Cross-Market Impact and Technology Adoption in the Sharing Economy

Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

in Industrial Administration (Marketing)

Yijin Kim

Tepper School of Business

Carnegie Mellon University

April 2020

Dissertation Committee:

Kannan Srinivasan (Chair)

Hui Li

Nitin Mehta

Param Vir Singh

Yan Huang (External Reader)

Acknowledgements

I would like to express my deepest gratitude to my committee members Kannan Srinivasan, Hui Li, Nitin Mehta, Param Vir Singh, and Yan Huang for their tremendous help and support. I also thank Lawrence Rapp and Laila Lee for their help facilitating this dissertation. Lastly, I'm extremely grateful to my family and friends. I would not be here without their continuous encouragement and unconditional love.

Abstract

The sharing economy is growing fast, with its global revenue estimated to grow from $14 billion in 2014 to $335 billion by 2025. The growth of the sharing economy has impacted different markets, and its impact is growing larger with technology advancement, which facilitates trade in the sharing economy by improving search, matching, pricing, etc. This dissertation provides insights to these key issues in the sharing economy. The first two chapters examine how the sharing economy affects other traditional industries, in the context of Airbnb and the housing market. In the third chapter, we explore whether and how artificial intelligence technology leveraged to facilitate transactions benefits players in the sharing economy, in the context of Airbnb's Smart Pricing algorithm. These researches are enabled by various quantitative methods, including structural modeling, machine learning, and quasi-experimental methods.

The first chapter examines the impact of Airbnb on the local rental housing market. Airbnb provides landlords an alternative opportunity to rent to short-term tourists, potentially causing some property owners to switch away from long-term rental, thereby affecting rental housing supply and affordability. Despite recent government regulations to address this concern, it remains unclear whether and what type of properties are switching. Combining Airbnb and American Housing Survey data, we estimate a structural model of property owners' decisions and conduct counterfactual analyses to evaluate various regulations. We find that Airbnb mildly cannibalizes long-term rental supply. Cannibalization is concentrated among lower priced and affordable units. Cities where Airbnb is more popular experience a larger reduction in rental supply, but they do not necessarily have a larger percentage of switchers. The counterfactual results suggest that limiting the number of days a property can be listed is a more desirable policy than imposing a linear tax. We propose a new concave tax and show that it outperforms existing policies in terms of reducing cannibalization, maintaining market expansion, limiting affordable units from switching, and allowing economically disadvantaged hosts to benefit from Airbnb. Finally, Airbnb and rent control can exacerbate the negative impacts of each other.

The second chapter investigates the impact of Airbnb on the real estate market. On the one hand, Airbnb may drive up demand for homes and raise prices. More people can, and are willing to, afford to buy homes as Airbnb serves as an extra income source and helps hosts pay mortgages, and many investors see Airbnb as a real estate investment. On the other hand, Airbnb may decrease demand and price because of the negative externality issues such as noise, nuisance, and safety. This chapter examines the direction of the Airbnb's impact and explores the underlying mechanism. By leveraging the Airbnb listings data and the zip code level and property level real estate transaction data, we perform aggregate and micro level analyses to study the impact on the real estate transaction volume and price. In the zip code level analysis using

difference-in-differences approach, we find that the impact of Airbnb on the transaction volume is different across different property types, which motivates the property level analysis. In the property level analysis, we construct an Airbnb performance measure for each property using machine learning to assess Airbnb's impact on the housing price. We find that suitability for entire place listings is associated with the increase in the real estate price, while the price-increasing impact from the suitability for private room listings is dominated by the negative externality impact. The size of the impact differs by housing market conditions, such as real estate supply and owner occupancy. Our findings provide policy implications on where and which Airbnb properties should be regulated to alleviate housing affordability problems.

In the third chapter, we examine the societal impact of machine learning deployment in the context of Airbnb's Smart Pricing algorithm. Based on rich data on lodging demand, Airbnb's Smart Pricing uses machine learning to provide price suggestions to hosts who have difficulty in pricing due to demand uncertainty. However, hosts face a trade-off from incentive misalignment when deciding whether to use the algorithm or not. On the one hand, hosts could benefit from the algorithm, as the price recommendations leverage the detailed data that are not available to shots. On the other hand, the algorithm may work in Airbnb's interest, rather than for hosts' interest. We study how the algorithm adoption affects hosts' profit by estimating a structural model of hosts' pricing and algorithm adoption behavior using the Airbnb listings data. In the model, hosts form beliefs about the algorithmic bias using the price suggestions as a signal, and make the adoption decision based on their beliefs. The results show that using the algorithm often reduces hosts' profit due to algorithmic bias. Providing more correct beliefs about algorithmic bias and more precise signals from the price suggestions reduces the adoption and increases hosts' profit.

