Title
Role of Reputation Mechanisms for Quality Assurance in a Sharing Economy Platform: Evidence from Airbnb

Author
Kim, Jooho

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Role of Reputation Mechanisms for Quality Assurance in a Sharing Economy Platform: Evidence from Airbnb

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Management

by

Jooho Kim

Dissertation Committee:
Professor Sanjeev Dewan, Chair
Assistant Professor Tingting Nian
Assistant Professor Behnaz G. Bojd

2021

DEDICATION

To

my parents, grandparents, mj, and friends

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my committee chair, Professor Sanjeev Dewan, who has the attitude and the substance of a genius: he continually and convincingly conveyed a spirit of adventure in regard to research, and an excitement in regard to teaching. Without his guidance and persistent help this dissertation would not have been possible.

I would like to thank my committee members, Professor Tingting Nian and Professor Behnaz G. Bojd, for their full supports and guidance.

In addition, a thank you to Professor Mingdi Xin, who devoted her time to join in my practice research presentation and to suggest wonderful feedback. I also thank you to Professor Vidyanand Choudhary and Professor Vijay Gurbaxani for their supports.

My life and experience at University of California, Irvine have been amazing with my colleagues. I had too many good friends to mention their names one by one here but I thank you to all my dear friends in our department.

Lastly, I dedicate this dissertation to my parents and to my grandmother who passed away when I started my PhD in 2015.

VITA

Jooho Kim

2012	B.B.A. in Business Administration, Sungkyunkwan University
2015	M.S. in Management Information Systems, Seoul National University
2015-21	Research and Teaching Assistant, The Paul Merage School of Business, University of California, Irvine
2021	Ph.D. in Management, The Paul Merage School of Business, University of California, Irvine

FIELD OF STUDY

Information Systems

ABSTRACT OF THE DISSERTATION

Role of Reputation Mechanisms for Quality Assurance in a Sharing Economy Platform: Evidence from Airbnb

by

Jooho Kim

Doctor of Philosophy in Management

University of California, Irvine, 2021

Professor Sanjeev Dewan, Chair

Sharing economy platforms bring individual buyers and sellers together to promote transactions between the two parties. Since most of the platforms include decentralized network of individual sellers who provide their own products or services, these lack standardized or established quality which may lead to quality uncertainty. Although sharing economy platforms rely on user-generated reviews and seller-provided information to provide trust between buyers and sellers, these would not completely reduce the quality uncertainty. In my dissertation, I examine the impact of platform-managed quality certification and simultaneous review system to find out how these mechanisms address quality uncertainty in the context of Airbnb. Leveraging a quasi-experimental design in combination with a machine learning algorithm, I find that the quality certification launched by Airbnb has differential impacts on consumers, property owners, and the platform. Also, I show how the reciprocity under the bilateral review system affects volume, valence, and semantic diversity of reviews. My findings have significant implications for researchers and practitioners who deal with quality management and review system designs especially within an online platform area.

CHAPTER 1

Introduction

Online marketplaces, including sharing economy platforms, crucially rely on review and reputation systems to provide trust between buyers and sellers, who are separated by time and space at the time of the transaction. Yet, review systems are not perfect and do not completely eliminate quality uncertainty, in part due to the documented problem of "review inflation" in bilateral review systems (see, e.g., Zervas et al. 2021, Filippas 2019). A variety of quality disclosure and assurance mechanisms are used in practice to complement to review systems, including price, brand name, warranties, licensing and third-party or government mandated quality certification (see Dranove and Jin 2010 for a survey). Considerable prior research has studied the theory and practice regarding quality assurance mechanisms in diverse industries, but only a few have looked at online marketplaces (Dewan and Hsu 2004, Edelman 2011, Lewis 2011, Barach et al. 2020), and none have studied the context of sharing economy platforms — which is the focus of our work.

Another departure from prior work on quality certification is that whereas most of the focus has been on third-party certification (Arora and Asundi 1999, Gao et al. 2010, Ozpolat et al. 2013), we look at self-certification of sellers by the platform itself. In this regard, Barach et al. (2020) is closest to our work, in that they study "steering" of buyers to recommended sellers, but their context is an online B2C labor market, whereas ours is one of the leading sharing economy platforms, namely Airbnb, whose revenues are projected to be $3.5 billion per year by 2020 (Fortune 2017). In particular, we examine the impact of the Airbnb Plus service, announced in February 2018, wherein properties can pay a one-time fee of $149 to be considered for a Plus certification. The properties need to have at least 4.8 out of 5 rating and the host needs to have above 90% response rate to guest questions, at least 80% of 5-star host reviews, and no cancellations in the last year. Additionally, properties have to undergo a 100-point quality inspection, covering how well properties are designed, equipped, and maintained. Properties that successfully clear all of these hurdles are listed with a Plus badge on the Airbnb site.[1]

Yet, one can question the credibility of the quality signal provided by a Plus certification, for a number of reasons. First, the certification is provided by the platform itself, which may be subject to a conflict of interest, since the platform has a stake in the revenue potential of the listings. For example, the platform may have the incentive to over-allocate Plus certifications in areas where it faces stronger competition from the hotel industry. Second, the fee for applying for a Plus certification is a one-time expense of $149, which is a fraction of the average daily rent of about $250 for Plus properties (Skift 2018a). However, the inspection fee might be a small fraction of the total cost of designing, equipping, maintaining and marketing required to raise the quality of a listing to the Plus level. The quality signaling literature calls for a costly signal, so that high-quality properties can more easily afford to convey it, relative to low-quality ones (see, e.g., Spence 1977). It is an open question whether the Plus signal is costly enough to cause separation between truly high-quality listings and pretenders who may do what it takes to obtain the certification, but then slack of, counting on the fact that they may not lose the certification. Third, industry reports

[1] According to the Airbnb website (airbnb.com/s/Plus_homes) the benefits of a Plus certification include a Plus badge and favorable placement in search results. However, as we discuss in Section 5, the search rank effect is rather weak, so the economic impacts of the certification can be primarily attributed to quality signaling.

suggest that Airbnb does not have the capacity to conduct ongoing quality inspection (Skift 2018b), so that once properties receive the certification, they may have the incentive to lower their quality (and therefore maintenance costs) over time, hoping on not losing their certification. Therefore, it is an open empirical question as to whether the Plus certification is just "cheap talk" or a truly credible signal of quality (see Barach et al 2020 for a similar point) — an issue that motivates our analysis.

We contribute to the growing literature on the sharing economy, where the popularity of the phenomenon has motivated researchers to examine various aspects of the sharing economy and its impact on the broader economy (see, e.g., Sundararajan 2016). An important stream of this literature has focused on how sharing economy industries affect their traditional economy counterparts; e.g., the impact of Uber on the taxi industry (Cramer and Krueger 2016) and the impact of Airbnb entry on hotel industry revenues (Zervas et al. 2017). Our work is related to the latter, in the sense that the Airbnb Plus service was launched in order for Airbnb to complete more effectively against the standardization and relatively lower quality uncertainty offered by the hotel industry. Indeed, Airbnb has more to gain from the marginal consumer who chooses an Airbnb Plus listing over a hotel room, as compared to the marginal consumer choosing a Plus listing over a non-Plus one. In this context, our research questions are as follows: what is the impact of a Plus certification on the booking rate of the listings receiving this certification (*direct effect*); what is the impact of a Plus certification on other non-Plus listings nearby (*externality effect*); finally, what is the overall impact of the launch of the Plus service on local platform revenues (*local platform effect*).

To address these questions, we have compiled daily panel data from Airbnb over the period of July-December, 2018, including around 50,000 listings in multiple cities around the world where Airbnb Plus service was recently launched. Since the Airbnb Plus service was launched in different cities at different times, and individual listings received the certification at different times, we are able to deploy a Difference-in-Difference (DD hereafter) strategy to empirically measure the direct and externality economic impacts of the quality certification. To deal with potential endogeneity of treatment (i.e., receiving a Plus certification) we use DD in conjunction with suitable matching methods (propensity score matching and look-ahead matching), as we discuss in detail below.

To summarize our results, we find positive direct effects, so that receiving a Plus certification raises a typical listing's booking rate by 7.6% on average. We also find significant externality effects, so that one or more Plus listings within a 2-kilometer circular zone depresses the booking rate of other non-Plus listings within that zone by about 1.5% each.[2] Putting these effects together, the net impact on local platform revenues is positive, to the tune of over $80,000 per 2-kilometer zone on an annual basis. In sum, our results provide prima facie evidence that Plus certification is a credible signal of quality, it creates a separation in the revenue potential of listings with and without the certification, and higher net revenues for the platform in areas where Plus listings are introduced. We discuss implications of our results in the concluding section.

Quality uncertainty is a major source of friction in any trading system (Akerlof 1970), especially in electronic markets where buyers and sellers are separated by time and space.

[2] We choose a 2-kilometer radius zone around a focal Airbnb listing to identify economic impacts of Plus certification, since competition and externality effects are likely to be localized. The results are not sensitive to the specific magnitude of the radius (i.e., 2 km).

Review and reputation systems play a key role in overcoming information asymmetries and establishing trust in online platforms (Dellarocas 2003, Dewan and Hsu 2004). The design of review systems, in turn, plays a central role in determining the informativeness of these systems, influencing the willingness of users to engage with them and to provide honest unbiased feedback (e.g., Avery, Resnick and Zeckhauser 1999). One specific aspect of the design of review systems that has drawn considerable research interest is reciprocity in bilateral review systems, which is increasingly common in sharing economy platforms, such as Airbnb and Uber, wherein buyers rate sellers, and vice versa (see, e.g., Dellarocas and Wood 2007, Bolton et al. 2012, Proserpio et al. 2018, Fradkin et al. 2021). Within this stream, researchers have examined a variety of issues, including reciprocity and retaliation (Proserpio et al. 2018), lack of participation and biased reviews (Dellarocas and Wood 2007) and review inflation (Filippas et al. 2018, Zervas et al. 2021), among other topics.

One way to deal with the issues caused by reciprocity and retaliation — which in turn results in biased reviews and review inflation — is to change the timing and visibility of the review system. The traditional design of bilateral review systems was asynchronous, wherein one party to the transaction posted their review, and then the other. In such a setting, the first party would be reluctant to post a negative review fearing retaliation by the second mover. Accordingly, a less-than-satisfied user would either refrain from posting any review whatsoever, or suppress any negative feedback, with detrimental consequences for the informativeness of the review system either way. One way to deal with this issue is to constrain the two-sided review system to become a one-sided one; a case in point is eBay, which forced sellers to only post positive buyer reviews thereby seeding the bilateral review system with positive reciprocity (see, e.g., Klein et al. 2016, Hui et al. 2019). Another approach is to convert an asynchronous review system into a simultaneous one, first suggested by Bolton et al. (2012), where both parties to a transaction have a limited amount of time in which to post their reviews in a double-blind manner, thereby reducing the retaliation problem. Indeed, most bilateral review systems have implemented such a simultaneous review regime, although the evidence on the impact of this policy change in prior research is somewhat mixed.

A few recent studies have focused on the review policy change from asynchronous to simultaneous system and its impact (Bolton et al. 2012; Mousavi and Zhang 2018; Fradkin et al. 2021). While these studies have examined the same issue, they had either mixed findings or different approaches to estimate the effect. Using a laboratory experiment, Bolton et al. (2012) showed that this policy change leads to relatively more negative reviews and reduction in review frequency. On the other hand, Fradkin et al. (2021) found somewhat opposing results even if the study was examining the same policy change and its impact on the valence and frequency of reviews. They proved that valence of the reviews became less positive, although the magnitude was small, and the frequency increased. They claim that "unveiling mechanism" significantly affected these results since this mechanism attracts people to leave their reviews in order to find out how their counter-parities had left the reviews to them. The authors point out that Bolton et al. (2012) lacks the "desire to unveil" in their lab setting and this might be one of the reasons why the authors generate the contrasting results. Mousavi and Zhang (2018) also have studied the same policy change but have used an observational data to examine the impact before and after the policy change. Specifically, they have applied Regression Discontinuity Design to measure the effect of simultaneous review system on contents of textual reviews using various text-mining approaches.

While the aforementioned studies have investigated the shift from asynchronous to simultaneous review system and its effects using different estimation strategies, their results are somewhat mixed in terms of direction and magnitude. Also, Mousavi and Zhang (2018) draw out limited findings from their textual analysis which needs further elaboration. According to Airbnb, they changed the review system from asynchronous to simultaneous in July 2014. Under the previous review system, a review went public as soon as either a guest or a host posted her review regardless of whether the other side wrote her review or not. This allowed for guests or hosts to wait until the other side submitted her review and write their own review after observing what the other side had written. Under this asynchronous review mode, both sides had the incentive to wait and see what the other side posted first, so that positive reviews are reciprocated and negative reviews tended to generate a corresponding negative counter-review. Of course, the other possibility is that dissatisfied users would not post any review at all, leading to suppressed negative feedback and artificially-positive rating. In introducing the simultaneous review policy, Airbnb claimed that the new system would encourage guests and hosts to generate honest and informative reviews — a claim which serves to motivate this study. Therefore, we revisit the simultaneous review system and its effect on valence, volume, and contents of the reviews in order to resolve this ambiguity.

We apply a cross-platform Difference-in-Differences (DID) approach to a sample of cross-listed properties from Airbnb and TripAdvisor; i.e., we match properties that are listed on Airbnb with the exactly same properties that are also listed in TripAdvisor, which has a one-directional review system in which only the guests review the hosts. In order to detect the cross-listed listings from both platforms, we match Airbnb listings with TripAdvisor listings based on hosts' first name that are shown from the websites. Based on the matched candidate pairs, we use the number of bedrooms and bathrooms (listing characteristics that are not going to change over time) as matching covariates to filter out more precise candidate pairs. For the third and the last step, we leverage latitude and longitude, and property images, respectively. We run image matching algorithm to find out the same posted images from candidate matched listings. This algorithm enables us to map multiple images from one listing to the other sets of images from the potential matched candidates. These fine-grained 4 step algorithms allows us to find the exact same properties that are cross-listed on the two platforms. Using Python, we scrape all the reviews, review and listing-level attributes for all the current listings from both platforms in multiple cities in U.S.

In terms of our results, we find that the introduction of the simultaneous review system reduced ratings by 0.13-0.17, on average, on a scale of 1 to 5 (rounded to the nearest 0.5). The result is robust to controlling for review and listing-level variables such as the number of pictures and review length. The impact of review system is more salient for listings with above-mean rating. Turning to the nature of the textual feedback, we find that the net positive word frequency reduced after the policy change, sentiment of the feedback reduced, as did the entropy (i.e., the reviews are more focused). The number of reviews went up, suggesting that feedback frequency is higher, consistent with Fradkin et al. (2021). Our topic modeling analysis finds that the reviews incorporate a smaller number of topics, after the policy change.

CHAPTER 2

Impact of Quality Certifications on Demand at a Sharing Economy Platform: Evidence from Airbnb

This chapter studies the impact of quality certifications in a sharing economy platform, especially in Airbnb contexts. To be specific, we examine how the demand of rental properties changes after the properties receive Plus certification badges which presumably signal high quality in terms of style, amenities, maintenance, and reliability of the hosts. We employ a Difference-in-Difference estimation strategy in combination with two-step matching schemes such as Coarsened Exact Matching and Machine Learning Binary Classification algorithm. We find that Plus certification increases the weekly booking rate of Plus properties by 7.6% more than that of non-Plus properties on average. We also find that Plus certification truly generates a value for the properties without any quality signals, but it may not provide additional value for the properties with other types of quality signals. Our analysis sheds light on a sharing economy area when other sharing economy platforms create quality-tier products or services within their platforms.

Literature Review

We start by relating our work to two streams: the general literature on quality assurance and disclosure mechanisms; and prior work on quality assurance in the sharing economy in particular. We then describe the conceptual framework that underlies our empirical analyses.

The seminal work of Akerlof (1970) coined the term "market for lemons" to describe how quality uncertainty can lead to market frictions, even market failure. In the stereotypical example of a used car market, if buyers cannot discern true quality, then quality uncertainty depresses customer willingness to pay, which in turn drives high quality cars out of the market — leaving only lemons behind. A variety of institutional mechanisms have emerged to mitigate the impact of quality uncertainty, including brands (Rao and Monroe 1989), advertising (Milgrom and Roberts 1986), warranties (Grossman 1981), review and reputation systems (Dellarocas 2003), licensing (Stigler 1971) and quality disclosure and certification, which can be voluntary, government mandated or third-party (see Dranove and Jin 2010 for a survey). These studies have examined a variety of industries, including education, health care, food services, sports and finance, among others.

With respect to electronic markets — most relevant to our present study — Dewan and Hsu (2004) study the impact seller reputation system in eBay stamp auctions. They show that realized auction prices on eBay reflect an adverse selection discount in the range of 10-15% of the nominal value of the goods. Seller reputation does mitigate this adverse selection problem, through both higher probability of sale and higher realized sales price. Although these effects are statistically significant, they are economically modest, leaving room for improvement. Dimoka et al. (2012) also study quality uncertainty in eBay and make the additional important point that if review systems are to be useful then they must account for both product uncertainty and seller uncertainty — a feature that we capture in our empirical design. Lewis (2011) also studied adverse selection on eBay focusing on the trading of used cars on eBay Motors. They find support for their hypothesis that voluntary disclosure of quality information (such as photos, product descriptions and maintenance logs) serves to overcome information asymmetry and mitigate adverse selection in the electronic market. While voluntary disclosure works to some extent, quality certification by neutral third parties is likely to be more effective. Prior work has studied the impact of third party certification (e.g., ISO 9000, 9001 standards in the software industry) on: firm share prices (Nicolau and Sellers 2002), exports (Gao et al. 2010), product price (Arora and Asundi 1999), and revenue or sales (Levine and Toffel 2010). They all generally found a positive relationship between quality certification and organizational performance. Looking at an online context, Özpolat et al. (2013), empirically examined the impact of third-party assurance seals on purchase conversion using online shopping data. They found that the presence of the assurance seals decreased product uncertainty and, in turn, increased demand in an online market. Edelman (2011) found a contrary result where third-party certification did not work as intended. Specifically, he found that TRUSTe certified websites were less than half as trustworthy as non-certified websites, presumably because the third party was issuing certifications without proper verification. This underscores the importance of the neutrality, diligence and proper incentive alignment of third-party certifiers if this mechanism is to be effective in overcoming quality uncertainty.

These issues come to the fore when the platform itself provides quality disclosures or certifications, which has been studied in some recent work. Barach et al. (2020) studied an online labor market which matches buyers (i.e., job openings) with freelance sellers. They conduct randomized experiments in which the platform steers high willingness-to-pay buyers to money-back guaranteed sellers. They found that such a mechanism increases the quality of matches, driven by the fact that the money-back guarantees are informative about uncertain seller quality. In a similar setting, Horton and Johari (2018) showed that buyer signaling of preferences for price and quality allows for the proper sorting of sellers to buyer type, improving matching efficiency and quality. In contrast to these papers, we examine platform self-certification in a sharing economy context, which raises some new issues for quality signaling, which we discuss next.

Sharing economy platforms typically rely on bilateral review systems wherein both sides of the platform review each other. Thus, on Airbnb, guests can provide reviews of hosts and their properties, and hosts can also provide reviews of guests after they have completed their stay. Proserpio et al. (2018) found that highly rated hosts could increase their property price. However, other studies have found review inflation in bilateral review systems; i.e., each side is fearful of leaving poor reviews, for fear of subsequent retaliation (Zervas et al. 2021, Filippas et al. 2019). With few exceptions, ratings and reviews are overwhelmingly positive on the Airbnb platform, limiting their ability to mitigate quality uncertainty. This has led sharing economy platforms to turn to other mechanisms for signaling quality, such as cancellation policy. Zalmanson et al. (2018) found that changing cancellation policy of a property from 'lenient' to 'strict' is viewed as a quality signal by potential guests and a stricter cancellation policy led to an increase in a property's booking rate. Professional photographs can also be an informative signal of quality as demonstrated by Zhang et al. (2017).

These quality-assurance mechanisms in sharing economy platforms still fall short when stacked up against the well-known brand names and long-standing reputations of hotels. Unlike other quality cues like photo quality and cancellation policy, a Plus certification operates in a more direct way to reduce information asymmetry, wherein the platform itself takes on the roles of quality inspector and certifier. The resulting quality badge is still subject to an adverse selection problem because the application fee of $149 for Airbnb Plus certification is relatively modest, and it may not deter lower quality properties from seeking the certification. Further, there is a moral hazard problem due to the fact that Airbnb does not have sufficient capacity to keep monitoring quality past the initial inspection, providing incentives for listings to let quality slip over time. Whether the Plus certification is able to mitigate these problems is ultimately an empirical question, which we study using the empirical framework described next.

Conceptual Framework

The conceptual framework for our analysis is depicted in Figure 2.1. We explore 1) how Airbnb Plus certification affects a property's booking rate — *Direct Effect*; 2) how the presence of Plus listing nearby affects non-Plus listings' booking rate — *Externality Effect*; and 3) how the introduction of Airbnb Plus service affects overall platform revenues in a neighborhood — *Local Platform Effect*. In doing so, we control for listing and neighborhood characteristics, as well as listing and city-time fixed effects.

