VRBO} + s_{NM\ user,Airbnb\ \&\ VSR=0} + exp(\Delta\gamma) \cdot s_{NM\ user,Airbnb\ \&\ 1_{VSR}}} \tag{24}$$

Similarly:

$$\begin{aligned}
\frac{s_{NM\ user,Airbnb\ \&\ 1_{VSR}}}{s_{VS\ user,Airbnb\ \&\ 1_{VSR}}} &= \frac{\frac{\sum_{j\in Airbnb\ \&\ 1_{VSR}} exp(\beta X+\gamma_{NM})}{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR})}}{\frac{\sum_{j\in Airbnb\ \&\ 1_{VSR}} exp(\beta X+\gamma_{NM}+\Delta\gamma)}{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR}+\Delta\gamma\cdot 1_{VSR})}} \\
&= \frac{\frac{\sum_{j\in Airbnb\ \&\ 1_{VSR}} exp(\beta X+\gamma_{NM})}{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR})}}{\frac{exp(\Delta\gamma)\sum_{j\in Airbnb\ \&\ 1_{VSR}} exp(\beta X+\gamma_{NM})}{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR}+\Delta\gamma\cdot 1_{VSR})}} \\
&= exp(\Delta\gamma)\frac{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR}+\Delta\gamma\cdot 1_{VSR})}{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR})} \\
&= exp(\Delta\gamma)\frac{1+\sum_{j\in Airbnb\ \&\ VSR=0} exp(\beta X)+exp(\Delta\gamma)\sum_{j\in Airbnb\ \&\ 1_{VSR}} exp(\beta X+\gamma_{NM})}{1+\sum_{j\in Airbnb} exp(\beta X+\gamma_{NM}\cdot 1_{VSR})} \\
&= exp(\Delta\gamma)\cdot\big(s_{NM\ user,VRBO} + s_{NM\ user,Airbnb\ \&\ VSR=0} + exp(\Delta\gamma) \\
&\quad \cdot s_{NM\ user,Airbnb\ \&\ 1_{VSR}}\big)
\end{aligned} \tag{25}$$

This implies:

$$\begin{aligned}
s_{VS\ user,Airbnb\ \&\ 1_{VSR}} &= \frac{1}{exp(\Delta\gamma)} \\
&\quad \cdot \frac{s_{NM\ user,Airbnb\ \&\ 1_{VSR}}}{s_{NM\ user,VRBO}+s_{NM\ user,Airbnb\ \&\ VSR=0}+exp(\Delta\gamma)\cdot s_{NM\ user,Airbnb\ \&\ 1_{VSR}}}
\end{aligned} \tag{26}$$

Plug these into the DID results:

$$\begin{aligned}
&(\gamma_{NM}+\Delta\gamma)\cdot s_{VS\ user,VRBO}\cdot s_{VS\ user,Airbnb\ \&\ 1_{VSR}} \\
&-\gamma_{NM}\cdot s_{VS\ user,VRBO}\cdot s_{VS\ user,Airbnb\ \&\ 1_{VSR}} = -0.03374/0.53
\end{aligned} \tag{27}$$

$$\begin{aligned}
&\frac{\gamma_{NM}+\Delta\gamma}{exp(\Delta\gamma)}\cdot\frac{s_{NM\ user,VRBO}\cdot s_{NM\ user,Airbnb\ \&\ 1_{VSR}}}{(s_{NM\ user,VRBO}+s_{NM\ user,Airbnb\ \&\ 1_{VSR=0}}+exp(\Delta\gamma)\cdot s_{NM\ user,Airbnb\ \&\ 1_{VSR}})^2} \\
&= -0.03374/0.53+\gamma_{NM}\cdot s_{NM\ user,VRBO}\cdot s_{NM\ user,Airbnb\ \&\ 1_{VSR}}
\end{aligned} \tag{28}$$

Because almost all users are normal users, the data gives us $s_{NM\ user,VRBO}$ (market share of $VRBO$), $s_{NM\ user,Airbnb\ \&\ VSR=0}$ (total market share of all normal Airbnb listings), and $s_{NM\ user,Airbnb\ \&\ 1_{VSR}}$ (total market share of all Airbnb VS listings). We also know $\gamma_{NM}$ from the utility regression. Thus, the only unknown in the above equation is $\Delta\gamma$. We can solve it easily and get $\Delta\gamma = -0.5332$.