Contents

Chapter I

Market Shifts in the Sharing Economy: The Impact of Airbnb on Housing Rentals

1 Introduction

Sharing economy platforms have affected marketing mix decisions (e.g., product, pricing, and distribution channels) by providing an additional channel for individuals to market their products and services. For instance, peer-to-peer marketplaces for short-term accommodations such as Airbnb, HomeAway, and VRBO have emerged as an alternative channel for landlords to market their properties to short-term tourists, in addition to the traditional long-term rental market for local residents. These home-sharing platforms have grown at an exponential rate in recent years. Airbnb, the most popular platform, had over six million listings around the world as of March 2019—more listings than the hotel rooms from the six largest hotel groups combined.[1] Given the opportunity to rent to short-term tourists, some property owners may switch away from the traditional channel of long-term rental to the new channel of Airbnb, as the yields can be two to three times higher on Airbnb than on the long-term rental market.[2] Such switching behavior could impact the rental housing supply and affordability.

A staggering rise in short-term rental platforms has prompted questions over whether they have contributed to rental housing shortages and the affordable housing crisis and whether they should be regulated to protect affordable housing. For example, the City of Los Angeles approved new rules for Airbnb-type rentals in December 2018, after more than 3.5 years of debate since the law had first been proposed.[3] Similarly, in San Francisco, there has been a controversial debate and change over the scope of the city's short-term rental regulation, which first went into effect in February 2015.[4] By 2019, many cities had imposed limits on the number of days that a property can be listed on short-term rental platforms (e.g., a

[1]See https://press.airbnb.com/airbnb-hosts-share-more-than-six-million-listings-around-the-world/

[2]See https://tranio.com/articles/airbnb_a_game-changer_for_the_commercial_property_market_4982/.

[3]See https://www.latimes.com/local/lanow/la-me-ln-airbnb-rental-ordinance-20181211-story.html

[4]In a ballot measure in San Francisco in 2015, 55% of voters rejected Proposition F, which would have reduced the number of days that owners can rent out their properties from 90 to 75. See https://www.theguardian.com/us-news/2015/nov/04/san-francisco-voters-reject-proposition-f-restrict-airbnb-rentals. Later, in 2016, San Francisco approved a new rule that requires short-term rental websites such as Airbnb to display each host's registration number next to their listings or email the information to the city's short-term rentals office. This rule supplements San Francisco's existing short-term rental regulations that require hosts to register with the city's short-term rentals office. See http://fortune.com/2016/06/07/sf-airbnb-new-rules/.

maximum of 90 days in San Francisco and 180 days in Los Angeles). However, most of these policies were launched without empirical evidence on the actual impact of Airbnb on the rental housing market.

In this paper, we seek to answer the following two questions. First, how does Airbnb affect rental housing supply and affordability? In particular, we examine how many units are taken off the rental market (i.e., the impact on rental supply) and what types of properties are taken off (i.e., the impact on rental affordability). Second, what is the impact of potential policies or regulations on short-term rentals? Answering these questions requires an understanding of the underlying trade-offs, or benefits and costs, for property owners regarding whether to rent on the long-term market or list on the short-term market. The benefits of renting can be directly observed from the prices and occupancy rates in the long-term market and on Airbnb. However, the costs of Airbnb and long-term rental hosting and how they differ with respect to demographics, properties, and cities are unknown.

We estimate a structural model of property owners' decisions using Airbnb listings data and American Housing Survey data. We recover the underlying heterogeneous costs of long-term renting versus Airbnb listing, which allow us to simulate market outcomes in the absence of Airbnb to examine its impact and evaluate market outcomes under different policies. In the model, property owners first make a discrete choice among Airbnb, long-term rental, and an outside option. The outside option refers to owner-occupied units, meaning that owner lives in the unit and chooses neither Airbnb nor long-term rental. This decision is usually made yearly, as rental leases are typically one year long. Second, if they choose Airbnb, owners decide the number of days to list their properties on Airbnb, which can be a monthly decision. The two decisions are linked in that the ex ante expected profit from the second decision affects the first decision. We also model hosts' availability as an exogenous factor that can impact the two decisions. The costs of Airbnb and long-term rental hosting, as well as host availability, are allowed to be heterogeneous by property characteristics, host demographics, various metro area characteristics (e.g., population, density, mortgage affordability, wage and employment in the accommodations industry, how long Airbnb has been present, and how favorable city regulations are to short-term rentals) and over time.

The results show that Airbnb mildly cannibalizes the long-term rental supply but creates a market expansion effect. The level of cannibalization varies significantly across metro areas. Interestingly, we find that although the reduction in the rental supply is larger in metro areas where Airbnb is popular, the percentage of switchers is not necessarily larger in these areas. For example, San Francisco has the highest Airbnb popularity and suffers the most from rental supply reduction, but the percentage of switchers is among the lowest. It suggests that most of the Airbnb listings are from market expansion rather than cannibalizing the rental supply in San Francisco. Policymakers need to take a holistic view when evaluating the impact of Airbnb. Although Airbnb causes a mild reduction in the overall rental supply, the breakdown of the decrease by property characteristics raises concerns about housing affordability. The cannibalization impact is largely concentrated among lower priced, affordable units rather than among higher priced, luxurious ones. A basic studio or one-bedroom apartment is more likely to be taken off the long-term rental market than a house with two or more bedrooms with amenities.