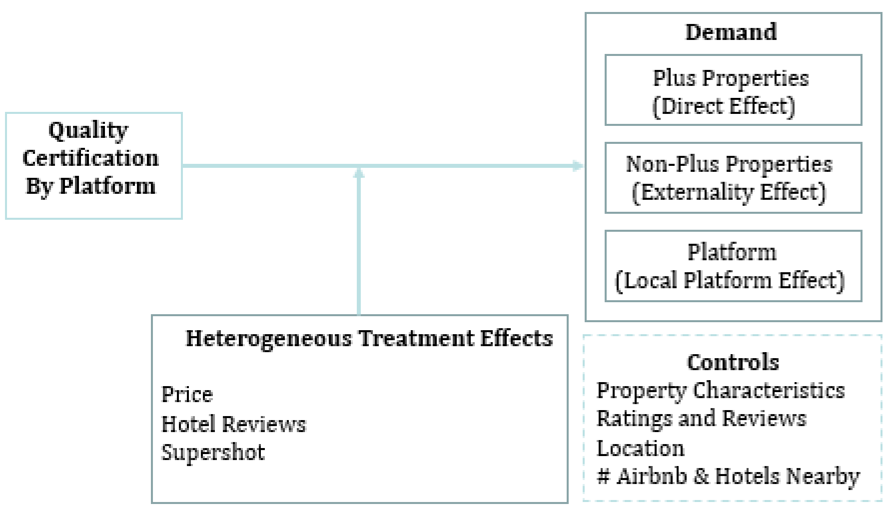

Figure 2.1 Conceptual Framework

We use a Difference-in-Difference estimation approach, in conjunction with propensity score matching and Coarsened Exact Matching, to generate causal results for the questions identified above. Before we describe our empirical approach and identification strategies, we describe our context, data set and key variables. Note that we discuss *Externality Effect* and *Local Platform Effect* in the next chapter.

Data and Key Variables

To address our research questions, we use a novel dataset collected from one of the largest short-term rental platforms, Airbnb, which has over 5 million listings across 81,000 cities around the world.[3] Airbnb Plus, introduced in February 2018, is a high-end tier consisting of listings that have been inspected and certified by the platform. According to Airbnb, Plus-certified listings need to satisfy several criteria. Specifically, listings should maintain an average rating of 4.8, be thoughtfully designed, tastefully furnished, well-equipped with amenities, and immaculately maintained. Hosts apply for certification after paying the inspection fee, which was $149 at the time of this study. Airbnb inspectors visit the property to check whether every qualifying criterion is satisfied.

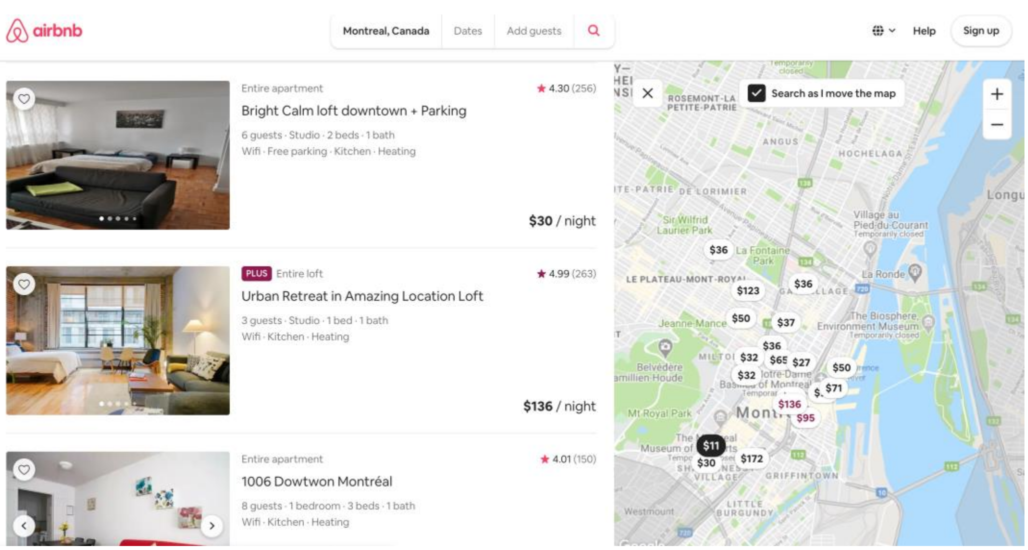

Figure 2.2 Plus Listing in Airbnb Search Impressions

[3] Airbnb is an online platform that connects people who want to rent out their homes, ranging from private rooms to the entire property, with people who are looking for accommodations in that locale.

Once a listing passes the inspection, it is certified as an Airbnb Plus property, and a Plus badge is added to the listing impression returned in search results (see Figure 2.2, second item on the list). We compiled a data set consisting of 49,850 listings in 8 cities (Auckland, Austin, Cape Town, Mexico City, Montreal, Phoenix, Seattle, and Sydney) from July 2018 to December 2018, using a web crawler.

We define the key variables deployed in our empirical analysis, noting that our unit of analysis is listing-week; i.e., the data consists of weekly observations for a set of listings in the 8 cities included in our study.

Our main dependent variable is *booking rate* as we expect that the most direct impact of receiving a Plus certification would be on the demand for a listing, all else equal. For a given listing on a given day, we first define daily booking ratio as the proportion of days in the following 30 days that the property is reported to be occupied. We then average these daily booking ratios in a given week to obtain the average weekly booking rate — our dependent variable.

For each property, we collect the geolocation (latitude and longitude), zip-code, the number of beds and baths, capacity (in terms of maximum number of guests that can be accommodated), the textual description of the property and the number of photos shown on Airbnb. The number of reviews for listings and hosts indicate the number of reviews received for each listing and host, respectively. Star rating is the average of rating the listing has received from customers, and it ranges between 1 and 5. Cancellation policy is categorized into several types, namely 'super strict', 'strict', 'moderate', and, 'flexible'. We create a binary cancellation policy dummy variable, which is equal to 1 if the policy is 'strict' or 'super strict', and it is set to 0 if the cancellation policy is 'moderate' or 'flexible'. Superhost is a dummy variable, coded as 1 if the host is a Superhost, and as 0 if not. Room type is coded as 1 if the listing is an entire house, or 0 if it is a private/shared room rental. Lastly, Listing Duration is defined as the number of weeks since the date that the property was first listed on the Airbnb platform. Last, but not least, we observe whether or not a listing has a Plus certification in a given week or not, coding a dummy variable DD accordingly (DD is set to 1 following the receipt of a Plus certification, and it is 0 otherwise). Figure 2.3 shows a visualization of the geolocation of Plus (red points) and non-Plus (green points) listings for four of the cities in our data set. We can see that both types of listings are geographically dispersed across the cities.

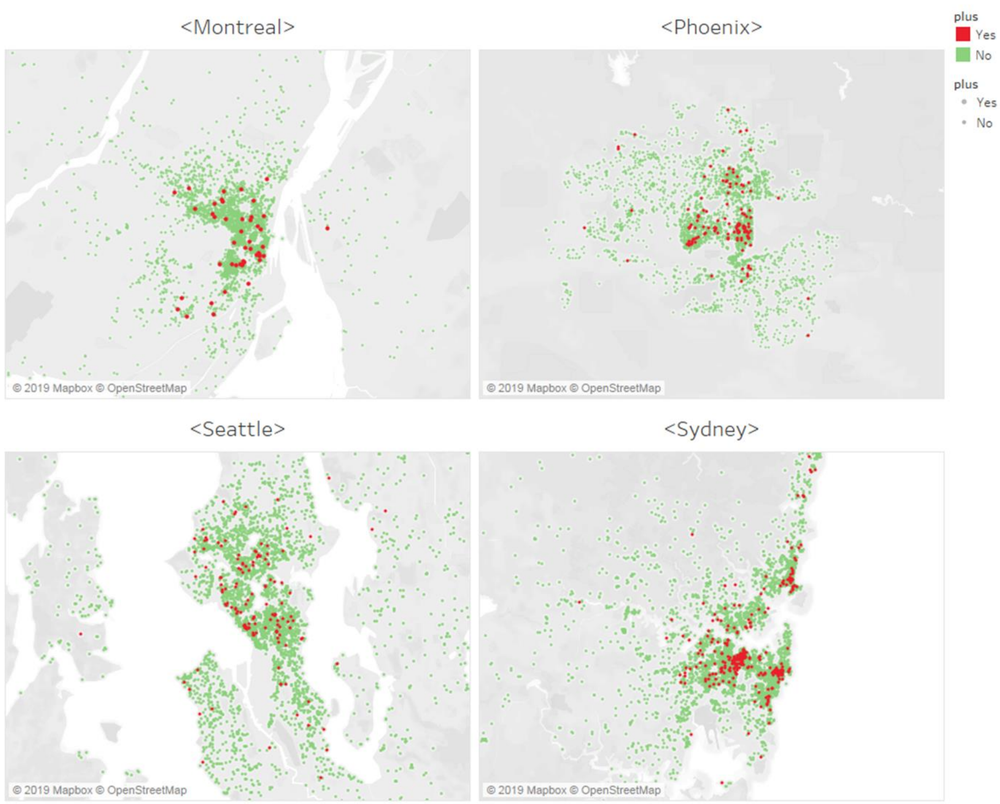

Figure 2.3 Visualization of Plus and Non-Plus Listings in Four Cities

In our analysis, it is important to control for the underlying popularity of the neighborhood in which an Airbnb listing is located, which in turn affects booking demand for the listing. We construct two popularity proxies as, follows. First, we determine the number of other Airbnb listings within a 2-kilometer radius range of the focal listing. Popularity is also reflected in hotel room bookings. While we do not have data on actual hotel reservations, we are able to observe the number of hotel reviews posted on TripAdvisor — which is likely to be correlated with the number of hotel bookings. Accordingly, based on the property's address, we use web-scraping to count the number of hotel reviews posted on TripAdvisor for all hotels located within a 2-kilometer radius range of the focal listing. In this way, we attempt to address the potential endogeneity in acquiring the Plus certification, as a function of neighborhood popularity, as we further discuss below.

Table 2.1 shows the summary statistics of selected variables in the data set, listed separately for Plus and non-Plus listings. For the Plus Listings there are two sets of columns, averaged for the time periods before and after receiving Plus certification, respectively. Plus listings, before or after certification, are much more likely to be a Superhost as compared to non-Plus listings (76-78% versus 37%), they receive more reviews and they have a higher star rating (4.98 versus 4.85). Plus listings are also located in more popular neighborhoods as reflected in both the number of Airbnb listings and number of hotel reviews within a 2 km radius range. Interestingly, the average daily price of Plus and non-Plus listings are almost exactly the same (around $218), but the booking rates of Plus listings after certification are substantially higher (67% versus 56%) resulting in higher revenues. Indeed, the real impact of Plus certification is an increase in booking rate, while price is essentially unchanged. The correlation matrix is presented in Table 2.2. Here, we see that price is negatively correlated with booking rate, while Number of Reviews, Star Rating and Superhost status are all positively correlated with booking rate — as we might expect. As we discuss in the following section, we use matching methods to restrict the comparison of Plus listings with similar non-Plus listings, which is at the heart of our identification strategy for the impact of Plus certification.

Table 2.1 Variable Names and Descriptions

Variable Name	Variable Description	Plus Listings						Non-Plus Listings		
		Before Plus Certification			After Plus Certification					
		Obs	Mean	S.D.	Obs	Mean	S.D.	Obs	Mean	S.D.
Booking_Rate	Average daily booking rate	10,426	0.58	0.29	23,275	0.67	0.26	809,378	0.56	0.35
Price	Average daily price	10,430	218.64	447.80	23,280	218.84	301.11	809,395	218.71	368.60
Num_Reviews	# Reviews a listing received	10,430	50.04	52.11	23,280	60.86	57.15	809,395	37.49	55.02
Star_Rating	Rating (1 to 5) for a listing	9,812	4.97	0.12	22,879	4.98	0.11	602,348	4.85	0.27
Num_Host_Reviews	# Reviews a host has received	10,430	215.59	356.15	23,280	242.36	437.12	809,395	289.12	1009.37
Superhost	= 1 if superhost; 0 otherwise	10,430	0.76	0.43	23,280	0.78	0.41	809,395	0.37	0.48
Cancellation Policy	= 1 if 'strict'; 0 if 'lenient'	10,430	0.46	0.49	23,280	0.56	0.49	809,395	0.48	0.5
The Number of Pictures	# Pictures a listing has posted	10,430	26.78	12.76	23,280	30.34	13.52	809,395	19.78	12.66
Listing_Duration	# Weeks since joined Airbnb	9,381	192.29	97.78	20,673	206.98	90.71	551,101	177.77	95.1
Listings in 2km	# Airbnb listings in 2km zone	10,430	711.59	553.89	23,280	699.77	514.03	809,395	520.96	546.69
Hotel Reviews in 2km	# Hotel reviews in 2km zone	8,646	258.21	335.85	19,263	285.47	357.38	579,327	250.13	365.95
Bathrooms	# Bathrooms	10,430	1.68	0.98	23,280	1.68	0.98	808,688	1.68	1.06
Bedrooms	# Bedrooms	10,430	1.94	1.43	23,280	1.94	1.43	809,227	2.00	1.42
Beds	# Beds	10,430	2.19	1.94	23,280	2.19	1.94	808,827	2.69	2.05
Capacity	# People accommodation	10,430	4.59	3.12	23,280	4.59	3.12	809,395	4.72	2.94
Room Type	1 if 'entire house'; 0 else	10,430	0.94	0.24	23,280	0.94	0.24	809,395	0.81	0.39

Table 2.2 Correlation Matrix

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1)DD	1								
(2)Booking_Rate	0.05	1							
(3)Price	0.05	-0.12	1						
(4)Revenue	0.06	0.09	0.78	1					
(5)Num_Reviews	0.03	0.10	-0.19	-0.14	1				
(6)Num_Host_Reviews	-0.01	-0.02	-0.00	0.01	0.12	1			
(7)Star_Rating	0.09	0.06	0.03	0.04	0.05	-0.14	1		
(8)Superhost	0.11	0.06	-0.12	-0.08	0.23	-0.10	0.40	1	
(9)Listing_Duration	0.05	0.03	0.05	0.05	0.25	0.13	0.00	0.03	1

Identification Strategy: Direct Effect

Our quasi-experimental empirical strategy exploits the fact that the Airbnb service was introduced in different cities at different points in time, and within each city, listings received the quality certification at different points in time. A specific listing is defined to be "treated", once it obtains an Airbnb Plus certification. The control group consists of listings that do not obtain Airbnb Plus certification in our observation window. Specifically, we employ a Difference-in-Difference (DD) approach to causally estimate how receiving a Plus certification affects a listing's own booking rate (*Direct Effect*), and the booking rate of other listings nearby (*Externality Effect*). Having access to observations of a large number of property listings over time, we are able to isolate a treatment group of listings that are non-Plus at the beginning of the observation window but get treated at some point; i.e., receive a Plus certification. By comparing the booking rate change before and after the acquisition of Plus certification with the booking rate change of non-Plus listings, we are able to identify the impact of quality certification on booking rate. However, properties that receive Plus certification may gain higher booking rate even in the absence of the certification. To address the heterogeneity across properties, we use the Difference-in-Difference (DD) approach to "difference out" the pre-treatment trend associated with Plus properties. This allows us to distinguish the component of the post-treatment changes attributable to the quality certification from the component that is attributable to the kind of properties that are more likely to become Plus listings. In other words, we can separate the effect that is truly due to the treatment from that due to "selection." We now describe our DD specifications.

We start with our models for estimating direct effect associated with Plus certification, followed by a discussion of our identification strategy to assure that our empirical estimates are causal. The direct effect is estimated using the following two-way fixed effects framework for listing *i* in city *j* in time period t:

$$(Booking_{Rate})_{ijt} = \alpha + \beta_1 DD_{it} + \beta_2 ln(Num\_Reviews)_{it} + \beta_3 ln(Num\_Host\_Reviews)_{it} + \beta_4(Star\_Rating)_{it} + \beta_5(Superhost)_{it} + \beta_6 ln(Listing\_Duration)_{it} + \beta_7 ln(Price)_{it-1} + X_i + V_{jt} + \varepsilon_{ijt} \quad (2.1)$$

In the direct effect version of this model, each focal listing i in the treatment group received a Plus certification during the observation window, so that the indicator variable DD_{it} is equal to 1 if listing *i* has a Plus listing in time period *t,* and it is set to 0 otherwise. That is, treatment in the direct effect model happens when the focal listing itself receives a Plus certification.

The coefficients β_1 measures the average treatment effect in the direct effect model. Note that our DD setting is different from the standard DD setting in the sense that each unit receives treatment at different points in time (Babar and Burtch 2020; Ozturk et al. 2016; Stevenson and Wolfers 2006; Zhang et al. 2017). Because price in the current period is endogenous — consumers may make decisions based on some unobserved variables that may also affect listing owners' pricing behavior — we use the lagged price $Price_{it-1}$ as an instrument for the current period price (Villas-Boas and Winer 1999). We include listing fixed effects, X_i , to control for unobserved property-level time-invariant factors. Given the temporal nature of our data, we need to control for differential time trends across cities, such as Airbnb's advertising efforts in a local market, seasonality in short-term rental markets, and so on. For instance, the demand for short-term rental may change differently overtime across cities, due to factors such as seasonality and major events (e.g. sports or entertainment events) that take place in a certain city. The supply of Airbnb listings may also change as a result of Airbnb's advertising campaign in a city, which might in turn affect Airbnb properties' booking rate. Therefore, we include city-time fixed effect, V_{jt}, to capture all unobserved time-varying factors that affect all properties in a city temporally. We also include property time-varying observables such as the number of property reviews, host reviews, and the star rating to control for the impacts of these user generated reviews on our outcome variables. $Superhost_{it}$ is the dummy variable equals to 1 if the host of the property is a Superhost or 0, if not. $Listing\_Duration_{it}$ measures how many weeks since listing i has been on Airbnb by the time period t.

As we saw in our discussion of descriptive statistics in the previous section, Plus listings are different from non-Plus listings in a number of ways; i.e., they have a larger volume and valence of reviews, they are more likely to be a Superhost, and they tend to be in more popular locations relative to non-Plus listings. Besides these observable characteristics, Plus listings are also likely to be different in terms of a number of unobservable or intangible characteristics, such as style, design, range of amenities, cleanliness, etc. If the decision to apply for a Plus certification depends on these observable and unobservable characteristics, then we may have an endogeneity problem. We deal with this endogeneity problem through suitable matching methods. We use Coarsened Exact Matching (CEM) on observable listing and neighborhood characteristics. To address potential endogeneity due to selection on unobservable characteristics, we use a novel Propensity Score Matching (PSM) method that utilizes natural language processing of listing textual reviews, as we discuss below.

In the direct effect model (see Figure 2.4), the treatment group consists of listings that received a Plus certification at some point in the observation window. For each focal listing in the treatment group, we first use Coarsened Exact Matching (CEM) on the observable listing and neighborhood characteristics to find candidate listings for the control group (Iacus et al. 2012).

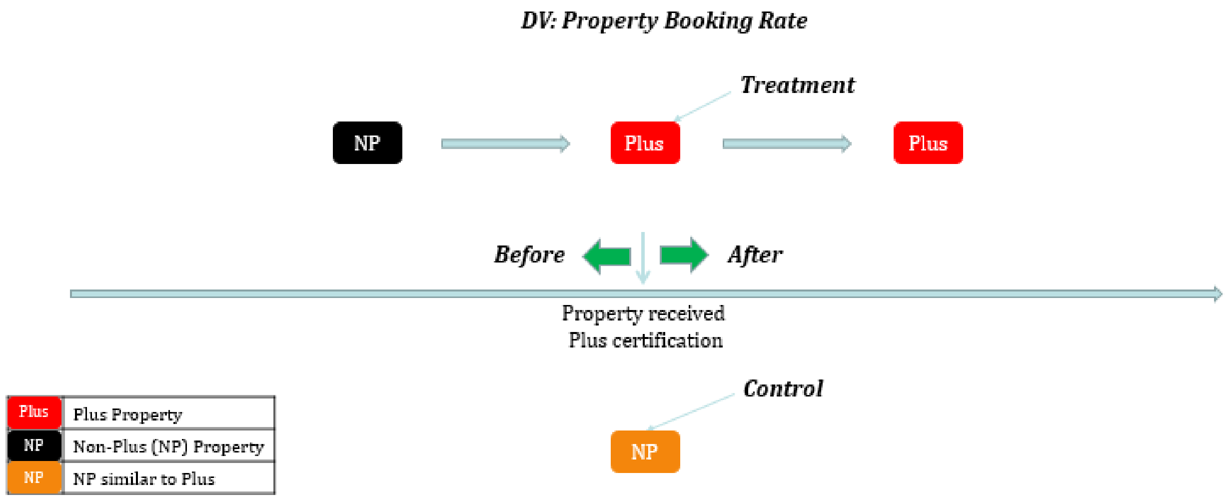

Figure 2.4 Treatment and Control Units in the Direct Effect Model

Matching covariates include listing characteristics such as the number of bathrooms, bedrooms, beds, capacity, cancellation policy, the number of pictures, tenure on the site, room type, and zip code. We also account for neighborhood popularity, which influences the incentives for seeking a Plus certification, and also Airbnb's inclination to certify listings in popular areas or regions with a higher concentration of hotels. We include two proxies for neighborhood popularity: the number of Airbnb listings inside a 2km radius range of a focal listing, and the number of hotel reviews posted on TripAdvisor within the same zone. Figure 2.5 shows that the imbalance between the treatment and control groups is substantially reduced by CEM, using the L1 measure of imbalance. For most variables the L1 distance between the treatment and control groups goes towards zero.

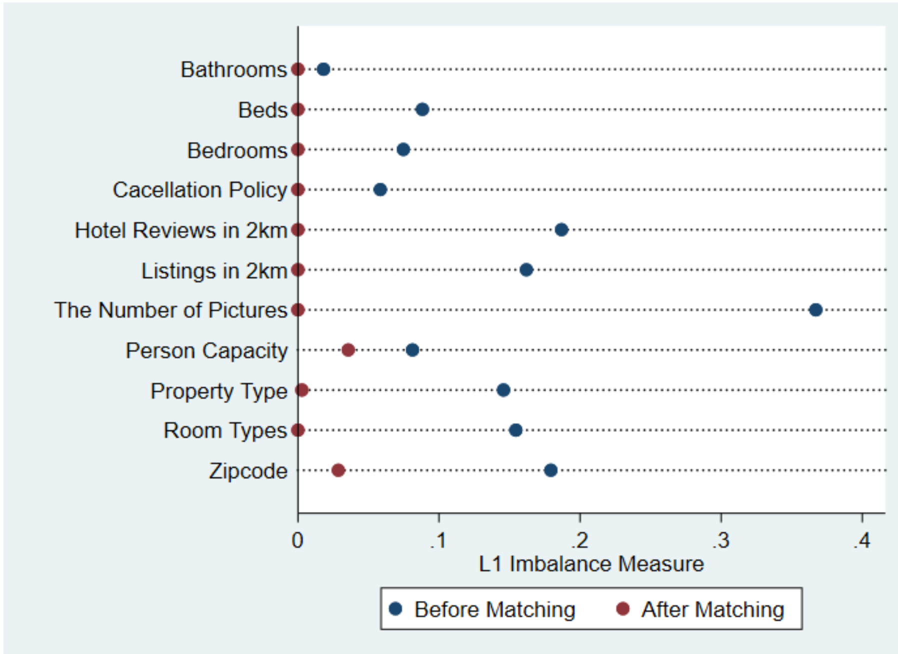

Figure 2.5 Reduction in Imbalance Due to CEM

The CEM method described above can mitigate the "selection on observables," however, it does not resolve potential endogeneity due to "selection on unobservables." According to Airbnb, listings need to satisfy certain requirements (e.g. attractive style, good design, well-equipped, well-maintained, and so on) to obtain the Plus certification. Researchers cannot directly observe these intangible factors, however, they are likely to be reflected in textual reviews from guests. To explicitly account for such unobserved qualitative features, we mine textual reviews from both Plus and non-Plus listings and feed the results into a classification model, which predicts the likelihood, or propensity, that a property would receive a Plus certification. The control group is constructed by matching the propensity score of the treated focal listing with that of one of the CEM-matched non-Plus listings with the closest propensity score.

To unpack these steps in more detail, we start by scraping textual review data for the listings in our data set. We then pre-process the data by removing stop-words and apply lemmatization. We compute tf-idf scores which enable us to rescale a set of features extracted from each listing's review data. To avoid overfitting, we randomly draw 2,236 listings from our unmatched CEM samples. After excluding the listings that have zero reviews, we split the CEM sample into training and testing subsamples in a 70:30 ratio, and train a logistic regression classification model. When applied to the test data, the classification model yields 75% accuracy, 74% precision, and 75% recall. We retrieve the probabilities of Plus classification, which can be interpreted as propensity scores, for both Plus and non-Plus listings. For each Plus listing in the treatment group we find the non-Plus listing in the matched CEM sample with the closest propensity score, and build out the control group that is matched 1:1 with the treatment group. In this way, we account for both observable characteristics and unobservable listing features embodied in textual reviews that might play a role in the listing's likelihood of obtaining Airbnb Plus certification. We assume that the acquisition of Airbnb Plus certification is orthogonal to any residual unobservables in booking rate.

Following previous studies (Agrawal and Goldfarb 2008; Zhang et al. 2017), we estimate the pre-treatment trends to check whether our models meet the parallel trends assumption, which is a key assumption for DD analysis. Since the treatment groups in our DD models receive treatment at different points in time, we normalize the time dimension as time periods before and after the treatment by assigning reverse negative integers to the pre-treatment periods, and positive integers in a sequential order to the post-treatment periods. For example, the normalized time periods look like as follows: (*n, ...,* -4,-3, -2, -1, 0, 1, 2, 3, 4, *..., m*) where 0 is the time period when the treatment takes place for a listing, and 1 is the second post-treatment time period. The specification for the relative time model is exactly the same as the one in Equation (1), except the $\beta_1 DD_{it}$ term is replaced by $\sum_{t=-n}^{n} \lambda_t Z_{it}$, where Z_{it} is the interaction term of the treatment indicator and the (pre-treat and post-treat) time dummy variable at time t and λ_t, is the corresponding coefficient. For example, Z(-1) can be expressed as *treatment*pre-treat*(-1) where *pre-treat*(-1) represents a dummy for one week prior to the treatment period. Therefore, every pre-treat and post-treat time dummy, *pre-treat*(-4)*, pre-treat*(-3)*, …, post-treat*(2) *, post-treat*(3) *, post-treat*(4)*, …* is interacted with the treatment variable so that we can identify the trends of outcomes before and after the treatment time period. Here, we consider an eight-week time interval — four weeks prior to the treatment period and four weeks posterior. We use *pre-treat(*-1*)* as the baseline period, the coefficient of which is normalized to zero. The externality model is similar, with the appropriate change in the interpretation of "treatment".

It is important to note that the coefficients for pre-treatment trends should not be statistically significant if the parallel assumption is to be satisfied in both direct effects and externality effects models. If the parallel trend assumption is met, we have confidence that there is no evidence of any differences in in the booking rate before the presence of Plus certification between the treated and control groups.

Results

We start with our baseline results for direct effect, followed by a number of additional analyses and robustness checks.

Table 2.3 shows the results for direct effect model (2.1) using the PSM-matched samples. As shown in the first column of the table, the coefficients of the key variable, *DD*, is positive and significant, indicating that receiving a Plus certification leads to a 7.6% increase in book rate. The estimated coefficients for the control variables have the signs and magnitudes one would expect; i.e., more reviews, and higher star rating are associated with higher booking rate, whereas price is negatively associated with booking rate.

Turning to the economic significance of these estimates, we calculate the revenue implications associated with the direct and externality effects. Starting with the former, the average increase in annual revenue associated with receiving a Plus certification amounts to \$3517.74 = 7.6% (direct effects coefficient) * \$218.64 (average price of a Plus listing pre-treatment) * 0.58 (average booking rate pre-treatment) * 365 (number of days in a year). These revenue impacts are nontrivial, suggesting that Plus certification is a credible signal of quality.

Table 2.3 Baseline Results for Direct Effect

	Direct Effect
DD	0.076*** (0.014)
ln(Num_Reviews)	0.101** (0.049)
ln(Num_Host_Reviews)	-0.006 (0.033)
Star_Rating	0.103* (0.060)
Superhost	-0.008 (0.033)
ln(Listing_Duration)	0.024 (0.115)
ln(Lag1_Price)	-0.107*** (0.037)
Constant	0.140 (0.707)
Observations	18,089
R-squared	0.629
Listing FE	Yes
Week*City FE	Yes

*Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$*
Standard errors are clustered by Listing ID.

A key identifying assumption of the Difference-in-Difference estimation strategy is the parallel trends assumption, in that there is no significant difference in booking rate in the pre-treatment period between listings that receive Plus certification, and those that do not. If this is true, then the average increase in booking rate for the treatment group can be taken to be a causal estimate of Plus certification on focal listings' revenues. As we discussed in the previous section the parallel trends assumption is verified using the relative time model, the results for which are presented in Table 2.4. We see that all of the pre-treatment coefficients are insignificant, indicating that the parallel trends assumption is indeed satisfied. At the same time, the post-treatment coefficients are all significantly positive, which empirically establishes our key identifying assumption.

Figure 2.6 shows the trend in booking rate before and after treatment. Specifically, t+0 on the x-axis represents the time period when listings become Plus. We normalize the coefficient in t-1 as our baseline so that we can interpret the coefficients in the other relative periods as the difference in booking rate between treatment and control groups compared to that in t-1. The solid line plots the coefficient estimates in our relative time model (Table 4) and the dashed lines show the 95% confidence interval for the corresponding coefficients. From this graph, we can confirm that the difference in booking rate between Plus and non-Plus listings is not significantly different from zero before treatment, whereas the difference becomes significant and positive after treatment. This clearly illustrates a significant impact of receiving a Plus certification on booking rates.

Table 2.4. Checking the Parallel Trends Assumption (Direct Effect)

	Direct Effect
Pre_treat(-4)	0.008 (0.013)
Pre_treat(-3)	0.001 (0.009)
Pre_treat(-2)	-0.006 (0.006)
Pre_treat(-1)	*Baseline* *Omitted*
Start_treat(0)	0.019*** (0.006)
Post_treat(1)	0.038*** (0.009)
Post_treat(2)	0.052*** (0.012)
Post_treat(3)	0.061*** (0.013)
Post_treat(4)	0.065*** (0.015)
ln(Num_Reviews)	0.103** (0.049)
ln(Num_Host_Reviews)	0.002 (0.033)
Star_Rating	0.106* (0.059)
Superhost	-0.008 (0.033)
ln(Listing_Duration)	0.039 (0.114)
ln(Lag1_Price)	-0.109*** (0.037)
Constant	0.026 (0.706)
Observations	18,089
R-squared	0.627
Listing FE	Yes
Week*City FE	Yes

*Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$*
Standard errors are clustered by Listing ID.

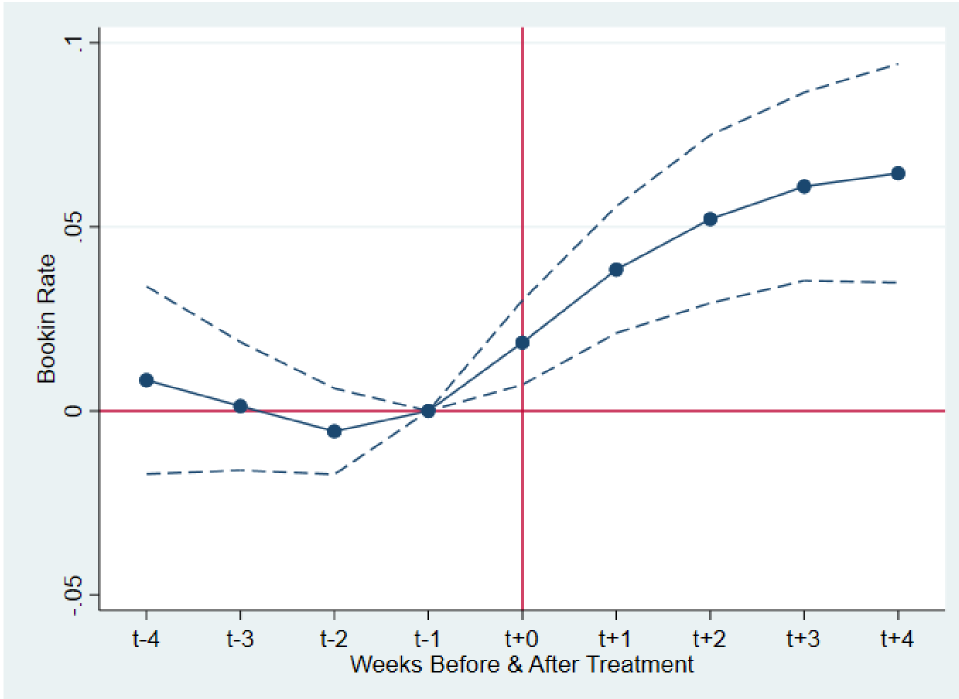

Figure 2.6 Trend in Booking Rate Before and After Plus Certification

Heterogeneous Treatment Effects

As is well known, online marketplaces are characterized by information asymmetry. This issue takes on greater importance on the Airbnb platform due to the nature of the service provided; i.e., it is an experience good, the value of which cannot be easily assessed before consumption (Nelson 1974). Extensive studies have investigated how price can enable firms to signal quality to imperfectly informed users (Milgrom and Roberts 1986), because only high-quality firms can afford to charge a high price and recoup the revenue loss due to the decreased demand for repeat purchase — if quality were actually low. If Airbnb Plus certification also serves as a signal of product quality, its impact should be smaller for properties with higher price. To investigate this mechanism and the heterogeneous impacts across different types of properties, we partition our sample into properties based on the median price and then estimate the interaction effects in our original DD specification.

Our second partition is motivated by the observation that Airbnb tends to be more appealing to casual travelers whereas hotels attract more business travelers who value hotel amenities, predictable and standardized quality, and convenience of location close to business or tourist centers. In areas with greater presence of hotels, Airbnb is probably less likely to be a popular option for prospective travelers. Therefore, properties receiving the Plus certification in these areas would not bring as much benefit as those in areas with lower density of hotels. We segment the properties by the density of hotels nearby, proxied by the number of hotel reviews on TripAdvisor in a 2km radius circle; we add an additional interaction term between DD and a dummy variable indicating a property with above-median nearby hotel reviews, High_hotel_reviews.

Besides price, Superhost can serve as another quality signal to consumers. Superhost is a status automatically granted by Airbnb if the property owner's account meets certain criteria.[4] If a property earns the Superhost status, a badge shows up in a prominent place next to the title of the property. In the third analysis, we split our sample into two subsamples. One subsample consists of properties with below-median price and without the Superhost status, whereas the other subsample consists of properties with above-median price and Superhost status. The latter group, given the presence of multiple quality signals, should have lower quality uncertainty and thereby less subject to the influence of Plus certification. The former group represents the group with highest quality uncertainty and Plus certification shall matter most to those properties.

[4] To earn the Superhost status, the property owner has to maintain a 90% response rate or higher, a 4.8 overall rating, a 1% cancellation rate or lower and complete at least 10 trips or 3 reservation that total at least 100 nights. (https://www.airbnb.com/help/article/829/how-do-i-become-a-superhost. Accessed on April 28th, 2020)

The results of the heterogeneous effects are presented in Table 2.5. The first column presents results of the model with an interaction effect between the DD variable and the dummy variable identifying listings with above-median price. We find that the coefficient on the interaction term, which captures the differential impact of Plus certification on the high-price properties compared to the low-price properties, is negative and statistically significant, in line with our intuition. In the second column, we interact the DD variable with

Table 2.5 Heterogeneous Treatment Effects

	Price	Hotel Reviews	Low Price & No Superhost	High Price & Superhost
DD	0.105*** (0.016)	0.094*** (0.017)	0.112*** (0.042)	0.025 (0.023)
DD*High Price	-0.065*** (0.018)			
DD*High_Hotel_Reviews		-0.027* (0.014)		
ln(Num_Reviews)	0.108** (0.049)	0.102** (0.048)	-0.012 (0.103)	0.183*** (0.063)
ln(Num_Host_Reviews)	-0.010 (0.034)	-0.008 (0.033)	0.098** (0.048)	-0.088 (0.105)
Star_Rating	0.109* (0.059)	0.098 (0.061)	0.090 (0.120)	0.186*** (0.067)
Superhost	-0.008 (0.033)	-0.009 (0.033)		
ln(Listing_Duration)	0.007 (0.113)	0.021 (0.114)	-0.364 (0.614)	-0.223 (0.346)
ln(Lag1_Price)	-0.100*** (0.037)	-0.105*** (0.037)	0.171 (0.385)	-0.140*** (0.045)
Constant	0.157 (0.694)	0.179 (0.701)	0.886 (3.972)	1.353 (1.818)
Observations	18,089	18,089	1,492	5,832
R-squared	0.630	0.629	0.660	0.648
Listing FE	Yes	Yes	Yes	Yes
Week*City FE	Yes	Yes	Yes	Yes

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Standard errors are clustered by Listing ID.

the dummy variable indicating the property has above-median hotel reviews in its neighborhood. As seen from the table, the coefficient on DD * High_Hotel_Reviews is negative and statistically significant, in line with our arguments above. In column 3 (Low Price & No Superhost) the DD variable is still positive and statistically significant; its magnitude is greater than what we saw in the baseline results of Table 3. The coefficient of the DD variable in Column 4 (High Price & Superhost) is no longer significant, as we argued above. Overall, we conclude that the Plus certification indeed brings an outsized benefit for properties that lack quality signals, but may not have provided added value for properties with already strong quality signals.

Robustness Checks

We used propensity score matching to generate our baseline results. As a robustness check we consider an alternative matching approach — Look-Ahead Matching (LAM), as also used in Bapna et al. (2018). The rationale behind this approach is that, those listings that are treated in an earlier period should be similar to those that receive treatment in a later period. The only difference between these two groups is the timing of treatment. Accordingly, LAM could mitigate the concern that treated listings are different from control listings with respect to unobservable characteristics.

To implement LAM, we define a new treatment group as listings which became Plus in the first half of our time window, and the new control group as listings which did not receive Plus certification until sometime in the second half of our time window. We further apply CEM to ensure that the treatment and control groups are similar in terms of observable listing and neighborhood characteristics. This LAM approach is extended to construct matching treatment and control groups for the externality model in a manner analogous to what we used in the PSM approach above.

Table 2.6 reports the estimation results using the LAM sample. As shown in the table, the coefficients on the *DD* variables are similar to those seen in Table 2.3. A Plus certification is associated with a 9.1% increase in booking ratio for the certified listing on average. These percentage effects are larger than those in Table 2.3, so our baseline estimates are relatively conservative.

Table 2.6 Robustness Check with Look Ahead Matching (Direct Effect)

	Direct Effect
DD	0.091^{*} (0.049)
ln(Num_Reviews)	-0.138 (0.128)
ln(Num_Host_Reviews)	0.194 (0.169)
Star_Rating	0.011 (0.137)
Superhost	-0.007 (0.055)
ln(Listing_Duration)	-0.256 (0.310)
ln(Lag1_Price)	-0.292^{***} (0.075)
Constant	2.834 (1.734)
Observations	2,701
R-squared	0.626
Listing FE	Yes
Week*City FE	Yes

Robust standard errors in parentheses *** *p<0.01,* ** *p<0.05,* * *p<0.1*
Standard errors are clustered by Listing ID.

To further strengthen the validity of our causal estimates, we implement a placebo test for the direct effect model. For our treatment group, we generate random pseudo treatment times and then assign them to the Plus listings. Every Plus listing has a random treatment time that is different from its actual treatment time. We estimate our DD models with the pseudo treatment times, the results for which should not be significant. This will also confirm that there are no other unobserved shocks that create similar effects on booking rate as does the acquisition of a Plus certification. Table 2.7 shows that, in the direct effect model, the coefficients of the placebo *DD* variables are not statistically significant, in either the direct or externality model. These results from the Placebo test further lend support for our causal claims.

Table 2.7 Robustness Check with Placebo Test (Direct Effect)

	Direct Effect
DD	0.002 (0.012)
ln(Num_Reviews)	0.112** (0.049)
ln(Num_Host_Reviews)	-0.000 (0.033)
Star_Rating	0.111* (0.059)
Superhost	-0.006 (0.032)
ln(Listing_Duration)	0.064 (0.114)
ln(Lag1_Price)	-0.104*** (0.037)
Constant	-0.140 (0.696)
Observations	18,089
R-squared	0.625
Listing FE	Yes
Week*City FE	Yes

*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*
Standard errors are clustered by Listing ID.

If users book in advance of their actual stay, then a property that was non-Plus at the time of booking may become Plus by the time they began their stay. Due to these lags, the demand impact of a Plus certification may not fully manifest itself for a week or two after certification. To address this concern, we consider scenarios in which users book rooms one or two weeks in advance, respectively. More specifically, we substitute our dependent variables with the one-week ahead and two-week ahead booking rates. The results shown in Table 2.8 (where we have included the baseline 0 Week-ahead results for the sake of easy comparison) indicate that the qualitative nature of our results are unchanged.

Table 2.8 Robustness Check for Advance Booking (Direct Effect)

	Direct Effect		
	0 Week Ahead (Baseline)	1 Week Ahead	2 Weeks Ahead
DD	0.076*** (0.014)	0.085*** (0.014)	0.082*** (0.015)
ln(Num_Reviews)	0.101** (0.049)	0.106** (0.048)	0.105** (0.046)
ln(Num_Host_Reviews)	-0.006 (0.033)	-0.032 (0.032)	-0.047 (0.030)
Star_Rating	0.103* (0.060)	0.120** (0.056)	0.111** (0.056)
Superhost	-0.008 (0.033)	-0.002 (0.030)	0.004 (0.028)
ln(Listing_Duration)	0.024 (0.115)	0.035 (0.123)	0.029 (0.126)
ln(Lag1_Price)	-0.107*** (0.037)	-0.112*** (0.042)	-0.094** (0.047)
Constant	0.140 (0.707)	0.112 (0.748)	0.181 (0.771)
Observations	18,089	17,194	16,302
R-squared	0.629	0.636	0.642
Listing FE	Yes	Yes	Yes
Week*City FE	Yes	Yes	Yes

*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*
Standard errors are clustered by Listing ID.