We highlight the usefulness of the structural model in identifying the actual potential switchers. It is not appealing to take the observed data and assume, without modeling the hosts' decisions, that an observed "full-time" ("part-time") Airbnb listing always implies cannibalization (market expansion). First, hosts who list full year on Airbnb does not necessarily mean that they are switchers from the long-term rental market. For instance, we find that senior hosts tend to list full time on Airbnb, yet they do not choose long-term rental in the absence of Airbnb, suggesting that they are not switchers. Second, hosts who list part of a year on Airbnb does not necessarily mean that they are not available for the rest of the months and cannot choose long-term rental. Hosts may choose to list for shorter if the Airbnb profit is large enough to allow them to list part time and still earn more than listing in the long-term rental market. Overall, one needs to systematically model the revenue and cost trade-offs of the hosts in order to recover who are switchers.

In the counterfactual analysis, we evaluate two sets of policies related to the supply and affordability of rental housing. The first set of counterfactuals are motivated by recent regulations on short-term rental. Policymakers are continuously searching for effective policies to prevent switching away from long-term rentals, especially in cities with tight housing markets such as San Francisco, New York, and Los Angeles. In addition to limiting the length of listings on Airbnb, local municipalities also require hosts to collect certain taxes from guests, similar to a hotel occupancy tax. We examine these two existing policies (days limit and a linear tax) and further propose a new concave tax that imposes a higher tax on affordable units and a lower tax on expensive units, which is motivated by our finding that cannibalization is largely concentrated among affordable units.

A desirable policy should maintain the positive impact of Airbnb (non-switchers or market expansion) and reduce the negative impact of Airbnb (switchers or cannibalization). Therefore, we assess the desirability of the three policies along four dimensions: (1) the ability to maintain the number of non-switchers while reduce the number of switchers, (2) the ability to reduce cannibalization rate, or percentage of switchers, (3) the ability to reduce the fraction of affordable units among switchers, and (4) the ability to maintain the number of economically disadvantaged hosts (e.g., low income, young, or low education hosts) among non-switchers and allow them to continue benefiting from Airbnb. The third measure is particularly relevant to affordable housing concerns and the fourth measure is relevant to social inequality because it captures potential differential policy impact on heterogeneous hosts. We find that the proposed concave tax outperforms the other two policies along all four dimensions. The days limit is the second-best policy and the linear tax performs the worst.

The second set of counterfactuals focus on rent control policy, which limits the rent in the long-term rental market. Economists are virtually unanimous in concluding that rent controls are destructive because they reduce the supply of available housing. When a rent control policy is imposed, property owners choose not to rent out their units for long-term rental. Despite the known adverse impacts, the states of California, Maryland, New Jersey, New York, Oregon, and the city of Washington D.C. still have some rent control or stabilization policies on the books (as of March 2019).[5] We show that this negative effect of rent control policy is aggravated when Airbnb is available, as Airbnb serves as an additional profitable option for property owners and can motivate them to further switch away from the long-term rental market.

The results have strong policy implications for short-term rentals and affordable housing. Airbnb has been debated and regulated in cities that it has entered. It is difficult for policymakers to assess the impact of Airbnb because it requires knowing whether the properties would have been in the rental market had Airbnb not been available. Our model can be used to assess the impact of Airbnb on rental supply and affordability. The results provide a detailed profile of potential switching hosts and properties, which can serve as a foundation for policy making. We also provide a thorough evaluation of the desirability of various short-term rental regulations and propose a new policy that can outperform existing ones. Finally, we show that rent regulation must be implemented with extra caution when Airbnb is available, as lower profits from long-term rental can cause landlords to switch to Airbnb.

2 Literature Review

This paper contributes to the recent literature on the sharing economy (see Einav, Farronato, and Levin 2016 for a review of the sharing economy). In particular, our work relates to literature on (1) the impact of Airbnb on housing and rental market, (2) the impact of sharing economy on traditional industries, (3) supply decisions in the sharing economy, and (4) how sharing economy affects marketing mix decisions.

There have been papers in Economics and Marketing that study Airbnb's impact on the housing market. Lee (2016) and Gurran and Phibbs (2017) provide descriptive analyses of Airbnb and the rental housing market in Los Angeles and Sydney, Australia, respectively. Other studies focus on how home-sharing affects housing prices and rents in a particular city. Horn and Merante (2017) find that a one-standard-deviation increase in Airbnb listings is associated with an increase in asking rents of 0.4% in Boston. Sheppard and Udell (2018) find that doubling the total number of Airbnb listings within 300 meters of a house is associated with an increase in house prices of 6% to 9% in New York City. Koster et al. (2019) find that

[5] See https://www.curbed.com/2019/3/8/18245307/rent-control-oregon-housing-crisis

the adoption of Home Sharing Ordinances reduced housing prices by 3% and rents by 3% in Los Angeles. Recently, researchers in Marketing also contribute to this topic. Barron, Kung, and Proserpio (2019) use a comprehensive dataset covering the U.S. and find that a 10% increase in Airbnb listings leads to a 0.42% increase in rents and a 0.76% increase in house prices.