Besides a Plus badge in the listing, another purported benefit of a Plus certification is higher placement in search results (see, e.g., Lodgify 2020). If the search rank effect is strong enough then it could potentially confound the effects of quality signaling, since both factors should affect booking rate in the same direction. Also, our research design relies on listing level rather than search session data, and it is therefore not suited to separate the two effects. To address this concern, however, we compared the rankings of Plus and non-Plus listings in search results for the eight cities in our data set. Specifically, we looked at the first 200 impressions in search results (about 10 pages on the desktop) in each of the eight cities, and conducted a difference-of-means test of the Plus and non-Plus listings. We found that in seven out of the eight cities in our data set, the differences in average search rank of Plus and non-Plus listings are not statistically significant in a one-tailed t test. Only Austin presented a marginally significant difference ($p < 0.1$) in ranking. Accordingly, we dropped observations for Austin and reran our direct and externality effects models. We found that the results (omitted in this draft) were essentially identical to the baseline results of Table 3. We can conclude that our results are robust to the search rank issue.

CHAPTER 3

Quality Certification and Demand: Externality and Local Platform Effects

This chapter covers the extended view of the effect of quality certification on the non-certified properties and on the platform itself. Previous sharing economy studies have considered properties with high quality photos (Zhang et al. 2017) and stricter cancellation policy (Zalmanson et al. 2018) as ways to signal the product quality to customers. Given that they have focused on the effect of seller managed signals, we uncover the effect of quality certification that is purely managed by the platform. In particular, this chapter emphasizes its effect on the other stakeholders other than customers and Plus listings. Who benefits and who is hurt by the launch of Plus certification in Airbnb? What is the effect on the counterparts (non-Plus listings)? What is the net effect on the Airbnb platform? These are the questions we need to address. Our findings show that the presence of one or more Plus listings in a neighborhood negatively affects the demand of non-Plus listing by 1.5% in the same neighborhood. However, Airbnb receives increased net benefits by 1.5% on average by creating Plus program within the platform. We conclude that quality certification has differential impacts on Plus, non-Plus listings, and on the platform. Other sharing economy platforms need to cautiously approach designing the quality-tier products or services since they may consider the tradeoff between the expected gains they can achieve and the potential risk of negative externality from their products or services.

Identification Strategy: Externality and Local Platform Effects

One of the noticeable differences between the previous and the current chapter is that the previous chapter has focused on the role of quality certification on its own demand whereas this chapter examines the spillover effects of quality certification on the demand of non-certified properties and on the revenue of the platform. In other words, does the presence of Plus listings in a neighborhood cannibalize non-Plus listings' demand? Or does it actually benefit these non-Plus listings nearby by generating positive spillovers? For example, if competition effect is at work, we would expect to see that the presence of Plus listings within a certain distance of a focal non-Plus listing *i* will negatively affect that listing *i*'s booking rate. On the other hand, it may positively impact the focal non-Plus listing's booking rate if Plus listings either help draw more customers from the hotel industry to Airbnb, in general, or shifting some existing Airbnb customers to this area due to the better perceived quality.

We start with our models for estimating externality effect associated with Plus certification, followed by a discussion of our identification strategy to assure that our empirical estimates are causal. Similarly, the externality effect is estimated using the following two-way fixed effects framework for listing *i* in city *j* in time period *t*:

$$(Booking\_Rate)_{ijt} = \alpha + \beta_1 Externality\_DD_{it} + \beta_2 ln(Num\_Reviews)_{it} + \beta_3 ln(Num\_Host\_Reviews)_{it} + \beta_4 (Star\_Rating)_{it} + \beta_5 (Superhost)_{it} + \beta_6 ln(Listing\_Duration)_{it} + \beta_7 ln(Price)_{it-1} + X_i + V_{jt} + \varepsilon_{ijt} \quad (3.1)$$

The identification strategy that we have adopted to address the externality effect model is different from that in direct effect model. In the externality effects model, each focal listing i in the treatment group is a non-Plus listing, but in this case the indicator variable DD_{it} keeps track of whether or not one or more of the neighboring listings (within a 2 km range) has received a Plus certification. That is, treatment in the direct effects model happens when the focal listing itself receives a Plus certification, whereas in the externality model, treatment corresponds to at least one of the other listings in a 2km range receiving a Plus certification. Both the direct and externality models include a control group of listings, which we will describe later.

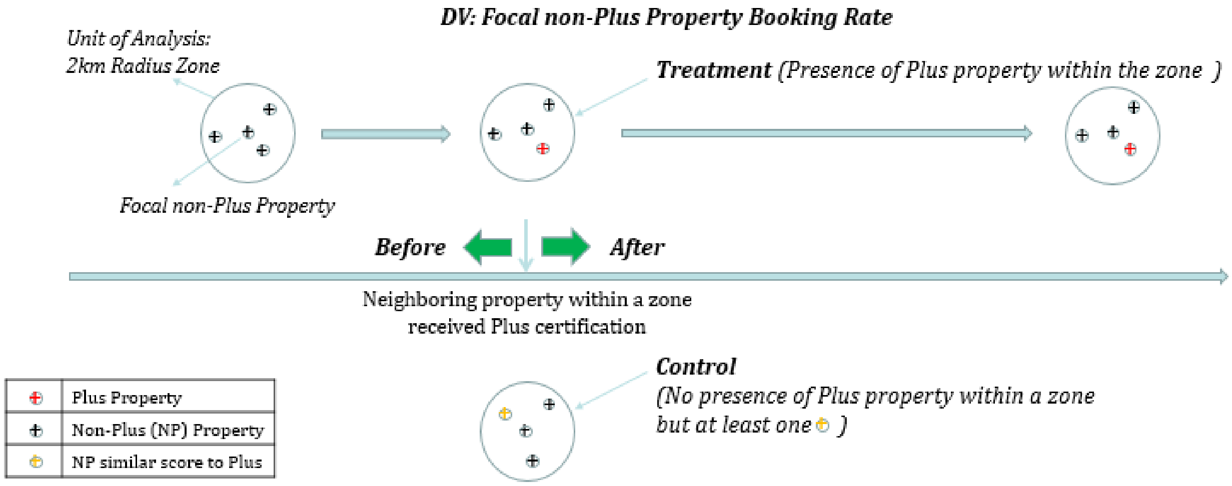

Figure 3.1 Treatment and Control Units in the Externality Effect Model

The construction of a matched control group for the externality effects model (see the second panel of Figure 4) is along the same lines, but with a few differences. Note that in the externality effects model, the treatment group consists of non-Plus listings that have one or more Plus listings in a 2km radius range. We construct the corresponding control group in two steps. We first apply CEM on the basis of listing and neighborhood characteristics to compile an initial set of candidate control listings. We then narrow that set using PSM to construct the final control group of listings, that do not have any Plus listings within a 2 km range of the focal listing — but one or more of them are classified as a Plus listing by the machine learning model. Note that we exclude focal listings in the control group if none of the listings in the 2 km range are predicted to be "Plus-like."

Results

Table 3.1 shows the results for externality effects models (3.1) using the PSM-matched samples. We find that the presence of one or more Plus listings within a 2km neighborhood reduces a focal non-Plus listing's booking rate by 1.5%, on average, consistent with a negative externality effect of Plus certification. The estimated coefficients for the control variables have the signs and magnitudes one would expect; i.e., more reviews, and higher star rating are associated with higher booking rate, whereas price is negatively associated with booking rate.

Turning to the economic significance of these estimates, the externality impact on a non-Plus listing of having a Plus listing nearby (i.e., within a 2km zone) is -$670.56 = -1.5% (externality effects coefficient) *218.71 (average price of non-Plus listing) * 0.56 (average booking rate of non-Plus listing) * 365 (number of days in a year).

Table 3.1 Baseline Results for Externality Effect

	Externality Effect
DD	-0.015* (0.009)
ln(Num_Reviews)	0.123*** (0.022)
ln(Num_Host_Reviews)	0.002 (0.017)
Star_Rating	0.042 (0.029)
Superhost	-0.009 (0.011)
ln(Listing_Duration)	-0.158*** (0.048)
ln(Lag1_Price)	-0.020 (0.034)
Constant	0.837*** (0.299)
Observations	81,394
R-squared	0.659
Listing FE	Yes
Week*City FE	Yes

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Standard errors are clustered by Listing ID.

The results for parallel trend assumption are presented in Table 3.2.

Table 3.2 Checking the Parallel Trends Assumption (Externality Effect)

	Externality Effect
Pre_treat(-4)	0.003 (0.010)
Pre_treat(-3)	0.010 (0.007)
Pre_treat(-2)	0.005 (0.005)
Pre_treat(-1)	*Baseline* *Omitted*
Start_treat(0)	-0.009* (0.005)
Post_treat(1)	-0.017** (0.008)
Post_treat(2)	-0.020** (0.009)
Post_treat(3)	-0.020** (0.010)
Post_treat(4)	-0.010 (0.011)
ln(Num_Reviews)	0.123*** (0.022)
ln(Num_Host_Reviews)	0.002 (0.017)
Star_Rating	0.042 (0.029)
Superhost	-0.009 (0.011)
ln(Listing_Duration)	-0.158*** (0.048)
ln(Lag1_Price)	-0.020 (0.034)
Constant	0.834*** (0.299)
Observations	81,394
R-squared	0.660
Listing FE	Yes
Week*City FE	Yes

*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*
Standard errors are clustered by Listing ID.

Robustness Checks

This LAM approach is extended to construct matching treatment and control groups for the externality model in a manner analogous to what we used in the PSM approach above. Table 3.3 reports the estimation results using the LAM sample. As shown in the table, the coefficients on the *DD* variables are similar to those seen in Table 3.1. A Plus leads to a 3.7% decrease in the booking ratio of neighboring non-Plus listings. These percentage effects are larger than those in Table 3.1, so our baseline estimates are relatively conservative.

Table 3.3 Robustness Check with Look Ahead Matching (Externality Effect)

	Externality Effect
DD	-0.037* (0.021)
ln(Num_Reviews)	0.142*** (0.046)
ln(Num_Host_Reviews)	-0.066 (0.048)
Star_Rating	0.073 (0.053)
Superhost	0.011 (0.022)
ln(Listing_Duration)	-0.166 (0.152)
ln(Lag1_Price)	-0.012 (0.064)
Constant	0.914 (0.819)
Observations	14,758
R-squared	0.668
Listing FE	Yes
Week*City FE	Yes

*Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1*
Standard errors are clustered by Listing ID.

Also, Table 3.4 shows that ,for the externality effects models, the coefficients of the placebo *DD* variables are not statistically significant, in either the direct or externality model. These results from the Placebo test further lend support for our causal claims.

Table 3.4 Robustness Check with Placebo Test (Externality Effect)

	Externality Effect
DD	0.005 (0.007)
ln(Num_Reviews)	0.124^{***} (0.022)
ln(Num_Host_Reviews)	0.002 (0.017)
Star_Rating	0.041 (0.029)
Superhost	-0.009 (0.011)
ln(Listing_Duration)	-0.160^{***} (0.048)
ln(Lag1_Price)	-0.020 (0.034)
Constant	0.841^{***} (0.300)
Observations	81,394
R-squared	0.659
Listing FE	Yes
Week*City FE	Yes

Robust standard errors in parentheses *** *p<0.01,* ** *p<0.05,* * *p<0.1*
Standard errors are clustered by Listing ID.

If users book in advance of their actual stay, then a property that was non-Plus at the time of booking may become Plus by the time they began their stay. Due to these lags, the demand impact of a Plus certification may not fully manifest itself for a week or two after certification. To address this concern, we consider scenarios in which users book rooms one or two weeks in advance, respectively. More specifically, we substitute our dependent variables with the one-week ahead and two-week ahead booking rates. The results shown in Table 3.5 (where we have included the baseline 0 Week-ahead results for the sake of easy comparison) indicate that the qualitative nature of our results are unchanged.

Table 3.5 Robustness Check for Advance Booking (Externality Effect)

	Externality Effect		
	0 Week Ahead (Baseline)	1 Week Ahead	2 Weeks Ahead
DD	-0.015* (0.009)	-0.015* (0.009)	-0.013 (0.009)
ln(Num_Reviews)	0.123*** (0.022)	0.135*** (0.022)	0.127*** (0.022)
ln(Num_Host_Reviews)	0.002 (0.017)	0.000 (0.016)	0.007 (0.016)
Star_Rating	0.042 (0.029)	0.039 (0.028)	0.032 (0.027)
Superhost	-0.009 (0.011)	-0.009 (0.011)	-0.009 (0.011)
ln(Listing_Duration)	-0.158*** (0.048)	-0.186*** (0.049)	-0.205*** (0.051)
ln(Lag1_Price)	-0.020 (0.034)	-0.038 (0.033)	-0.071** (0.034)
Constant	0.837*** (0.299)	1.045*** (0.302)	1.318*** (0.312)
Observations	81,394	76,971	72,569
R-squared	0.659	0.667	0.675
Listing FE	Yes	Yes	Yes
Week*City FE	Yes	Yes	Yes

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Standard errors are clustered by Listing ID.

We have seen that a Plus certification has a positive impact on the booking rate of listings receiving the certification, but a negative impact on other non-Plus listings nearby. This raises the key question of how these effects balance out in terms of the net impact on platform revenues. To address this question, we again adopt a Difference-in-Difference specification, that is a bit different from our baseline direct and externality effects models. Specifically, we use a circle-level analysis where the goal is to determine whether the Plus certification of one or more listings in a 2km circle results in a higher or lower local (i.e., within the circle) net revenues for the Airbnb platform. We measure net revenue effects at the granularity of a restricted local region (i.e., 2km radius circle). This micro level analysis is appropriate given that Plus listings only account for a small portion of the total listings in our sample, and competition is likely to be most intense for those listings in close geographic proximity to each other. The outcome variable is $ln(Avg\_Revenue)_{ijt}$, measuring the average weekly revenue for all the listings that are located within a 2km-radius range of the focal listing *i* in City *j* and Week *t*. The covariates used to explain this outcome variable are also aggregated to the circle-week level, resulting in the following specification.

$$ln(Avg\_Revenue)_{ijt} = \alpha + \beta_1 LP\_DD_{it} + \beta_2 ln(Avg\_Num\_Reviews)_{it} + \beta_3 ln(Avg\_Num\_Host\_Reviews)_{it} + \beta_4 (Avg\_Star\_Rating)_{it} + \beta_5 (Avg\_Superhost)_{it} + \beta_6 ln(Avg\_Listing\_Duration)_{it} + X_i + V_{jt} + \varepsilon_{ijt} \tag{3.2}$$

Here LP_DD_{it} is a dummy variable which equals to 1 if there is at least one Plus listing within a 2 km radius range of Listing *i* (in City *j*) at time *t;* and 0 otherwise. $ln(Avg\_Revenue)_{ijt}$ is the average revenue of all the listings that are located within a 2 km radius range of *i* at time *t*.

To properly identify the impact of Plus listings on platform revenue, we use matched treatment and control circles, where a circle is always defined as a 2 km circle around some focal listing *i*. As before, we first match the two groups on observable listing and neighborhood characteristics, however, these variables are calculated as 2km circle averages. For example, one of our matching covariates is the average number of bedrooms for all listings within the circle. After CEM on observables we conduct PSM based on textual reviews, along the lines of our externality model; that is, we restrict our final control group to those 2km circles in which at least one or more non-Plus listings have similar propensity scores to those of the Plus listings in the corresponding treatment circle. The results of the circle level analysis for local platform effects are displayed in Table 9, which shows that the appearance of Plus listings in a circle results in a net revenue increase of 1.5% for the Airbnb platform in that circle.

Table 3.6 Local Platform Effect

	Local Platform Effect
DD	0.015^{***} (0.004)
ln(Avg_Num_Reviews)	-0.646^{***} (0.041)
ln(Avg_Num_Host_Reviews)	-0.060^{***} (0.013)
Avg_Star_Rating	0.800^{***} (0.100)
Avg_Superhost	0.082^{*} (0.047)
ln(Avg_Listing_Duration)	-0.030 (0.097)
Constant	3.612^{***} (0.620)
Observations	114,322
R-squared	0.973
Listing FE	Yes
Week*City FE	Yes

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Standard errors are clustered by Listing ID.

Translating the percentage effect to absolute dollars, we note that the introduction of Airbnb Plus listings in a neighborhood generates additional net revenues of $80,702.20 = 1.5%(Local Platform Effects Coefficient) * $218.64(average daily price) * 0.58 (average daily booking rate) * 365 (number of days in a year) * 726.48 (average number of listings in a 2km circle) * 16% (Airbnb Service Fee 3% from hosts and 13% from guests) per circle, on average.[5] We can conclude that the initial evidence on the launch of the Airbnb Plus service generates positive net revenues for the platform, and presumably helps the platform to more effectively compete against the hotel industry.

[5] https://www.airbnb.com/help/article/1857/what-is-the-airbnb-service-fee

CHAPTER 4

Impact of Simultaneous Bilateral Ratings on Volume, Valence, and Diversity of Reviews: Evidence from Airbnb

This chapter analyzes the impact of a change in a review system from asynchronous to a simultaneous regime, using a quasi-experimental approach. Specifically, our research design uses a difference-in-difference approach that exploits the fact that Airbnb changed its review system from asynchronous to simultaneous in July 2014, whereas TripAdvisor remained asynchronous throughout. Applying this identification strategy to a set of properties cross-listed on both platforms, we find some significant causal impacts of the review system change on volume, valence, and diversity of listing reviews on Airbnb. Compared to the prior asynchronous regime, we find that after the change to simultaneous reviews: guests are more likely to leave a review, resulting in a 12% increase in review volume per unit time; on average, the valence of reviews reduces by 0.13-0.17 on a scale of 1-5, consistent with a reduction in "review inflation"; the sentiment of reviews is less positive following the change in review system; and the reviews are less ambiguous and focused around a smaller set of topics. These results are consistent with the notion that under a simultaneous review regime, guests are less concerned about retaliation effects, and thereby engage more frequently and honestly with the review system. Our findings have significant implications for research and practice dealing with review systems design, especially with respect to bilateral review systems common in the sharing economy.

Literature Review

In this section we discuss the closest papers to ours, and describe how our work is differentiated. A number of recent studies have examined the impact of a policy change from asynchronous to simultaneous double-blind reviews. Bolton et al. (2012) conducted lab studies in a simulated eBay platform to examine the impact of such a change. They showed that the policy change would result in relatively more negative feedback (as dissatisfied users would no longer be dissuaded to share their negative feedback). They also showed that feedback frequency (likelihood of a user to leave feedback) would go down, due to the fact that satisfied users have a reduced incentive to post a review, as the positive review cannot result in a reciprocal positive review from the other party, in a simultaneous review system. Fradkin et al. (2021) examine the same policy change (i.e., asynchronous to simultaneous) in a field study on the Airbnb platform using a randomized experimentation approach. They report that under the simultaneous review system the valence of reviews becomes less positive, but the magnitude of the effect is small, and the policy change does not result in a reduction in adverse selection either. In contrast to Bolton et al. (2012), however, they find that feedback frequency goes up, not down. They attribute this last result to a "desire to unveil" counter-party ratings; i.e., when a user is informed that the counter-party has left a review, they are motivated to submit their own review quickly in order to uncover the review left about them. They claim that the lab setting of the Bolton et al. (2012) study does not replicate this "desire to unveil" and therefore generates the opposite result with respect to feedback frequency.

Another study that examined the policy change by Airbnb is Mousavi and Zhang (2018), but they use an observational study design applied to data from the platform before and after the policy change. Specifically, this study used a Regression Discontinuity Design and panel data analysis to examine the impact of the policy change on feedback valence and frequency. They also use textual analysis and topic modeling to examine if the policy change influenced information content (review depth, topic coverage, etc) and sentiment in the posted reviews. They find that the policy change led to less positive reviews that were also lengthier and more objective in their personal opinions. The study did not examine the impact of the policy change on feedback frequency.

While there is a consensus in prior work that a shift from asynchronous to simultaneous review system would result in less positive reviews, there are mixed results with respect to the magnitude and direction of the impact on feedback frequency. Also, with the exception of Mousavi and Zhang (2018) there is little that we know about the impact of the policy change on the content and nature of the feedback itself. This study hopes to fill the gap in several ways. First of all, similar with Mousavi and Zhang (2018), we adopt an observational quasi-experimental approach that examines before and after data surrounding Airbnb's policy change from asynchronous to simultaneous review in July 2014. Unlike the previous study, however, we examine cross-listed properties from two short-term rental platforms, Airbnb and Trip Advisor, in order to control for unobserved heterogeneity that is listing-specific and time-varying. In other words, we compare the reviews from Airbnb to those from Trip Advisor for the exact same properties that are listed on both platforms. Since all the reviews on Airbnb platform are affected by the policy change, we identify the comparable reviews, not influenced by the Airbnb's new review policy, for the same properties that are also listed on the other platform, Trip Advisor. Comparing reviews within each same property naturally eliminates the possibility of biased results coming from unobservables at a property level. Second, our study overcomes the issue of limited time and scope of randomized experiments by leveraging observational data. While both studies have employed rigorous experiment settings, they had somewhat opposing results to each other. One of the possible reasons might be that experimental approaches choose relatively a short period of time for their experiments which can lead to different outcomes. Also, many experimental studies are context-specific which are at a risk of generalizability issue. Therefore, we take the advantage of using observational data which allows us to capture the actual effect of simultaneous review policy on the outcome changes. It also gives a possibility for us to estimate the causal effect, on average, which covers wider range of a time window before and after the policy change. Lastly, we hope to add to the evidence on the causal impact of the policy change on feedback valence, frequency, sentiment and diversity. Our results additionally include embedded semantic structures of textual reviews using detailed analysis of topic modeling approach and the combination of sentiment analysis and topic modeling. Although, Mousavi and Zhang (2018) worked on these, they have provided limited findings such as changes in the number of topics and etc. Our study takes one step further to derive out which topics or contents are mentioned by reviews and the changes in review valence in which topics. Recent IS literature have adopted a topic modeling approach to study the information contents that are semantically hidden in e-commerce reviews (Khern-am-nuai et al. 2018), posts and replies in a microblogging website (Geva et al. 2019), reviews in online healthcare website (Saifee et al. 2020), and online crowdsourcing communities (Hwang et al. 2019). Motivated by the existing studies, we uncover the qualitative dimensions of customer reviews and examine which and to what extent dimensions are affected by the simultaneous review policy. However, we take the different approach in terms of using topic modeling. Specifically, we employ one of the semi-supervised approaches, Anchored Correlation Explanation (Anchored CorEx), which is far more efficient in dealing with the issue of uninterpretable topics by merging multiple micro-grained topics into a more general single topic and by effectively separating mixed overlapping themes into distinct sets of topics.

Data and Matching Procedures

We collect customer reviews and property characteristics for all the available listings from both Airbnb and TripAdvisor platforms in multiple cities in U.S. The list of the cities includes Los Angeles, New York City, Broward County, Austin, Washington D.C. and New Orleans, Destin, Gatlinburg, Kauai, Kissimmee, Miramar, Orange Beach, Panama Beach, Pigeon Forge, and Sevierville. These cities have enough number of vacation rental properties on both platforms which allows us to identify the properties that are listed on both platforms. We explain our unique matching procedure to obtain the exact cross-listed properties from both platforms as listed below.

First of all, for each same city, match Airbnb and TripAdvisor properties based on host names and keep all the possible candidate pairs. Second, based on the possible candidate pairs from Step 1, match based on the number of bedrooms and bathrooms, and preserve all the possible candidate pairs which have the same numbers of bedrooms and bathrooms. Third, based on the possible candidate pairs from Step 2, we identify closest distance candidate Airbnb property to the focal TripAdvisor property by calculating the square distance of longitude and latitude of both properties. Lastly, based on the final set of candidate pairs from Step 3, we leverage the image similarity index, and assign similarity scores to candidate pairs. Above the threshold of 1.06, the pairs should be the exact cross-listed properties. Below the threshold of 1.03, the pairs are less likely to be the exact cross-listed properties. The candidate pairs within the range of 1.03 to 1.06, we manually check whether a candidate pair is the exactly the same properties.

Given the various numbers of images uploaded by hosts, we design a procedure to identify the maximal bound of similarity by iteratively comparing one image *i* from a set of images A from Airbnb property *m* and one image *j* from a set of images *T* from TripAdvisor property *n*.

$$s_{i,j}^{m,n} \coloneqq Sim_{i \in A, j \in T}(i, j) \tag{4.1}$$

Then the maximal similarity index for the pair of property *m* from Airbnb and the property *n* from TripAdvisor is defined as:

$$S^{m,n} \coloneqq \max\left(0, s_{i,j}^{m,n}\right), \forall i \in A, \forall j \in T \tag{4.2}$$

We assign $S^{m,n}$ to the candidate pair of properties m from Airbnb and n from TripAdvisor to the candidate pool from step 3 and we denote $Sim(\cdot,\cdot)$ as the image similarity function we choose, but in principle it could be other image similarity functions. Given this similarity measure, we do not require the number of images from Airbnb property m and the number of images from TripAdvisor property n to be the same. In addition, we can also tolerate the irrelevant pictures uploaded by the hosts, including local restaurants and scene. Note that hosts might enter Airbnb and TripAdvisor at different time points, they would upload pictures to both platforms with various angles, quality, amount and size. In this scenario, the average similarity or minimum similarity between the set of pictures on Airbnb and the set of pictures on TripAdvisor, only depends on management efforts of hosts, other than listing characteristics. Only the maximum value of picture sets similarity $S^{m,n}$can help to determine probability of exactly same listing across Airbnb and TripAdvisor.

We choose host names, the number of bedrooms and bathrooms, latitude and longitude, and pictures as the matching scheme criteria for the following reasons. First, hosts are usually required to verify government IDs and provide their actual names on their profiles. We anticipate that first name of host for a given cross-listed property on Airbnb and TripAdvisor are consistent to each other. Second, the same property should exhibit strictly equal number of bedrooms and bathrooms even if they are listed on different platforms. Third, the property location is always fixed, we also expect that the geographic coordinates of cross-listed properties on Airbnb and TripAdvisor should be the same. Lastly, we further compare the property pictures uploaded by hosts. We believe that same hosts of the same properties post same or similar set of property pictures on the two platforms. Following these four-step matching algorithm, we could successfully detect cross-listed properties on Airbnb and TripAdvisor that are exactly the same properties. Figure 4.1 shows a cross-listed properties on both Airbnb and TripAdvisor. On Airbnb, the property name is "*Loft @ Hollywood Vine w/ Rooftop Lounge 31+ Days*", however, the name of same property that are listed on TripAdvisor is "*Loft Living at Hollywood Vine - 31+ Night Rental*". Also, the daily price on Airbnb is $96 whereas the price is listed $145 on TripAdvisor. In addition, some pictures are in different angles and colors even though they depict the same objects and features. However, our unique matching procedure is robust and efficient in terms of classifying these pictures as same pictures.

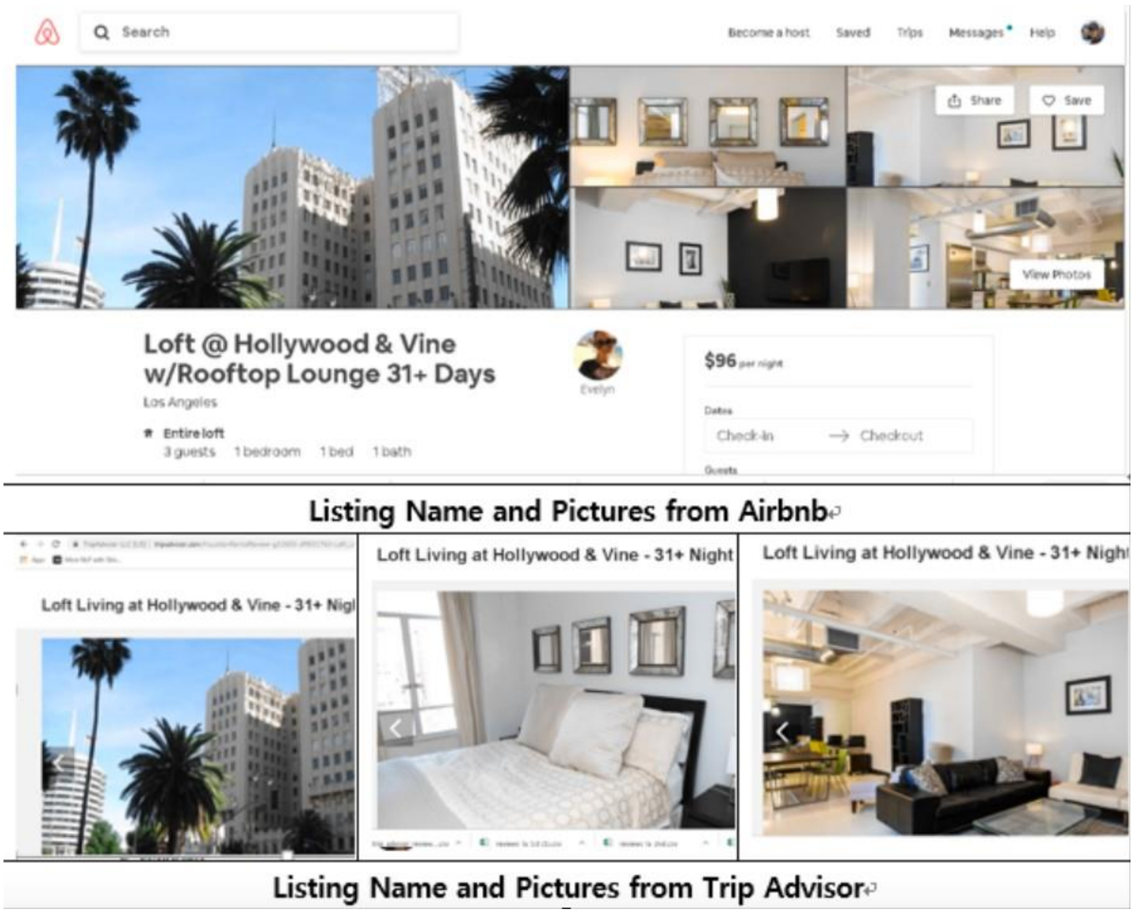

Figure 4.1 Cross-Listed Property on Airbnb and Trip Advisor

Variable Descriptions

After we go over the steps identifying the cross-listed properties, we retrieve 25,486 reviews for the 758 cross-listed properties. We could obtain textual reviews, rating, and other property and host characteristics such as the number of pictures for each property, review length counted by the number of words, the first date when a property was registered, and so on. We also construct a panel data at property-monthly level for our analysis. To describe the detailed definitions of the variables. R*ating* is defined as a cumulative average score that a property has received by guests from Airbnb and TripAdvisor in a given month. *No. Pictures* represents the number of pictures that a property has posted on Airbnb or TripAdvisor.

Reviewer Activeness means the total number of reviews a reviewer (guest) posted on each platform. In other words, it can be interpreted as the intensity of reviewer activity or participation in writing online reviews. For example, if a reviewer has posted many reviews, this reviewer tends to actively use the platform and adopt a higher standard on evaluation a new listing. *Reviewers Tenure* indicates how many months has passed since a reviewer joined on each platform. *Host Tenure* is constructed in a similar manner. It is defined as the number of months since a host has joined each platform. *Review Length* is the total number of words counted from each review posted on each platform. We believe that a reviewer is more likely to give longer reviews if she or he has experienced an extremely satisfactory or unsatisfactory service from the property and we include this measure to control for any factors that would confound our analysis.

Our treatment group and control group are defined as the property-monthly level reviews that belongs to either Airbnb or TripAdvisor, respectively. Note that these reviews are from the exact same property which is posted on both platforms. In other words, the treatment group includes reviews come from Airbnb for a property and the control group includes reviews from TripAdvisor for the same property. Post treatment is coded as 1 if the review posted date is after July 10th, 2014 and is 0 otherwise. Given that our samples (Exact same properties) already rule out property related counterfactuals, we would not need to implement any matching methods to construct comparable property groups. Therefore, we only incorporate extra covariates to increase estimation preciseness.

Table 4.1 and Table 4.2 show the summary statistics of our data set. We can see that Rating for the treatment group decreases from 4.85 to 4.79 after the introduction of the new review policy from Airbnb. The Rating difference between treatment and control groups in pre-treatment period is 0.23 whereas the value in post-treatment period becomes 0.06.

Table 4.1 Summary Statistics for Treatment and Control Group (All Years)

		Treatment Group			Control Group		
Variable	Pre/Post	Obs	Mean	SD	Obs	Mean	SD
Rating	Pre	631	4.85	0.44	541	4.62	0.72
No. of Pictures		631	26.73	22.14	541	34.70	14.08
Review Length		631	85.87	72.82	541	406.92	435.26
Host Tenure		631	498.65	297.45	541	920.26	594.66
Reviewer Tenure		630	240.27	310.00	273	643.96	558.67
Reviewer Activeness		630	13.67	52.58	273	17.55	22.09
Communication with Host		631	0.22	0.41	541	0.02	0.15
Location		631	0.73	0.43	541	0.20	0.39
Equipment		631	0.31	0.46	541	0.27	0.44
Cleanliness		631	0.75	0.43	541	0.48	0.50
Check-in Experience		631	0.65	0.46	541	0.40	0.47
Recommendation		631	0.56	0.50	541	0.21	0.41
Rating	Post	20,098	4.79	0.59	4,216	4.73	0.68
No. of Pictures		20,095	31.66	19.74	4,556	33.57	13.52
Review Length		20,098	107.51	177.72	4,217	322.07	348.00
Host Tenure		20,095	1,158.24	651.64	4,212	829.81	643.47
Reviewer Tenure		19,971	711.49	641.91	1,198	1,169.39	1,218.81
Reviewer Activeness		19,971	8.58	36.78	1,264	14.44	57.69
Communication with Host		20,098	0.22	0.41	4,204	0.08	0.27
Location		20,098	0.51	0.49	4,204	0.14	0.34
Equipment		20,098	0.27	0.44	4,204	0.19	0.39
Cleanliness		20,098	0.64	0.48	4,204	0.45	0.50
Check-in Experience		20,098	0.49	0.48	4,204	0.24	0.40
Recommendation		20,098	0.41	0.49	4,204	0.18	0.39

We combine these to hypothesize that the value of Airbnb Rating might have decreased in respect to that of TripAdvisor Rating after the Airbnb's review policy change. To link the causal relationship between the review policy change and the Rating change, we need to construct the causal model to verify the effect of review policy change on customers' rating and review behavior. We also incorporate the summary statistics for another set of samples. Table 4.2 shows a similar pattern compared to Table 4.1. Note that samples from Table 4.2 only include the reviews and rating before and after 3 years of policy change in order to construct a more balanced sample data set.

Table 4.2 Summary Statistics for Treatment and Control Group (-/+3 Years)

		Treatment Group			Control Group		
Variable	Pre/Post	Obs	Mean	SD	Obs	Mean	SD
	Pre						
Rating		611	4.85	0.44	408	4.68	0.66
No. of Pictures		611	26.10	18.85	408	34.96	14.45
Review Length		611	86.08	73.49	408	372.78	427.07
Host Tenure		611	510.22	294.87	408	1045.27	603.99
Reviewer Tenure		610	245.42	312.85	214	735.53	585.73
Reviewer Activeness		610	13.58	53.30	214	18.71	24.28
Communication with Host		611	0.22	0.42	408	0.02	0.15
Location		611	0.73	0.43	408	0.24	0.42
Equipment		611	0.31	0.46	408	0.28	0.45
Cleanliness		611	0.75	0.43	408	0.49	0.50
Check-in Experience		611	0.65	0.46	408	0.42	0.48
Recommendation		611	0.56	0.50	408	0.23	0.42
	Post						
Rating		5,655	4.79	0.57	960	4.71	0.68
No. of Pictures		5,655	31.00	23.67	960	32.91	14.61
Review Length		5,655	93.91	151.92	960	327.30	364.24
Host Tenure		5,655	947.24	570.34	960	636.69	643.74
Reviewer Tenure		5,633	513.41	503.98	472	989.83	1,101.51
Reviewer Activeness		5,633	11.96	53.07	472	15.24	37.81
Communication with Host		5,655	0.26	0.44	960	0.10	0.30
Location		5,655	0.63	0.47	960	0.27	0.43
Equipment		5,655	0.33	0.47	960	0.27	0.44
Cleanliness		5,655	0.72	0.45	960	0.53	0.50
Check-in Experience		5,655	0.62	0.47	960	0.33	0.44
Recommendation		5,655	0.49	0.50	960	0.24	0.43

Model-Free Evidence

Figure 4.2 plots the rating distributions before and after the policy change for both platforms. The left histogram depicts the rating distribution for Airbnb and TripAdvisor in pre-treatment period whereas he right one shows the rating distribution (for Airbnb and TripAdvisor) in post-treatment period. Both histograms prove that most of the reviews tend to leave high ratings (above 4 and 5). However, we can clearly notice that the portion of 5 star rating is decreased and the portion of 1-4 star rating is increased in the post-treatment period with respect to the pre-treatment period. Based on the result, we may provide an evidence that the simultaneous review system changed a way guests review their stays. However, we need to carefully interpret the result since model free evidence lacks causal inference.

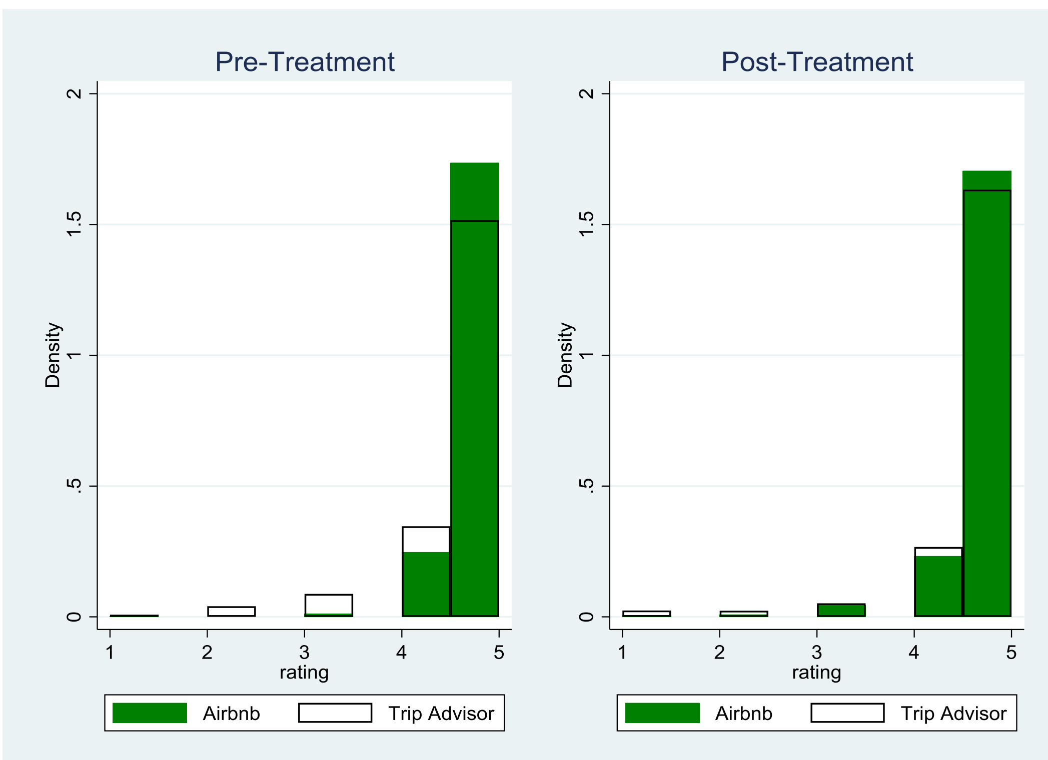

Figure 4.2 Rating Distribution for Airbnb and Trip Advisor

Empirical Models

Our research design employs the Difference-in-Differences (DID) identification, as shown in Equation (4.3). Note that it is also described using conceptual framework in Figure 4.3 below.

$$Rating_{it} = \beta_1 * After_t + \beta_2 * Airbnb_i + \beta_3 * After_t * Airbnb_i + \beta_4 * X_{it} + \varepsilon_{it} \qquad (4.3)$$

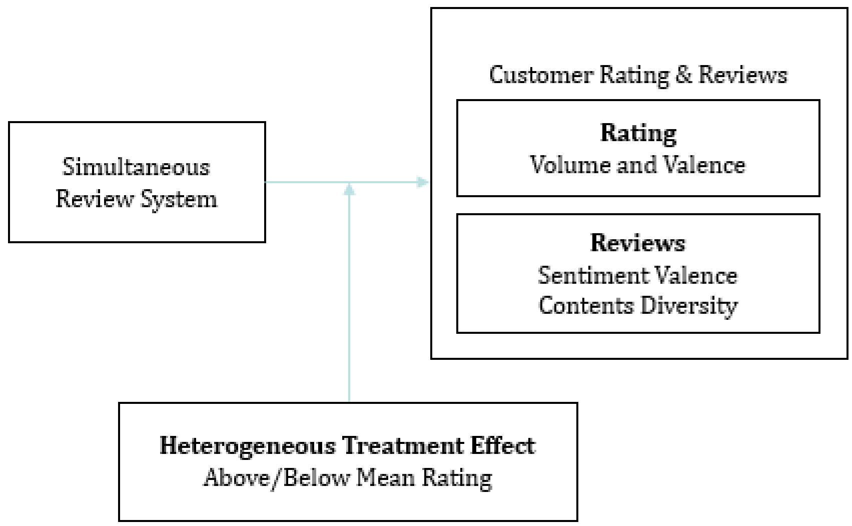

Figure 4.3 Conceptual Framework (Simultaneous Review System)

In Equation (4.3), β_3 is the parameter of interest which indicates the change of cumulative average rating from Airbnb compared to that from TripAdvisor for the exact same properties after the simultaneous review policy. β_1 captures the effects of policy change in terms of time and β_2 denotes the mean differences in rating between Airbnb and TripAdvisor for the same properties. Even with exact cross-listed properties, we still control the pre-determined covariates at property and reviewer levels which would not be affected by the treatment, to better fulfill the assumptions of DID estimator. In an econometric perspective, controlling extra pre-determined covariates would be helpful to increase estimation precision, and in this way, it is more likely to fulfill the assumption of conditional independence. However, we would also to assess the robustness of our DID estimator, by comparing the estimation with and without control variables X_{it} in Table 4.3.

Table 4.3 Main Results (All Years)

	Average Rating	
Airbnb	0.237*** (0.058)	0.185*** (0.060)
After	0.103* (0.056)	0.065 (0.060)
Airbnb*After	-0.171*** (0.060)	-0.129** (0.065)
No. of Pictures (Log)	0.194 (0.169)	0.044** (0.017)
Review Length (Log)	0.011 (0.137)	-0.018*** (0.006)
Reviewer Activeness (Log)	-0.007 (0.055)	0.029*** (0.006)
Host Tenure (Log)	-0.256 (0.310)	-0.009 (0.008)
Reviewer Tenure (Log)	-0.292*** (0.075)	0.005 (0.005)
Constant	2.834 (1.734)	4.583*** (0.108)
Observations	25,469	22,051
R-squared	0.005	0.009

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

We additionally test our DID model using different sets of samples. Table 4.3 shows the results from the samples in different time length. First column adopts the samples 3 years before and after the policy change. We constrain the samples in a similar manner (4 and 5 years before and after the policy change) and the corresponding results are shown in the second and third columns. Even if we limit the samples to generate the equal length of pre and post treatment period, we still obtain the consistent results compared to the main results in Table 4.2. Therefore, we claim that the change of Airbnb review policy has decreased the rating, on average, for Airbnb properties compared to the rating for TripAdvisor properties.