We contribute to the literature and differentiate from previous studies in important ways. First, while most studies focus on the impact of Airbnb on housing and rental prices, we focus on the supply choices of the hosts and assess how many and what types of properties are taken off the rental market, which are the causes of the change in housing and rental prices studied by previous works. To the best of our knowledge, we are the first to systematically and formally model the hosts' decisions and recover the underlying cost of hosting. Second, other studies mostly focus on a particular city. We leverage data on a wide variety of cities and show that there is significant heterogeneity across cities, which is helpful for localized policy making. Third, while other studies mostly provide descriptive analysis or use a difference-in-difference approach, we use structural models to explicitly analyze the underlying trade-offs faced by individual hosts. The framework allows us to conduct counterfactual analysis. The results have strong policy implications regarding short-term rental regulations and rent control.

More broadly, our paper contributes to the literature on how sharing economy affects traditional industries and incumbent firms. For instance, ride-sharing services are found to affect earnings of tax drivers (Berger, Chen, and Frey 2018), automobile ownership (Gong, Greenwood, and Song 2017), alcohol-related motor vehicle fatalities (Greenwood and Wattal 2017), and local entrepreneurial activity (Burtch, Carnahan, and Greenwood 2018). On the subject of Airbnb, in a pioneering work, Zervas, Proserpio and Byers (2017) study the impact of Airbnb's entry on hotels in Texas and find that Airbnb mildly cannibalizes hotels, with lower price hotels being the most affected. Li and Srinivasan (2019) study how Airbnb's flexible supply changes the way in which the industry accommodates seasonal demand and how incumbent hotels with fixed capacity should respond.

Our work also relates to the stream of literature on supply choices in the sharing economy. Zhang, Mehta, Singh, and Srinivasan (2018) model Airbnb hosts' decision of whether to operate or block the listing, along with the listing quality decision (e.g., the image quality in the listing description and the host's service effort). Li, Moreno and Zhang (2016) study the pricing decisions of Airbnb hosts and find that a substantial number of Airbnb hosts are unable to optimally set prices. We contribute to the literature by studying the property owners' decisions of whether and how long to list on Airbnb.

Finally, our paper contributes to the literature on how sharing economy affects marketing mix decisions (e.g., product choice, pricing, and distribution channels). Jiang and Tian (2018) study sharing economy enabled collaborative consumption and find that when the firm strategically chooses its retail price, consumers' sharing of products with high marginal costs is a win-win situation for the firm and the consumers. Tian and Jiang (2017) study how consumer-to-consumer product sharing affects the distribution channel and find that the sharing market tends to increase the retailer's share of the gross profit margin in the channel. Dowling et al. (2019) study two common pricing strategies in car sharing services, pay-per-use and flat-rate pricing. They find a prevalent and time-persistent pay-per-use bias because of underestimation of usage, a preference for flexibility, and the influence of physical context (e.g., weather). They suggest that the pay-per-use bias may be the prevalent tariff choice bias in the Sharing Economy.

3 Data

3.1 Data Description

The two main data sets used in this study are the 2015 and 2017 American Housing Survey (AHS) and Airbnb listings data for 9 representative metropolitan areas.[6] First, the AHS is the most comprehensive longitudinal national housing survey in the U.S. that gathers detailed property-level data on properties that are either owner-occupied, renter-occupied, or vacant in a metropolitan area. Each observation includes a housing unit, its property characteristics (e.g., number of bedrooms and bathrooms, amenities, property type), occupant demographics (e.g., age, education, income, gender, marital status), tenure information (whether the unit is owner-occupied, renter-occupied, or vacant), and rent if applicable. As the survey is conducted biennially, we utilize the most recent two years' data at the time of this study. These two years also have stronger Airbnb's presence than previous years as Airbnb continues to grow over time.

The second data set contains information on every Airbnb property listed on Airbnb in 2015 and 2017, collected by AirDNA, a third-party company that specializes in data collection and analysis. Each property record contains monthly performance information, such as the number of days available for booking, average daily rate, and occupancy rate. It also includes over 20 property characteristics such as location (zip code), property type (e.g., house, apartment), listing type (entire place or private/shared room), number of bedrooms and bathrooms, and amenities such as a kitchen, air conditioning, heating, washer, dryer, fireplace, and parking space. We also collect data on when a property is first listed on Airbnb to distinguish between none listing and new listing.[7] In the 9 representative metro areas, there were 169,338 properties listed on Airbnb in 2015 and 252,459 properties in 2017.

[6]The U.S. Office of Management and Budget (OMB) refers to a metropolitan area as a Core Based Statistical Area (CBSA), which corresponds to an urbanized core area containing a substantial population and its adjacent communities having a high degree of economic and social integration with that core. For convenience, we denote a metro area by its principal city in the CBSA (e.g., New York for New York-Newark-Jersey City, NY-NJ-PA).

[7]For instance, if a property was first listed in February 2015, we exclude January 2015 when estimating the host's second-stage decision of how many days to list in a month because zero day listed in January 2015 is due to not entering Airbnb yet instead of choosing not to list.

Combining the two data sets provides a comprehensive data set on every property in the selected area. A property is either listed on Airbnb (units in the Airbnb data set), on the long-term rental market (renter-occupied or for-rent vacant units in the AHS data set), or neither (owner-occupied units in the AHS data set). Hereafter, we refer to choosing neither Airbnb nor long-term rental as the outside option.[8]

We focus on three sets of covariates in the empirical analysis: property characteristics, demographics, and market characteristics. The property characteristics are available at the property level in both the AHS data and the Airbnb data. Demographic information is available for each property in the AHS but not for the Airbnb listings. We collect zip code-level demographics from the American Community Survey (ACS) and impute the host demographics for the Airbnb properties using the local zip code-level demographics. The metro area characteristics include the metro area-level population, density, employment and wage in the accommodation industry from the ACS data and mortgage affordability information from the Zillow Mortgage Affordability Index. We also collect data on an additional set of metro area-level variables that serve as covariates in the estimation of hosts' choices and instruments in the hedonic regressions of revenues. These variables include rent-to-own ratio, unemployment rate, number of air passengers to the city, Airbnb regulation score, and Airbnb history. Rent-to-own ratio and unemployment rate are collected from the ACS. The number of air passengers to the city is from the T-100 Market (All Carriers) database published by the Bureau of Transportation Statistics.[9] Airbnb regulation score, which measures how friendly city policies are to short-term rentals, is published by R Street Institute.[10] Lastly, Airbnb history, measured by the time since Airbnb reached 10% of the total rooms supplied by hotel and Airbnb in a city, is computed using our Airbnb data set and the hotel data from tourism-related reports and articles.[11]

3.2 Data Patterns

In this subsection, we describe the observed data patterns that motivate our empirical model specifications. In particular, we present the percentage of Airbnb, long-term rental, and outside option units, which relates to the first-stage decision of whether to list, and the listing patterns for the Airbnb units, which relates to the second-stage decision of how many days to list.

Table 1 presents the percentage of Airbnb, long-term rental, and outside option units by year and the summary statistics of properties choosing each option. In 2015, 0.70% of the properties chose Airbnb, 40.86%

[8] Note that the properties in the Airbnb data set can overlap with the properties in the AHS data set (for instance, if the AHS owner-occupied properties are listed as "private rooms" on Airbnb). However, the observed characteristics in the two data sets do not allow us to distinguish whether an AHS property is listed on Airbnb. Given that the number of Airbnb listings as a fraction of the total number of AHS properties is very small (0.70% in 2015 and 1.02% in 2017), we assume that there are no overlaps and combine the two data sets without removing overlapped properties.

[9] See https://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=111.

[10] See https://www.rstreet.org/wp-content/uploads/2016/03/RSTREET55.pdf.

[11] See, for example, https://washington.org/dc-information/washington-dc-facts.

Table 1: Summary Statistics by Airbnb, Long-Term Rental, and Outside Option

	Airbnb	Long-term Rental	Outside Option
2015: Number of Units	169,338	9,864,703	14,110,939
2015: Proportion (%)	0.70	40.86	58.44
2017: Number of Units	252,459	9,846,744	14,558,423
2017: Proportion (%)	1.02	39.93	59.04
Number of Bedrooms	1.38 (0.96)	1.95 (0.98)	3.14 (0.98)
Number of Bathrooms	1.31 (0.70)	1.69 (1.22)	3.12 (1.41)
Apartment (%)	71.88 (44.96)	73.45 (44.16)	13.29 (33.94)
Kitchen (%)	90.60 (29.18)	98.88 (10.52)	99.97 (1.84)
Air Conditioning (%)	80.28 (39.79)	87.13 (33.48)	93.48 (24.69)
Heating (%)	86.97 (33.67)	99.42 (7.62)	99.71 (5.33)
Washer (%)	62.36 (48.45)	45.17 (49.77)	91.72 (27.56)
Dryer (%)	60.44 (48.90)	42.22 (49.39)	90.60 (29.19)
Fireplace (%)	12.01 (32.50)	10.87 (31.13)	45.56 (49.80)
Parking Space (%)	41.79 (49.32)	30.79 (46.16)	76.28 (42.54)
Private or Shared Room (%)	43.9 (49.6)		
Airbnb Daily Price ($)	199.79 (1,566.05)		
Airbnb Occupancy Rate (%)	25.4 (35.9)		
Monthly Rent ($)		1,362.05 (1,269.75)	
Rental Occupancy Rate (%)		91.9 (27.3)	

Note: Standard deviations are shown in parentheses.

chose long-term rental, and 58.44% chose the outside option. The numbers changed to 1.02%, 39.93%, and 59.04%, respectively, in 2017 as the number of Airbnb properties increased by nearly 50% from 2015 to 2017. We find that the Airbnb units are comparable in property characteristics to the long-term rental units. For example, both have smaller numbers of bedrooms and a larger proportion of apartment units than the outside option properties. This suggests that properties on Airbnb and in the long-term rental market could come from the same pool. We also find that Airbnb could generate more rental income per month than a long-term rental, as implied by Airbnb's average daily price ($199.79) and occupancy rate (25.4%) and the average monthly rent for long-term rentals ($1,362.05) and occupancy rate (91.9%). The payoff difference could motivate hosts to switch from long-term rental to Airbnb.