Table 4.4 Main Results (-/+3 Years)

	-/+ 3 Years	-/+ 4 Years	-/+ 5 Years
Airbnb	0.129** (0.051)	0.173*** (0.063)	0.179*** (0.062)
After	0.053 (0.052)	0.065 (0.062)	0.068 (0.061)
Airbnb*After	-0.130** (0.059)	-0.141** (0.067)	-0.140** (0.066)
No. of Pictures (Log)	-0.018 (0.027)	0.017 (0.023)	0.038** (0.019)
Review Length (Log)	-0.040*** (0.011)	-0.035*** (0.008)	-0.022*** (0.006)
Reviewer Activeness (Log)	0.031*** (0.009)	0.025*** (0.007)	0.030*** (0.006)
Host Tenure (Log)	-0.004 (0.012)	-0.003 (0.010)	-0.007 (0.009)
Reviewer Tenure (Log)	0.012 (0.008)	0.014* (0.007)	0.007 (0.005)
Constant	4.862*** (0.159)	4.686*** (0.136)	4.609*** (0.117)
Observations	6,928	12,140	19,131
R-squared	0.015	0.012	0.009

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

In order to obtain a valid estimate, the control group and treatment group would need to satisfy the parallel trends assumption, and constant treatment effects at β_3 . Before the change of review system policy, the cumulative average rating between treatment group and control group should not be observed with significant differences. We test this assumption using relative time models with and without covariates showing in Equation (4.4). After the policy change, we anticipate to gradually capture the deviation of cumulative average rating between treatment group and control group. We test the parallel trend assumption by running a relative time model, following the below Equation (4.4):

$$Rating_{it} = \sum_{r=-6}^{-1} \beta_t * I[t = r] * Airbnb_i + \sum_{r=0}^{+6} \beta_t * I[t = r] * Airbnb_i + \varepsilon_{it} \quad (4.4)$$

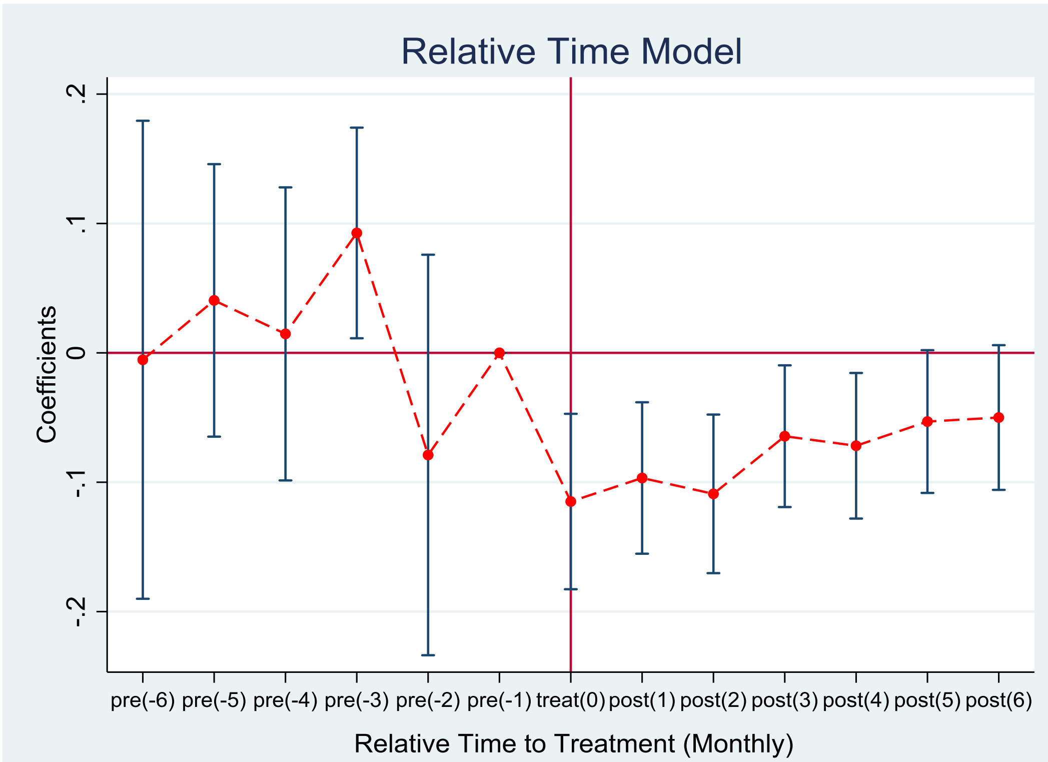

Figure 4.4. Visualization of Relative Time Model

From Equation (4.4), we require the cumulative average ratings of cross-listed properties between Airbnb and TripAdvisor are statistically indifferent during pre-treatment period. *r* represents the relative time from the policy implementation month July 2014: $r = -2$ denotes 2 months before the policy change and $r = +2$ stands for 2 months after the policy implementation. We report our estimation result in Figure 4.4 and Table 4.5, while the coefficient of $r = -1$ is omitted as a benchmark. We do not see any significant difference between the cumulative average rating in Airbnb compared to that in TripAdvisor before the

Table 4.5 Parallel Time Trends

	Average Rating	Average Rating
Airbnb	0.092*** (0.018)	0.091*** (0.025)
After	0.023 (0.035)	-0.004 (0.028)
pre-treat(-6)	0.047 (0.099)	-0.007 (0.094)
pre-treat(-5)	0.089 (0.061)	0.039 (0.054)
pre-treat(-4)	0.053 (0.065)	0.013 (0.058)
pre-treat(-3)	0.139*** (0.049)	0.091** (0.041)
pre-treat(-2)	-0.034 (0.081)	-0.080 (0.079)
pre-treat(-1)	*Baseline omitted*	*Baseline omitted*
treat(0)	-0.101*** (0.033)	-0.115*** (0.035)
post-treat(1)	-0.088*** (0.029)	-0.096*** (0.030)
post-treat(2)	-0.106*** (0.031)	-0.109*** (0.031)
post-treat(3)	-0.058** (0.027)	-0.064** (0.028)
post-treat(4)	-0.065** (0.029)	-0.072** (0.029)
post-treat(5)	-0.050* (0.028)	-0.053* (0.028)
post-treat(6)	-0.049* (0.029)	-0.050* (0.028)
No. of Pictures (Log)		0.042** (0.017)
Review Length (Log)		-0.015*** (0.006)
Reviewer Activeness (Log)		0.029*** (0.006)
Host Tenure (Log)		-0.024*** (0.009)
Reviewer Tenure (Log)		0.003 (0.005)
Constant	4.699*** (0.037)	4.747*** (0.090)
Observations	25,469	22,051
R-squared	0.007	0.012

Robust standard errors in parentheses

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

review system change. In other words, almost all the $\beta_{(-)}$ coefficients are insignificant during the pre-treatment period with and without control variables. However, after the policy implementation, the $\beta_{(+)}$ coefficients turn negative which confirms that the cumulative average rating in Airbnb significantly decreases in terms of that in TripAdvisor. We strongly claim that our DID analysis does not suffer from the violation of parallel trend assumption.

Heterogeneous Treatment Effects

Since most of the properties in Airbnb platform have received high ratings, one can conclude that there is a limited variation in terms of rating (Zervas et al. 2021). Although this might be true, it would be interesting to find out how the introduction of the new policy differently influences properties with different average rating. To do so, we further investigate whether the treatment effect on rating varies under different rating scores. For example, properties with high rating can be more affected by the simultaneous review system since those properties have received mostly "good" ratings. It would be a good opportunity to find out if this new review policy is more effective at reducing the degree of rating of the properties which already have accumulated extremely positive rating. In order to measure this, we split our samples into two different groups (listings with above/below mean rating) and run the same analysis over the two group separately. Table 4.6 shows that our main variable of interest, $After \times Airbnb$, is not significant for the below mean rating samples. On the other hand, the amount of rating is significantly decreased by 0.7-0.08 after the policy implementation for the above mean rated properties. These results lead us to interpret how the review policy change has a differential impact on the properties with different rating. As mentioned by Zervas et al. (2015), most Airbnb properties have received extremely high rating and this would blur the potential customers to identify the actual quality of the properties, our results give us a critical insight to conclude simultaneous review system works in a way that significantly decrease the excessive high rating of properties. If we see a reduction in below mean rating properties, this would generate an issue of exit of those properties from the platform. In other words, the low rated properties may be discouraged to be affected by the policy change and, in turn, exit the platform (e.g. close the service on the platform). This raises a reduction of seller side of the platform and would influence the potential matchings between the pool of buyers and sellers in the future.

Table 4.6 Heterogeneous Treatment Effects (Sub-Sample Analysis)

	Below Mean Rating		Above Mean Rating	
	Full Sample	-/+ 3 Years	Full Sample	-/+ 3 Years
Airbnb	0.289** (0.113)	0.211 (0.129)	0.032* (0.018)	0.024 (0.019)
After	-0.452*** (0.108)	-0.255* (0.138)	0.054*** (0.017)	0.042** (0.019)
Airbnb*After	0.190 (0.130)	0.048 (0.158)	-0.081*** (0.019)	-0.068*** (0.021)
No. of Pictures (Log)	0.044 (0.040)	-0.110** (0.054)	0.007 (0.006)	-0.006 (0.009)
Review Length (Log)	-0.077*** (0.014)	-0.106*** (0.024)	-0.004** (0.002)	-0.008*** (0.003)
Reviewer Activeness (Log)	0.106*** (0.019)	0.111*** (0.023)	-0.001 (0.002)	-0.002 (0.003)
Host Tenure (Log)	0.056** (0.026)	0.050 (0.033)	-0.007** (0.003)	-0.005 (0.004)
Reviewer Tenure (Log)	0.032* (0.018)	0.023 (0.027)	0.002 (0.001)	0.003 (0.003)
Constant	3.403*** (0.242)	4.164*** (0.294)	4.963*** (0.034)	5.012*** (0.052)
Observations	3,342	1,113	18,709	5,815
R-squared	0.063	0.080	0.005	0.007

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

Textual Review Analysis

We have focused on the effect of review policy change on guests' rating behavior. Prior literature regarding customer reviews included not only rating scores but also written reviews in many online platforms. Analyzing only the star rating does not fully capture the underlying customer behavior. Leveraging the customer (guest) textual reviews, we enrich our empirical analysis in order to examine how the policy change has affected the valence of the reviews. We turn our attention to the change of review sentiment, specificity, and on which subjects/topics customer mention when they are affected by the new review policy. We further combine the sentiment analysis with the topic modeling approach to find out how the valence of the reviews has changed for each topic over time.

We tokenize all the reviews in our data set and count the number of each token to construct the word frequency measure. To examine how the valence of the reviews has changed after the policy change, we count positive and negative terms for each document (review) and subtract the number of negative terms from the number of positive terms to generate net positive term frequency. We use the pre-trained dictionary to determine whether a term is classified as positive or negative. If the net positive term frequency is positive, a review contains higher number of positive terms than the number of negative terms. If it is negative, a review includes more negative terms than positive terms. Table 4.7 shows the following result. For both the full and balanced (3 years before/after the policy) data sets, net positive term frequency decreased in Airbnb reviews compared to that in TripAdvisor reviews after the simultaneous review policy has implemented. In other words, the valence of guest reviews became negative showing that guests evaluated their stays less positively after the policy change. To further support this result, we implement sentiment analysis. Numerous past studies in IS and marketing have analyzed sentiment of customer reviews and blog posts (Deng et al. 2018; Homburg et al. 2015; Luo et al. 2013; Rhue and Sundararajan 2019). We adopt Valence Aware Dictionary and sEntiment Reasoner (VADER) which is a lexicon and rule-based approach suited for social media product reviews (Hutto and Gilbert 2014). It generates the "compound score", a metric which calculates the sum of all the lexicon ratings which are normalized between -1 (extremely negative) and +1 (extremely positive).

Table 4.7 Word Frequency (Positive-Negative)

	Full Sample	-/+ 3 Years
Airbnb	6.752*** (1.095)	8.703*** (1.383)
After	3.620*** (0.719)	5.283*** (1.036)
Airbnb*After	-6.236*** (1.041)	-6.533*** (1.225)
No. of Pictures (Log)	-0.176 (0.452)	0.299 (0.778)
Review Length (Log)	-0.041 (0.117)	0.942*** (0.259)
Reviewer Activeness (Log)	1.112*** (0.148)	0.871*** (0.187)
Host Tenure (Log)	-0.166 (0.223)	0.422 (0.433)
Reviewer Tenure (Log)	0.316*** (0.069)	0.387*** (0.103)
Constant	1.012 (2.430)	-9.854** (4.954)
Observations	22,051	6,928
R-squared	0.048	0.063

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

Table 4.8 shows the sentiment analysis result. We find that sentiment score is decreased after the policy change, which is consistent with the findings from the net positive term frequency analysis.

Table 4.8 Sentiment and Entropy Scores

	Sentiment		Entropy	
	Full Sample	-/+3 Years	Full Sample	-/+3 Years
Airbnb	0.232*** (0.040)	0.236*** (0.039)	0.004 (0.031)	0.066* (0.037)
After	0.151*** (0.039)	0.171*** (0.040)	-0.016 (0.022)	0.025 (0.029)
Airbnb*After	-0.200*** (0.041)	-0.194*** (0.042)	-0.050 (0.033)	-0.071* (0.038)
No. of Pictures (Log)	-0.006 (0.008)	-0.008 (0.011)	-0.030** (0.013)	-0.001 (0.020)
Review Length (Log)	-0.002 (0.003)	0.008 (0.005)	0.054*** (0.004)	0.082*** (0.007)
Reviewer Activeness (Log)	0.026*** (0.002)	0.019*** (0.003)	0.010*** (0.004)	-0.001 (0.005)
Host Tenure (Log)	-0.002 (0.004)	0.003 (0.005)	0.011** (0.005)	0.025*** (0.007)
Reviewer Tenure (Log)	0.012*** (0.002)	0.008*** (0.003)	0.015*** (0.002)	0.010*** (0.003)
Constant	0.651*** (0.055)	0.618*** (0.068)	0.081 (0.068)	-0.231** (0.103)
Observations	22,051	6,928	22,051	6,928
R-squared	0.031	0.040	0.048	0.077

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

This is also aligned with the reduction in star rating but we could further examine the change of the tone of reviews. We also investigate the effect of review policy change on the ambiguity of reviews. The emotional entropy measures the amount of conflicting valence, which the values indicate unpredictability and surprise based on the inconsistency of the emotional language in a given string. We can say the reviews with higher value of emotional entropy are expressing mixed messages. From the third and fourth columns in Table 4.8, our results suggest that after the change of Airbnb policy, the reviews are more precise and less ambiguous in terms of valence and consistency. And we interpret the decline of emotional entropy as that reviewers are specifying the recommendation or drawback for the facilities or services.

We are motivated to not only understand the valence of textual reviews but also discover the latent semantic structure of a large set of customer reviews. Specifically, we attempt to uncover the embedded topics in the review corpus to identify on what aspects guests mention on both platforms and how these are different by the platforms and how these change over time. We leverage Anchored Correlation Explanation (Anchored CorEx) to extract the hidden topics of the reviews. CorEx is another approach that can be applied to topic modeling which does not require generative models (Gallagher et al. 2017). Probabilistic generative models, such as Latent Dirichlet Allocation (LDA), often involve specific assumptions and heavy computational costs due to parameter tuning. However, CorEx releases the generative model assumptions by learning topics without realizing the pre-existence of those topics. In other words, the model searches for the topics which give the maximum information regarding a set of documents based on the information theoretical framework. One of the noticeable differences between CorEx and other generative models is that CorEx can efficiently deal with the issue of uninterpretable topics by effectively splitting apart a topic with mixed multiple themes embedded into distinct sets of topics. In a similar way, it can collapse multiple granular topics into more general single topic. Also, CorEx considers the mutual information between documents and topics. In other words, it allows us to figure out to what extent each document is associated with the set of topics. Similar with other topic models, CorEx assigns the probability of each topic for a document to understand how much each document belongs to the each topic. Note that the sum of these probability can go beyond 1. For example, review A might have 0.8 of probability for topic 1 and 0.6 of probability for topic 2 at the same time. This enables us to investigate which review mentions which topics with different weights. We apply CorEx approach to derive the topics from the customer reviews. Figure 4.5 shows the top 30 words that are frequently mentioned from the review corpus. We can see that location and neighborhood related, property related, and positive words (great, enjoy, beautiful, and etc) are frequently shown within the reviews.

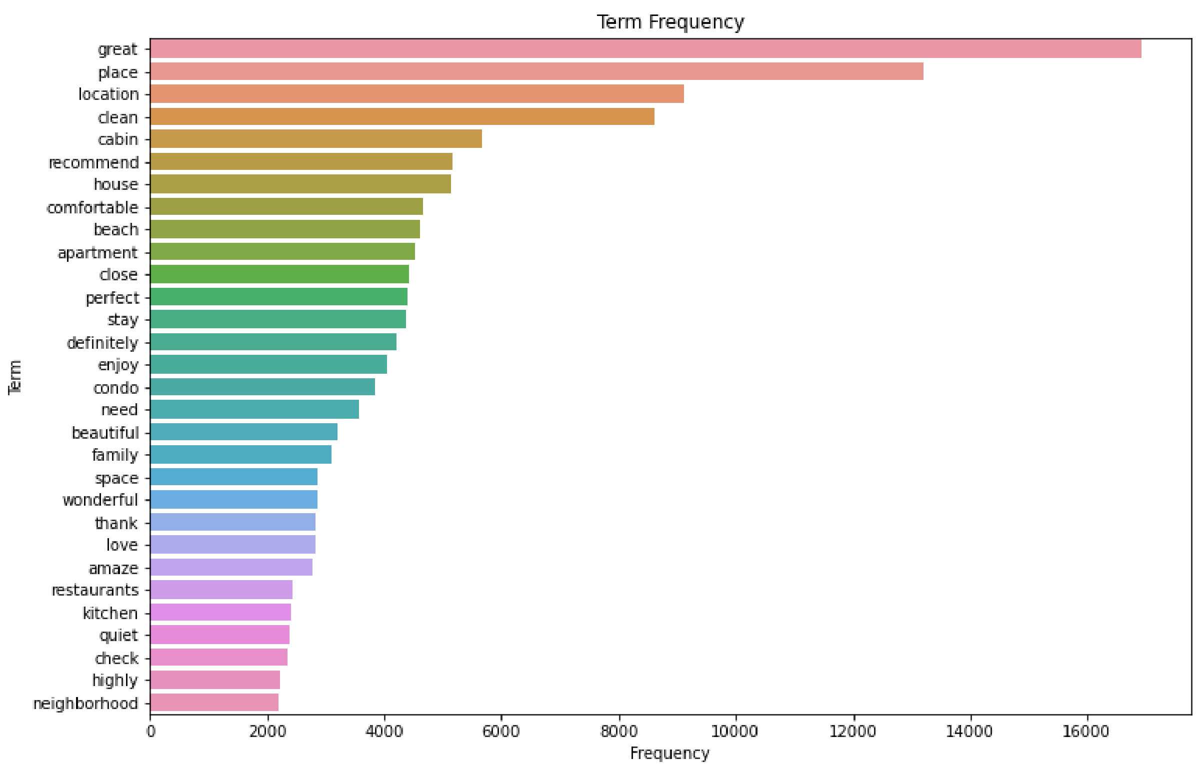

Figure 4.5 Top 30 Term Frequency

Since this model-free evidence only gives a simple overview of the word distribution over the corpus, we need to implement model-based strategy to identify the actual underlying topics that are embedded in the reviews. Table 4.8 shows our Anchored topic modeling analysis.

Table 4.9 Identified Topics and Top 10 Corresponding Keywords

Topics	Anchored Words	Top 10 Keywords
Communication with Host	Question, respond	Question, respond, quick respond, answer question, answer, quick, respond question, respond quickly, quickly, message
Location	Distance, Neighborhood	Neighborhood, distance, walk, walk distance, restaurants, apartment, street, place, quiet, quarter
Equipment and Amenities	Kitchen, bathroom, bedroom	Kitchen, bedroom, bathroom, dryer, washer dryer, washer, stock, live, towel, fully
Cleanliness	Clean, comfortable	Clean, comfortable, spacious, place clean, clean comfortable, cabin clean, house clean, condo clean, super clean, clean spacious
Check-in Experience	Check, arrive	Check, arrive, stay ,provide, need, night, experience, thank, little, look
Recommendation	recommend, highly	Recommend, highly, highly recommend, recommend place, recommend stay, friends, property manager, spotless, rental, condition

CorEx works as a semi-supervised topic modeling approach in a way that it enables us to choose "anchor words" that might represent potential topics that the algorithm had missed or had to consider earlier. After running CorEx model on our corpus and identifying the topics and the corresponding keywords, we choose a few keywords as anchors that we expect these anchors to identify more keywords that are more relevant to the labeled topics. As a result, we conclude 6 topics which represents the review corpus. The topic "Communication with Host" can be interpreted as the topic regarding hosts whereas the rest of the topics, "Location", "Equipment and Amenities", "Cleanliness", "Check-in Experience", and "Recommendation" represent the topic for property aspects. We choose six topics as the optimal number of topics based on the total correlation measures. We let the number of topics begins from two to ten and calculate the total correlation for each case. We stop the calculation at six because the accumulated total correlation does not significantly increase after six. These topics also mostly align with the Airbnb's and TripAdvisor's sub-rating categories. Therefore, we determine that these six topics generally represent what guests mention in their reviews. Figure 4.6 depicts keyword probabilities for each topic. Top 10 keywords in Table 4.8 correspond to the keywords in x-axis of each plot. The only difference is that each keyword has its own probability which explains how important each keyword is to a topic. Using the topic modeling results, we focus on whether customers discuss more or less number of topics on their reviews after the policy change. Specifically, we examine on which topic dimensions customers discuss more or less by measuring the probability of each topic assignment to each document. Note that multiple topics can be assigned to each document at the same time and the sum of these probabilities can be greater than 1 (different from the keywords probabilities to each topic which need to be summed up to 1 as mentioned in the previous paragraph). Table 4.9 shows the result when the outcome variable is the number of topics per document. It means how many topic related contents are described in each document (review).