We further examine the listing patterns of properties that choose Airbnb. When property owners list their properties on Airbnb, they can choose the dates when the property is available for booking (i.e., listed) or blocked from accepting reservations (i.e., not listed). We find that the listing pattern is heterogeneous across hosts and also across months within a host. Figure 1 plots the monthly number of days listed for two representative Airbnb properties. We find that hosts often choose not to list at all for a particular month. If a property is listed, it is more likely to be listed for the full month than for some part of the month. This is also supported by the histogram of the number of days listed in a month (Figure 2). The histograms for both 2015 and 2017 show a bi-modal pattern with the two modes at "no listing" and "full listing". In addition, we find that hosts are more likely to list their properties longer in 2017 than in 2015. Among all property-month observations, the percentage of full listing increased from 28.6% in 2015 to 42.8% in 2017 while the percentage of none listing decreased from 39.3% in 2015 to 34.2% in 2017.

We also explore the total number of days in a year that a property is listed on Airbnb, as the total revenue generated per year is more informative when compared with long-term rentals. Figure 3 shows the histogram of the percentage of days that a property is available for booking by year. In 2015, 51.9% of the properties are listed for less than half of the year. These "part-time" Airbnb hosts may list their properties only when they are not utilizing the property, for example, when they are away on vacation. In contrast, some properties are listed most of the time. The data shows that 33.7% of the observations are listed for more than 70% of the year, 26.5% are listed for more than 80% of the year, and 18.3% are listed for more than 90% of the year. Some of these properties may have been in the long-term rental market had Airbnb not been available. The listing pattern in 2017 shows very similar pattern with a slight shift to the right (i.e., listing longer).

Figure 1: Representative Airbnb Listing Patterns

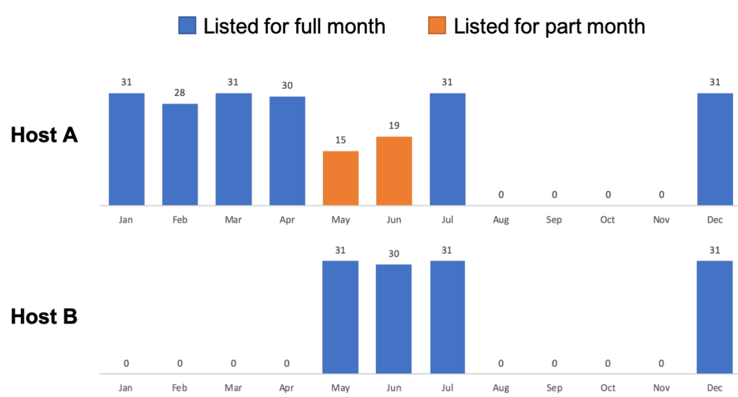


Figure 2: Histogram of Monthly Number of Days Listed

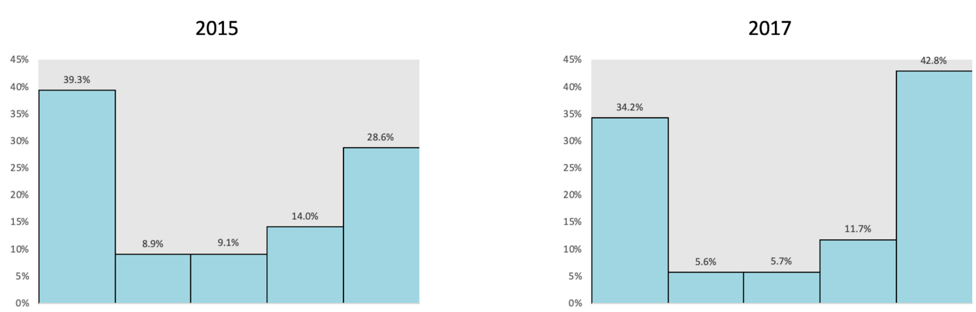


Figure 3: Histogram of Percentage of Days Listed in Each Year

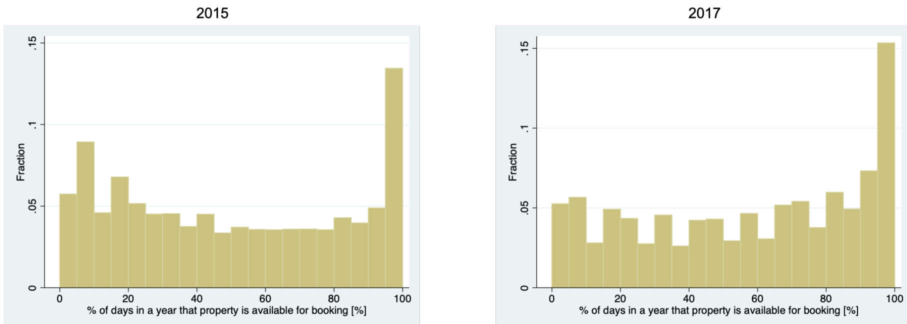


Table 2: Percentage of Units: By Metro Area

		Percentage of Units [%]		
Metro Area	# Units	Airbnb	Rental	Outside
Boston-Cambridge-Newton, MA-NH	1,842,184	0.83	37.41	61.76
Chicago-Naperville-Elgin, IL-IN-WI	3,484,175	0.41	35.15	64.44
Dallas-Fort Worth-Arlington, TX	2,438,499	0.23	40.83	58.94
Detroit-Warren-Dearborn, MI	1,673,407	0.10	29.03	70.87
Miami-Fort Lauderdale-West Palm Beach, FL	2,040,563	1.39	40.28	58.33
New York-Newark-Jersey City, NY-NJ-PA	7,450,000	1.28	47.15	51.58
Phoenix-Mesa-Scottsdale, AZ	1,588,120	0.47	36.03	63.50
San Francisco-Oakland-Hayward, CA	1,708,001	1.51	44.89	53.60
Washington-Arlington-Alexandria, DC-VA-MD-WV	2,166,332	0.80	36.12	63.08

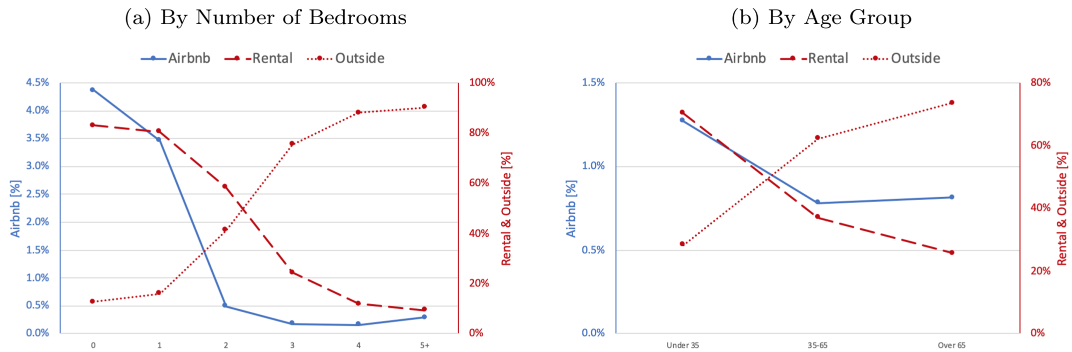
3.2.1 Heterogeneity

The data patterns vary by metro area, property characteristics, and demographics. We present the data patterns averaged across years in this subsection, as they do not qualitatively change over time. Table 2 and Figure 4 show the observed percentages of Airbnb, long-term rental, and outside option units by metro area, by number of bedrooms, and by age group. First, these percentages vary significantly across metro areas. The percentage of Airbnb properties ranges from 0.10% in Detroit to 1.51% in San Francisco. The top three metro areas with the highest proportion of Airbnb properties are San Francisco, Miami, and New York. Second, the percentage of units choosing each option differs by property characteristics such as the number of bedrooms. As the number of bedrooms increases, the proportions of Airbnb properties and long-term rental properties both decrease, except that the proportion of Airbnb increases from 0.16% for 4 bedroom units to 0.29% for 5+ bedroom units. Third, the percentages of Airbnb units and long-term rental units both decrease with age, except that the percentage of Airbnb increases for seniors of age over 65. This is consistent with Airbnb's report that seniors are the fastest-growing demographic of Airbnb hosts.[12]

We also find that the Airbnb listing behavior varies by metro area, property characteristics, demographics, and season. Figure 5 shows the histogram of the number of days available for booking in a month by metro

[12]See https://www.airbnbcitizen.com/seniors-airbnbs-fastest-growing-most-loved-demographic/

Figure 4: Percentage of Units: By Property Characteristics and Demographics



area, number of bedrooms, age, and season. Each observation represents a property-month combination. The overall bi-modal pattern in Figure 2 holds for all subgroups, with variations across subgroups. For example, in terms of metro area, the percentage of no-listing months is 26.9% in Detroit and 39.8% in San Francisco. In terms of property characteristics, properties with more bedrooms are less likely to have no-listing months and more likely to have full-listing months. In terms of demographics, senior hosts are more likely to list longer than others. Finally, properties are less likely to be listed in fall and more likely to be listed in spring and winter.