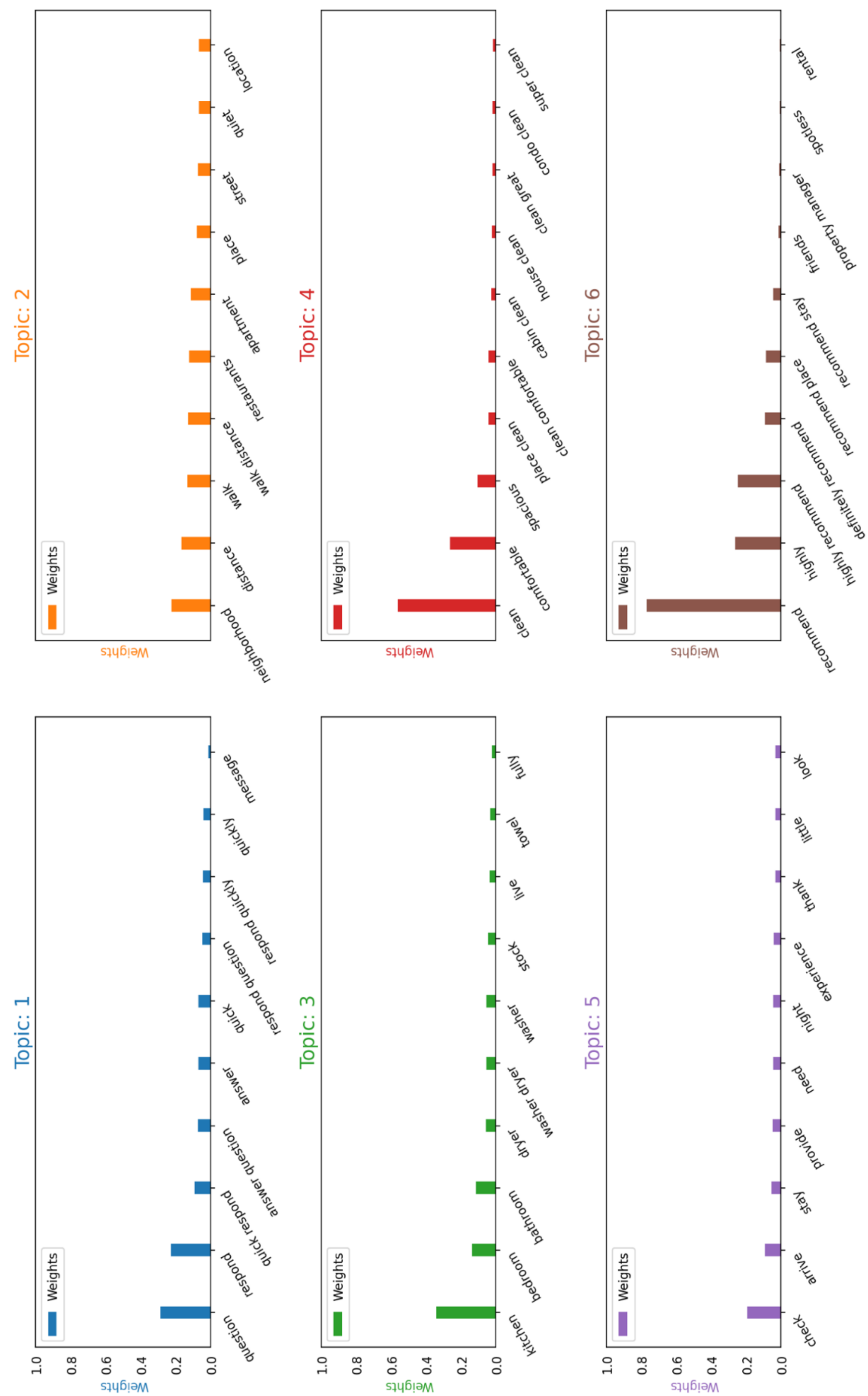

Figure 4.6 Keyword Probabilities (Weights) per Topic

Each document has the probability of each topic assignment and each probability is coded as 1 if the probability is greater than 0.5 and 0 otherwise. For example, review A has 6 different probabilities for all six topics. If the probability of topic 1 and 2 are 0.8 and 0.7 respectively, and the probability of topic 3, 4, 5, and 6 are 0.3, 0.1, 0, and 0.4 respectively, then the topic 1 and 2 are coded as 1 whereas the rest of the others are coded as 0 since these do not meet the threshold 0.5. Therefore, the number of topics for review A is 2 since the dominant topics that are discussed in review A are topic 1 and 2.

Table 4.10 Results for the Number of Topics per Document

	Number of Topics			
	Diff-in-Diff Output		Poisson Output	
	Full Sample	-/+3 Years	Full Sample	-/+3 Years
Airbnb	2.049*** (0.292)	2.622*** (0.335)	1.017*** (0.140)	1.192*** (0.164)
After	0.677*** (0.175)	1.187*** (0.237)	0.430*** (0.119)	0.634*** (0.148)
Airbnb*After	-1.383*** (0.281)	-1.499*** (0.327)	-0.692*** (0.140)	-0.743*** (0.167)
No. of Pictures (Log)	-0.182 (0.111)	0.014 (0.141)	-0.074 (0.046)	0.005 (0.048)
Review Length (Log)	-0.108*** (0.035)	0.209*** (0.065)	-0.044*** (0.014)	0.075*** (0.022)
Reviewer Activeness (Log)	0.320*** (0.033)	0.211*** (0.040)	0.120*** (0.011)	0.070*** (0.013)
Host Tenure (Log)	-0.023 (0.058)	0.162* (0.088)	-0.010 (0.024)	0.059* (0.031)
Reviewer Tenure (Log)	0.142*** (0.020)	0.104*** (0.029)	0.066*** (0.010)	0.039*** (0.011)
Constant	0.991* (0.593)	-2.230*** (0.841)	0.056 (0.256)	-1.046*** (0.314)
Observations	22,051	6,928	22,051	6,928

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

We find that first and second columns show negative coefficients for DID term which we interpret that customers mention fewer number of topics, tend to bring up narrower topics when they leave reviews. We confirm that the result is consistent when we run the Poisson model with the same data set. Poisson model is an appropriate identification strategy for modeling count outcome variable. In our contexts, the outcome variable is the number of topics, therefore, we adopt the Poisson model and find the similar results in terms of coefficient sign and magnitude.

Table 4.11 DV as the Topic Diversity Index

	Topic Diversity	
	Full Sample	-/+ 3 Years
Airbnb	0.755*** (0.098)	0.950*** (0.113)
After	0.259*** (0.064)	0.440*** (0.087)
Airbnb*After	-0.520*** (0.097)	-0.557*** (0.115)
No. of Pictures (Log)	-0.067* (0.039)	-0.004 (0.048)
Review Length (Log)	-0.045*** (0.013)	0.071*** (0.023)
Reviewer Activeness (Log)	0.112*** (0.011)	0.068*** (0.013)
Host Tenure (Log)	-0.008 (0.020)	0.055* (0.029)
Reviewer Tenure (Log)	0.051*** (0.007)	0.038*** (0.010)
Constant	0.266 (0.204)	-0.841*** (0.282)
Observations	22,051	6,928
R-squared	0.075	0.088

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

Similarly, we construct a topic diversity index using entropy diversity function so that we can measure variety of topics within each review. The diversity index is higher as a particular review mentions much contents regarding many topics whereas the index value is lower if another review only discusses contents that are related to fewer number of topics. In other words, a review receives a low value of diversity index if that review focuses on specific topics only. Table 4.10 shows the corresponding results. As you can see, the topic diversity index is reduced with respect to that of TripAdvisor reviews after Airbnb has changed the review policy which is consistent with the results from Table 4.9. If the variety of topics mentioned in the reviews is reduced, on which topic customers talked less?

Table 4.12 Seemingly Unrelated Regressions

	Probability Weight for Each Topic					
	Communication	Location	Equipment	Cleanliness	Check-in	Recommendation
Airbnb	0.198*** (0.030)	0.561*** (0.034)	0.211*** (0.032)	0.358*** (0.035)	0.365*** (0.035)	0.349*** (0.036)
After	0.084*** (0.027)	0.182*** (0.031)	0.115*** (0.029)	0.176*** (0.032)	0.042 (0.032)	0.087*** (0.033)
Airbnb*After	-0.077** (0.032)	-0.416*** (0.037)	-0.164*** (0.034)	-0.289*** (0.038)	-0.206*** (0.037)	-0.235*** (0.038)
No. of Pictures (Log)	-0.007 (0.006)	-0.080*** (0.006)	-0.032*** (0.006)	-0.020*** (0.007)	-0.040*** (0.007)	-0.002 (0.007)
Review Length (Log)	-0.003 (0.003)	-0.069*** (0.003)	0.034*** (0.003)	-0.034*** (0.003)	-0.001 (0.003)	-0.035*** (0.003)
Reviewer Activeness (Log)	0.023*** (0.003)	0.080*** (0.004)	0.042*** (0.003)	0.049*** (0.004)	0.069*** (0.004)	0.057*** (0.004)
Host Tenure (Log)	-0.010*** (0.003)	0.007* (0.004)	0.003 (0.004)	-0.007 (0.004)	-0.005 (0.004)	-0.011*** (0.004)
Reviewer Tenure (Log)	0.007*** (0.002)	0.034*** (0.003)	0.023*** (0.002)	0.023*** (0.003)	0.033*** (0.003)	0.021*** (0.003)
Constant	0.034 (0.043)	0.344*** (0.049)	-0.161*** (0.046)	0.424*** (0.050)	0.150*** (0.050)	0.204*** (0.051)
Observations	22,051	22,051	22,051	22,051	22,051	22,051
R-squared	0.012	0.101	0.024	0.031	0.044	0.034

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

In order to address this, we let the probability of each topic assigned to each review as an outcome variable. In this case, there are six different outcomes from six different topics and we employ Seemingly Unrelated Regression to estimate six equations simultaneously due to the concern of error term correlations. By observing that distribution of the probability of topic distribution across documents shows bimodal, depicting there are no values in the middle but the bins are toward at the end of both sides. We can interpret that many topic probabilities can be either nearly 0 and 1.

Table 4.13 Seemingly Unrelated Regressions (Logit)

	Seemingly Unrelated Logit					
	Communication	Location	Equipment	Cleanliness	Check-in	Recommendation
Airbnb	2.696*** (0.537)	3.048*** (0.521)	1.180*** (0.278)	1.567*** (0.245)	1.598*** (0.297)	1.636*** (0.275)
After	1.747*** (0.474)	1.269*** (0.441)	0.622*** (0.198)	0.745*** (0.182)	0.185 (0.217)	0.476*** (0.179)
Airbnb*After	-1.716*** (0.548)	-2.362*** (0.515)	-0.878*** (0.284)	-1.281*** (0.252)	-0.876*** (0.288)	-1.095*** (0.277)
No. of Pictures (Log)	-0.048 (0.080)	-0.348*** (0.127)	-0.171 (0.118)	-0.086 (0.096)	-0.170 (0.109)	-0.008 (0.085)
Review Length (Log)	-0.019 (0.027)	-0.300*** (0.043)	0.185*** (0.028)	-0.150*** (0.033)	-0.008 (0.037)	-0.149*** (0.027)
Reviewer Activeness (Log)	0.140*** (0.037)	0.359*** (0.038)	0.210*** (0.035)	0.224*** (0.030)	0.293*** (0.035)	0.244*** (0.031)
Host Tenure (Log)	-0.060 (0.060)	0.027 (0.063)	0.016 (0.062)	-0.029 (0.044)	-0.024 (0.057)	-0.049 (0.049)
Reviewer Tenure (Log)	0.049** (0.022)	0.151*** (0.023)	0.131*** (0.023)	0.098*** (0.017)	0.141*** (0.022)	0.097*** (0.020)
Constant	-3.914*** (0.674)	-1.254* (0.755)	-3.407*** (0.673)	-0.310 (0.502)	-1.528*** (0.593)	-1.428*** (0.480)
Observations	22,051	22,051	22,051	22,051	22,051	22,051

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

By accounting for this outcome distribution, we combine logit into our original SUR DID model. Table 4.12 gives us the logit DID model when all the six equations are estimated simultaneously and Table 4.11 represents the original DID model using SUR framework.

Surprisingly, both results are very consistent, meaning that all the coefficients of interests decreased. We believe that customers reduced mentioning contents that are related to all the topics in their reviews. To sum up, we have adopted the topic modeling analysis to uncover latent topics from the review corpus and find those topics are related property and host dimensions. We also find that customers tend to focus on specific topics when they review their stays. Not mentioning a variety of qualitative aspects regarding property and host, customers become focused on narrower topics when they leave reviews. Specifically, they discuss less amount of contents in all the property and host related topics and the results are very robust when we repeat the analysis using different estimation schemes.

Table 4.14 DV as the Sentiment Scores per Each Topic

	Sentiment Analysis for each Topic					
	Communication	Location	Equipment	Cleanlines s	Check-in	Recommenda tion
Airbnb	0.559* (0.319)	0.077 (0.049)	0.176*** (0.066)	0.220*** (0.052)	0.217*** (0.048)	0.158*** (0.049)
After	0.495 (0.319)	0.049 (0.049)	0.106 (0.067)	0.170*** (0.053)	0.154*** (0.049)	0.129*** (0.050)
Airbnb*After	-0.522 (0.319)	-0.066 (0.049)	-0.130* (0.067)	-0.186*** (0.053)	-0.181*** (0.049)	-0.146*** (0.050)
No. of Pictures (Log)	0.003 (0.004)	0.008** (0.003)	0.006 (0.011)	0.007 (0.005)	0.013*** (0.005)	0.000 (0.004)
Review Length (Log)	-0.014*** (0.004)	-0.007** (0.003)	-0.022*** (0.005)	-0.013*** (0.002)	-0.023*** (0.003)	-0.008** (0.003)
Reviewer Activeness (Log)	0.009*** (0.002)	0.006*** (0.002)	0.008** (0.004)	0.012*** (0.002)	0.009*** (0.003)	0.009*** (0.003)
Host Tenure (Log)	0.002 (0.003)	-0.002 (0.002)	-0.006* (0.004)	-0.004* (0.002)	-0.001 (0.003)	-0.000 (0.002)
Reviewer Tenure (Log)	0.005*** (0.002)	0.003** (0.001)	0.001 (0.003)	0.002* (0.001)	0.002 (0.002)	0.003* (0.001)
Constant	0.434 (0.320)	0.898*** (0.053)	0.910*** (0.087)	0.781*** (0.057)	0.815*** (0.055)	0.834*** (0.057)
Observations	4,544	10,973	5,831	14,036	10,696	8,896
R-squared	0.058	0.010	0.036	0.041	0.048	0.023

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

We combine the sentiment analysis with topic modeling approach to study how the valence of reviews has changed in each topic. Our overall sentiment analysis results give us the negative sentiment scores, showing that the overall reviews become less positive in terms of tone. We further analyze the sentiment of reviews within each topic category. Table 4.13 represents the results. The valence of the reviews has decreased across all the topics but the sentiment regrading "Equipment and Amenities", "Cleanliness", "Check-in", and "Recommendations" category reviews are significant only. We do not find any significance on the "Communication" and "Location" related reviews.

Robustness Checks

Table 4.15 Listings with Number of Reviews Greater Than 10 & 20

	>=10	>=20	>=10	>=20
	Full Sample		-/+ 3 Years	
Airbnb	0.196*** (0.063)	0.200*** (0.065)	0.138** (0.054)	0.154*** (0.054)
After	0.164*** (0.061)	0.165*** (0.062)	0.097* (0.051)	0.113** (0.051)
Airbnb*After	-0.224*** (0.066)	-0.216*** (0.067)	-0.172*** (0.058)	-0.177*** (0.058)
No. of Pictures (Log)	0.036** (0.018)	0.033* (0.020)	-0.027 (0.027)	-0.016 (0.030)
Review Length (Log)	-0.011* (0.006)	-0.010* (0.006)	-0.032*** (0.011)	-0.025** (0.012)
Reviewer Activeness (Log)	0.027*** (0.006)	0.028*** (0.006)	0.028*** (0.009)	0.026*** (0.010)
Host Tenure (Log)	-0.005 (0.009)	-0.004 (0.009)	0.000 (0.012)	-0.003 (0.012)
Reviewer Tenure (Log)	0.006 (0.005)	0.005 (0.005)	0.016** (0.008)	0.015* (0.008)
Constant	4.541*** (0.115)	4.542*** (0.126)	4.809*** (0.165)	4.753*** (0.175)
Observations	20,569	18,258	6,613	6,248
R-squared	0.007	0.007	0.015	0.012

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

We adopt the different set of samples throughout the analysis, namely Full Sample and Samples before/after 3 years of policy change (-/+ 3 years). To construct more balanced set of samples, we restrict our sample set by removing the properties with number of reviews less than 10 and 20. By doing this, we filter out the sample not only by balancing the equal length of pre and post treatment periods but also by balancing the sample-wise dimension by dropping the properties with very few reviews. Using the balanced sample sets, we still maintain the consistent results as shown in Table 4.14 compared to our main results.

Following the robustness check of specification in (Pu, Chen, Qiu, & Cheng, 2020), we implement a regression discontinuity design (RDD) for robustness check, as an alternative specification to Difference-in-Differences (DID). Based on the time of policy change, we estimate the local average treatment effects centered around the time of policy change. We modify the model using the following equation:

$$Rating_{it} = \beta_0 + \beta_1 * After_t + f(r_t) + \varepsilon_{it} \tag{4.5}$$

Where r_t measures the number of months that from month t to July 2014, when the policy change occurs, and $f(\cdot)$ denotes a flexible function controlling the endogeneity between ε_{it} and the policy change. The alternative specification of RDD can help us to mitigate the concern that our estimated effects of DID come from the time trends around the policy change. However, the RDD assumes the time-varying confounding factors change smoothly over the policy change window (Hausman & Rapson, 2018). We implemented a non-parametric version of regression discontinuity (Calonico, Cattaneo, & Titiunik, 2014) and a parametric version, and show the results in Table 4.15 and Table 4.16.

Table 4.16 Regression Discontinuity Design (Local Linear)

	Airbnb	TripAdvisor	Airbnb
Policy Change	-0.1456*** (0.0563)	-0.0647 (0.0657)	-0.1449** (0.0645)
Control Variables	No	No	Yes
Observations	20,729	6,709	20,597

Standard errors in parentheses, clustered in nearest neighbors
**** p<0.01, ** p<0.05, * p<0.1*

Given that we had both observations from Airbnb and TripAdvisor, we aggregate the monthly average rating for each platform, and implement the RDD for both platforms separately. We expect to see a significant decrease in Airbnb, consistent to our DID estimate.

Table 4.17 Regression Discontinuity Design (Global Parametric)

	Airbnb	
Policy Change	-0.0924** (0.0369)	-0.0762* (0.0403)
First-order Polynomial	--	--
Review Length (Log)		0.0047 (0.0130)
No. of Pictures (Log)		-0.0251 (0.0331)
Reviewer Activeness (Log)		0.0375*** (0.0095)
Host Tenure (Log)		-0.0040 (0.0123)
Reviewer Tenure (Log)		0.0069 (0.0088)
Constant	4.8158*** (0.0378)	4.7657*** (0.1661)
Observations	6,266	6,242
R-squared	0.2429	0.0375
Listing FE	Yes	No
City FE	No	Yes

Standard errors in parentheses, clustered in nearest neighbors
**** p<0.01, ** p<0.05, * p<0.1*

However, the RDD estimate for TripAdvisor serves as a placebo test, which should not exhibit any significant change around the policy implementation window. We visualize the average ratings for both Airbnb and TripAdvisor in Figure 4.7 and Figure 4.8, in which the upper panel is for Airbnb and the bottom panel is for TripAdvisor.

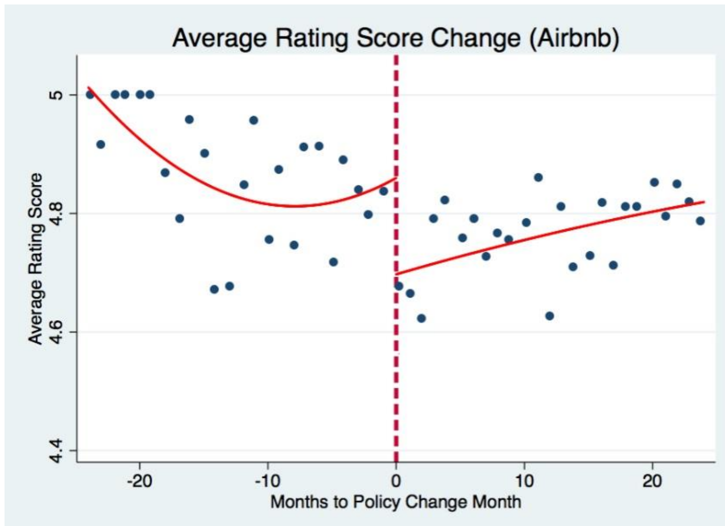

Figure 4.7 Average Rating Score Change on Airbnb

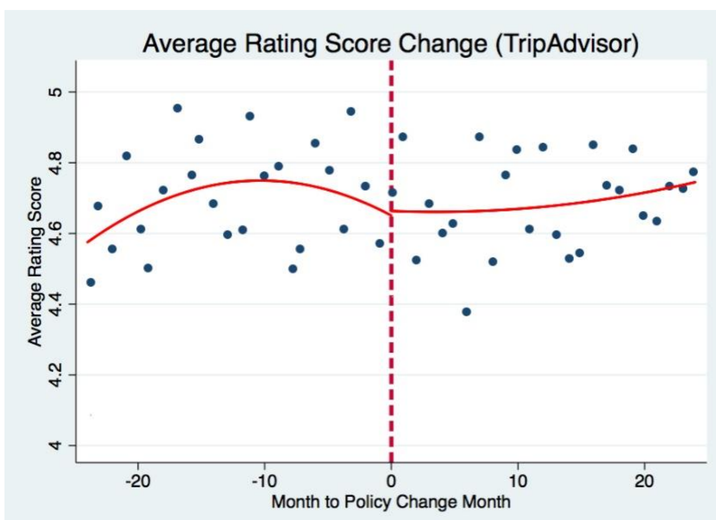

Figure 4.8 Average Rating Score Change on Trip Advisor

Round the red dash line, the Airbnb ratings show a decrease immediately after the policy implementation point, whereas the TripAdvisor ratings are smooth across the policy implementation point. In Table 4.16, we report the estimates for RDD of Airbnb and TripAdvisor ratings. After the policy change, the average ratings for Airbnb decrease by 0.15, which is of similar magnitude to our DID estimate (Column (1)). Consistent to our prediction of placebo test in Column (2), the policy change implemented at Airbnb would not affect the average ratings for TripAdvisor across the interruption time. To improve the estimation efficiency for non-parametric RDD, we also include the covariates in Column (3), and our results remain robust in similar magnitude. To estimate the policy effects in a parametric form, we use first-order polynomial in $f(\cdot)$, and we still obtain a significantly negative impact from the policy change on average ratings of Airbnb.

CHAPTER 5

Conclusions

The rise of sharing economy has disrupted existing business models in industries ranging from transportation (e.g. Uber, Lyft), short-term accommodation (e.g. Airbnb) and freelancing market (e.g. Upwork). One defining feature of sharing economy is the provision of a much wider array of product options due to the decentralized network of suppliers. Airbnb, as a marketplace that facilitates the match between property owners and travelers, also provides a more varied choice of accommodation to fulfill customers' needs. Compared with the traditional lodging business, these short-term rentals are typically characterized by a personal touch, and customized experience with an authentic feel. However, in contrast to large hotel chains with trusted brands and standardized services, Airbnb faces one pressing problem — information asymmetry consumers face when booking a stranger's home. That is, hotels and Airbnb listings differ in one crucial aspect: whereas hotels largely offer a standardized and predictable quality, Airbnb properties offer relatively uncertain quality.

To solve this issue, Airbnb has implemented a variety of steps including background screening, insurance for both guests and hosts, and online word-of-mouth to mitigate risks faced by both guests and hosts. However, consumer ratings are believed to be subject to "reputation inflation" on many sharing economy platforms (Filippas et al. 2019). In this paper, we investigate the impact of a specific quality certification system, Airbnb Plus, a platform-endorsed certification. In light of the overwhelmingly positive reviews on Airbnb, Airbnb Plus allows hosts to generate a quality signal on top of host and property reviews that could effectively differentiate themselves from the competition. Compared with other quality cues like photo quality and cancellation policy, Plus certification operates in a more direct way to reduce information asymmetry, wherein the platform itself plays the role of inspector and certifier — with potential conflicts of interest that may detract from its effectiveness. What is new about this type of quality certification, as compared to prior work (see, e.g., Dranove and Jin 2010), is that the quality tier is created by the platform itself. In order to acquire the Plus certification, properties have to be verified through in-person inspection to ensure quality and design while hosts have to qualify a set of criteria such as great reviews and commitment to high quality services.

This quality certification system is subject to both adverse selection (low quality properties could pay the low inspection cost) and moral hazard problems (lack of long-term quality monitoring capacity), so it is an open question as to whether the quality certification is a credible signal of quality or just "cheap talk" (see, e.g., Barach et al. 2020). Even if the quality certification is credible, the positive demand effects for certified properties could be more than offset by the negative externality effects for non-certified listings nearby, so the net revenue impacts for the platform is ambiguous. Motivated by these open issues, we develop a causal empirical framework for estimating both direct and externality effects of Plus certification, as well as local platform revenue effects. Our identification strategy relies on differential timing of the launch of the Plus service across cities and differential adoption of Plus certification across properties — in conjunction with suitable matching methods. We find that obtaining a Plus certification raises booking rate by 7.6%, which translates to an additional revenue of over $3500 for the average Plus-certified listing. On the other hand, a non-Plus listing in the neighborhood of a Plus listing sees a decline in booking rate of 1.5%, which is a reduction in annual revenues of about $670. Netting out the two effects, we do find a positive net revenue impact for the platform, to the tune of $80,000 of additional revenue for a 2 km zone that contains one or more newly certified properties. Accordingly, we find prima facie evidence that the Plus certification is a credible signal with durable economic impacts — at least for several months following certification.

Our findings have direct implications for property owners, consumers, and the platform itself. For consumers, by creating this quality certification system, Airbnb has provided a solution to potential travelers who have reservations about the quality of the properties. Consumers could not only gain access to a wide variety of homes that could fulfill their unique needs, but are also assured of having a high-quality stay. This new quality certification resolves consumers' confusion and mitigates their uncertainty brought by the ubiquity of positive ratings and reviews. The Plus certification would steer users from other non-certified listings, but more importantly, the quality assurance could convince hotel customers at the margin to choose quality-certified Plus listings instead. Our results indicate that the Plus service is indeed generating net revenue gains for the platform, presumably at the expense of the hotel industry.

The implications for property owners are rather evident. If the property is certified by Airbnb Plus, such quality certification will likely lead to increase in demand and revenue. Property owners should be aware of such a factor and the extent to which it may affect consumers' evaluation and decision making. As found in our paper, such positive influence is subject to a few moderating factors that lead to heterogeneous effects on demand. These factors include alternative quality cues such as price, Superhost and reviews. Because the benefits of Plus certification truly accrue to those properties that will gain from a significant reduction in information asymmetry, property owners should weigh the marginal value of

Plus certification against its cost, while deciding whether to spend efforts in acquiring Plus certification. Property owners may find it worthwhile if the benefits offset the time and monetary cost of acquiring Plus certification. For those properties that are negatively impacted from nearby Plus properties, they could either find alternative dimension (e.g. warranties, word-of-mouth) to differentiate from others, or invest in acquiring the same certification as competition necessitates it.

Turning to the platform, the positive revenue effects might indicate that the Plus service is an effective tool to compete with the hotel industry. Quality certification reduces the frictional impacts of information asymmetry and allows Airbnb to counteract the brand names and reputations of the traditional hotel industry. This would suggest a greater rate of Plus certification in an increasing number of cities worldwide. However, the "unravelling effect" theorized in the quality certification literature (see Dranove and Jin 2010) suggests that more and more properties would opt for certification, reducing the signaling value of the certification (Anderson et al. 1999). If this is true, then the platform would need to be careful about the tradeoff between the effectiveness of this quality signal and the net profit gain afforded by the Plus program. Moreover, our findings mainly pertain to a short-time period after the launch of Plus certification. How the effects evolve in the longer run is still an open question, depending on the extent to which the Plus certification system affects potential entry and exit of listings. The negative externality effect on non-Plus listings may further drive them to drift away from the platform, and raise the bar for future entry, resulting in potential revenue loss and lack of platform participation.

Our study is not free of limitations. First, our analysis is based on secondary archival data, which makes identification of causal effects challenging. Selection bias may unavoidably exist in our study due to some unobservable factors that we fail to control for. However, we believe the chance is relatively low, given that we construct comparable control groups using extensive matching methods, including natural language processing of qualitative review text. We have also deployed a variety of robustness checks to assure the validity of our findings. Second, our analysis is based on a specific sharing economy platform, so care should be exercised in generalizing our findings to other contexts. Finally, our analysis is geared toward short term impacts, over several months following certification, so the results should be carefully extended to the longer term. How the various economic impacts play out in the longer term is an open question, which could be a fruitful direction for future research. Although there are a few limitations in this study, we believe that our research generates useful and robust empirical regularities that have important implications or both research and practice with respect to quality assurance in the sharing economy.

Majority of online marketplaces adopt online review systems for potential customers to be informed about their products and services. Especially sharing economy platforms have been dealing with how to manage their reputation systems in order to promote efficient matches and transactions between buyers and sellers. However, Zervas et al. (2021) points out a critical issue within a sharing economy platform – review inflation. The study claim that reviews on sharing economy platforms, particularly in Airbnb, tend to be overwhelmingly high which distract customers to observe true quality of products. Recent studies have concentrated on the "designing" aspect of reputation systems. Among those studies, several researchers have focused on the reciprocity within bilateral review system where buyers and sellers evaluate each party. Under the bilateral review system, some studies have paid great attention to the simultaneous review mechanism where both parties, buyers and sellers in platforms, could see how they have reviewed each other after both had completed submitting their reviews (Bolton et al. 2012; Fradkin et al. 2020; Mousavi and Zhao 2018). While these prominent studies have generated insightful results in terms of the review policy change (from asynchronous to simultaneous) and the corresponding impacts, their results are conflicting and could not lead to any consensus since different studies have adopted different estimation approaches and data. These ambiguities motivate us to reconsider the simultaneous review mechanism and its impacts on rating and textual reviews. Our study contributes to the sharing economy literature which is particularly focused on the designs of review system. First, we cover the effect on both rating and textual reviews to find out how the valence, volume, and hidden semantic structure change after review policy has been changed to simultaneous system. While Bolton et al. (2012) and Fradkin et al. (2020) mostly analyzed the impact on speed of reviews and valence change in rating, we further incorporate the textual dimension of the reviews as well. Mousavi and Zhan (2018) have worked on this in a similar manner, however, our study covers more sophisticated textual analysis using a combination between topic modeling and sentiment analysis. Our findings from textual analysis show that valence of the reviews became less positive in host-related reviews and in some property-related reviews (For the details, see our results). We also find that valence of the rating has decreased and the number of reviews has increased which is consistent with the prior studies. These additional analysis would bring significant insights to not only Airbnb but also the other sharing economy platforms where they are willing to build and improve the design of their reputation systems. Second, we claim that our quasi-experimental approach and the use of observational data from both platforms, Airbnb and Trip Advisor, may overcome some of the issues coming from inconsistency of the results from similar studies and limited time and scope of the existing identification strategies. Our cross-platform difference-in-difference based on cross-listed property samples allows us to eliminate the unobserved property characteristics since we compare the reviews from both platforms but within the same properties. This enables us to control for property-level unobserved heterogeneity which may bias our results. Also, our study leverages observational data which span longer time window before and after the policy change compared to the fact that the prior studies using experimental designs have dealt with relatively shorter time range such as a couple of month for their experiments. In this way, we can identify the actual effects of review policy change in a long-run. We hope to contribute to the sharing economy literature in a way that our study additionally builds on the current studies related to design of review systems and the underlying mechanisms.

REFERENCES

Agrawal, A., & Goldfarb, A. (2008). Restructuring research: Communication costs and the democratization of university innovation. *American Economic Review*, *98*(4), 1578-90.

Akerlof, G. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics,* 84(3), 488–500.

Anderson, S. W., Daly, J. D., & Johnson, M. F. (1999). Why firms seek ISO 9000 certification: regulatory compliance or competitive advantage?, *Production and operations management*, *8*(1), 28-43.

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

Arora, A., & Asundi, J. (1999). Quality Certification and the Economics of Contract Software Development A Study of the Indian Software Industry. *National bureau of economic research,* (No. w7260).

Avery, C., P. Resnick, and R. Zeckhuser (1999). The market for evaluations. *American Economic Review*, 89 (3), 564-584.

Babar, Y., & Burtch, G. (2020). Examining the heterogeneous impact of ride-hailing services on public transit use. *Information Systems Research*, *31*(3), 820-834.

Bapna, R., Ramaprasad, J., & Umyarov, A. (2018). Monetizing Freemium Communities: Does Paying for Premium Increase Social Engagement?. *MIS Quarterly,* 42(3), 719-735.

Barach, M. A., Golden, J. M., & Horton, J. J. (2020). Steering in online markets: the role of platform incentives and credibility. *Management Science*, *66*(9), 4047-4070.

Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., & Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics* 9(1), 247-274.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6), 2295-2326.

Chen, W., B. Gu, Q. Ye, and K. X. Zhu (2019). Measuring and managing the externality of managerial responses to online customer reviews*. Information Systems Research* 30(1), 81–96.

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

Cramer, J., & Krueger, A. B. (2016). Disruptive change in the taxi business: The case of Uber. *American Economic Review*, *106*(5), 177-82.

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

Dellarocas, C. and C. Wood (2007). The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science*, 54 (3), 460-476.

Deng, S., Huang, Z. J., Sinha, A. P., & Zhao, H. (2018). The interaction between microblog sentiment and stock return: An empirical examination. *MIS quarterly* 42(3), 895-918.

Dewan, S. and V. Hsu (2004). Adverse selection in electronic markets: Evidence from online stamp auctions. *The Journal of Industrial Economics* 52(4), 497–516.

Dimoka, A., Hong, Y., & Pavlou, P. A. (2012). On product uncertainty in online markets: Theory and evidence. *MIS quarterly*, 395-426.

Dranove, D., & Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, *48*(4), 935-63.

Edelman, B. (2011). Adverse selection in online “trust” certifications and search results. *Electronic Commerce Research and Applications*, *10*(1), 17-25.

Feng, J., X. Li, and X. Zhang (2019). Online product reviews-triggered dynamic pricing: Theory and evidence*. Information Systems Research* 30(4), 1107–1123.

Filippas, A., J. J. Horton, and J. Golden (2018). Reputation inflation. *In Proceedings of the 2018 ACM Conference on Economics and Computation,* pp. 483–484. ACM.

Fortune. (2017). Airbnb’s Profits to Top $3 Billion by 2020. (http://fortune.com/2017/02/15/airbnb-profits/).

Fradkin, A., E. Grewal, and D. Holtz (2021). Reciprocity and Unveiling in Two-Sided Reputation Systems: Evidence from an Experiment on Airbnb. *Working Paper*, Boston University.

Gallagher, R. J., Reing, K., Kale, D., & Ver Steeg, G. (2017). Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Transactions of the Association for Computational Linguistics* 5, 529-542.

Gao, G., Gopal, A., & Agarwal, R. (2010). Contingent effects of quality signaling: evidence from the Indian offshore IT services industry. *Management Science*, *56*(6), 1012-1029.

Geva, H., Oestreicher-Singer, G., & Saar-Tsechansky, M. (2019). Using retweets when shaping our online persona: Topic modeling approach. *MIS Quarterly*, 43(2), 501-524.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics*, *24*(3), 461-483.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20(1), 25–46.

Hausman, C., & Rapson, D. S. (2018). Regression discontinuity in time: Considerations for empirical applications. *Annual Review of Resource Economics* 10, 533-552.

Homburg, C., Ehm, L., & Artz, M. (2015). Measuring and managing consumer sentiment in an online community environment. *Journal of Marketing Research* 52(5), 629-641.

Horton, J. J., & Johari, R. (2018). Buyer signaling improves matching: Evidence from a field experiment. *Available at SSRN 3245704*.

Hutto, C., & Gilbert, E. (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*, 8 (1), 216-225.

Hwang, E. H., Singh, P. V., & Argote, L. (2019). Jack of all, master of some: Information network and innovation in crowdsourcing communities. *Information Systems Research*, 30(2), 389-410.

Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 1-24.

Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2018). Extrinsic versus intrinsic rewards for contributing reviews in an online platform. *Information Systems Research*, 29(4), 871-892.

Klein, T. J., C. Lambertz, and K. O. Stahl (2016). Market transparency, adverse selection, and moral hazard. *Journal of Political Economy* 124(6), 1677–1713.

Levine, D. I., & Toffel, M. W. (2010). Quality management and job quality: How the ISO 9001 standard for quality management systems affects employees and employers. *Management Science*, *56*(6), 978-996.

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

Li, X. and L. M. Hitt (2008). Self-selection and information role of online product reviews.

Information Systems Research 19(4), 456–474.

Lodgify. (2020). Airbnb Plus: Everything Hosts Want to Know. (https://www.lodgify.com/blog/airbnb-plus/).

Luca, M. and G. Zervas (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science* 62(12), 3412–3427.

Luo, X., Zhang, J., & Duan, W. (2013). Social media and firm equity value. *Information Systems Research* 24(1), 146-163.

Mayzlin, D., Y. Dover, and J. Chevalier (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review* 104(8), 2421–55.

Milgrom, P., & Roberts, J. (1986). Price and advertising signals of product quality. *Journal of political economy*, *94*(4), 796-821.

Mousavi, R. and K. Zhao (2018). Examining the impacts of Airbnb's review policy change on listing reviews. *SSRN Working Paper.*

Muchnik, L., S. Aral, and S. J. Taylor (2013). Social influence bias: A randomized experiment. *Science* 341(6146), 647–651.

Nelson, P. (1974). Advertising as information. *Journal of political economy*, *82*(4), 729-754.

Nicolau, J. L., & Sellers, R. (2002). The stock market's reaction to quality certification: Empirical evidence from Spain. *European Journal of Operational Research*, *142*(3), 632-641.

Özpolat, K., Gao, G., & Jank, W., and Viswanathan, S. (2013). The Value of Third-party Assurance Seals in Online Retailing: An Empirical Investigation. *Information Systems Research,* 24(4), 1100-1111.

Ozturk, O. C., Venkataraman, S., & Chintagunta, P. K. (2016). Price reactions to rivals' local channel exits. *Marketing science*, *35*(4), 588-604.

Proserpio, D. and G. Zervas (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science* 36(5), 645–665.

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

Pu, J., Chen, Y., Qiu, L., & Cheng, H. K. (2020). Does identity disclosure help or hurt user content generation? Social presence, inhibition, and displacement effects. *Information Systems Research* 31(2), 297-322.

Rao, A. R., & Monroe, K. B. (1989). The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of marketing Research*, *26*(3), 351-357.

Rhue, L., & Sundararajan, A. (2019). Playing to the Crowd? Digital Visibility and the Social Dynamics of Purchase Disclosure. *MIS Quarterly* 43(4), 1127-1141.

Saifee, D. H., Zheng, Z., Bardhan, I. R., & Lahiri, A. (2020). Are Online Reviews of Physicians Reliable Indicators of Clinical Outcomes? A Focus on Chronic Disease Management. *Information Systems Research*, 31(4), 1282-1300.

Skift. (2018a). Airbnb Plus and Everything CEO Brian Chesky Just Announced. (https://skift.com/2018/02/22/airbnb-plus-and-everything-ceo-brian-chesky-just-announced/).

Skift. (2018b). Airbnb Plus' Risky Bet to Push Home Sharing to a New Level. (https://skift.com/2018/04/02/airbnb-plus-risky-bet-to-push-homesharing-to-a-new-level/).

Spence, M. (1977). Consumer misperceptions, product failure and producer liability. *The Review of Economic Studies*, *44*(3), 561-572.

Stevenson, B., & Wolfers, J. (2006). Bargaining in the shadow of the law: Divorce laws and family distress. *The Quarterly Journal of Economics*, *121*(1), 267-288.

Stigler, G.J. (1971). The Theory of Economic Regulation. *Bell Journal of Economics,* 2(1), 3-21.

Sundararajan, A. (2017). *The sharing economy: The end of employment and the rise of crowd-based capitalism*. Mit Press.

Villas-Boas, J. M., & Winer, R. S. (1999). Endogeneity in brand choice models. *Management science*, *45*(10), 1324-1338.

Wang, C., X. Zhang, and I.-H. Hann (2018). Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Information Systems Research* 29(3), 641–655.

Zalmanson, L., Proserpio, D., & Nitzan, I. (2018). *Cancellation Policy as a Signal of Trust and Quality in the Sharing Economy: The Case of Airbnb*. *Working Paper* (available at https://papers. ssrn. com/sol3/papers. cfm.

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

Zervas, G., Proserpio, D., & Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. *Journal of marketing research*, *54*(5), 687-705.

Zhang, S., Lee, D. D., Singh, P. V., & Srinivasan, K. (2017). How much is an image worth? Airbnb property demand estimation leveraging large scale image analytics. *Working Paper* (available at https://ssrn.com/abstract=2976021.