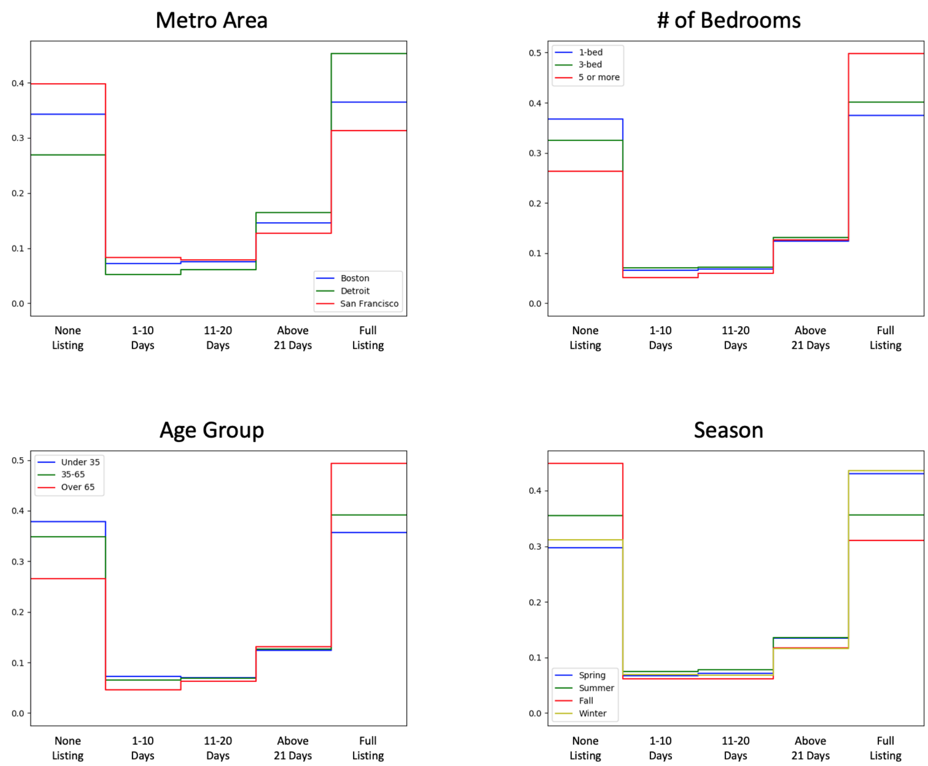
4 Model Setup

Property owners make endogenous decisions of whether and how long to list their properties based on cost-benefit trade-offs. Their decisions can be further affected by exogenously determined availability. Intuitively, a need for self-use of the property (i.e., unavailability) can prevent them from renting out their property. For instance, a host on Airbnb may block his/her calendar and not listing at all in a month if he/she needs to use the property and host his family during holidays; such unavailability may explain the bi-modal listing pattern we observe in the data, especially the mode of none listing. Also, a property owner may choose not to rent on either the long-term rental market or Airbnb if he/she needs to live in the property. The observed listing decisions are thus a result of both the endogenous decisions of the hosts and the exogenous availability of the hosts. Given that we do not directly observe hosts' availability, we take a probabilistic view and assume that hosts' availability is determined by an availability shock which arrives with some probability.

In the model, property owners make decisions in two stages:

First-stage decision. At the beginning of each year, the property owners select the use of their properties among three options: (1) Not rent out, or the outside option. It allows property owners to live in the property for self-use or keep the property vacant without listing it on any market; (2) Rent on the long-

Figure 5: Histogram of Monthly Number of Days Listed: By Metro Area, Property Characteristics, Demographics, and Season



term rental market and commit to be available (i.e., no self-use) during the next 12 months with some cost. The cost can include hosting cost and commitment cost that relates to the costs or efforts for keeping the property available when self-use need arises, for instance, staying in a hotel if the property owner temporarily needs to use the property; (3) Rent on Airbnb and receive an availability shock (i.e., self-use need or not) with some probability each month. Receiving the shock of being unavailable will prevent them from listing in that month.

Second-stage decision. For property owners who choose Airbnb in the first stage, the availability shock is realized at the beginning of each month. If not available, the property owner does not list in that month. If available, the property owner decides the number of days to list on Airbnb. For property owners who choose long-term rental or the outside option in the first stage, there are no further decisions to make in the second stage, as long-term rental hosts are bound by long-term leases during the lease period.

This two-stage setup is a representation of the fact that property owners usually decide the use of their property at the year level and the number of days to list on Airbnb at the month level. Although the availability shock only realizes in the second stage, hosts take it into account when making their first stage decisions. We allow property owners to be heterogeneous in the costs and benefits of each option and the probability of receiving the availability shock.

Property owners make the decisions to maximize their profits given their expectations about the rent and occupancy rate they can obtain in the long-term rental market and the price and occupancy rate they can obtain on Airbnb. We assume that the expectation is formed using a hedonic approach, which we detail in Section 4.3. We normalize the profit from the outside option to zero.

In addition to the revenues, property owners also consider the costs of renting on the long-term rental market versus Airbnb. The costs can include both *tangible* costs (e.g., property maintenance) and *intangible* costs (e.g., hassle from dealing with renters, living with Airbnb travelers). Specifically, in the first stage, property owners may incur the cost of long-term rental and the fixed cost of Airbnb hosting. The cost of long-term rental may include fees, taxes, insurance, maintenance costs, and the cost of committing to be available throughout the year. The fixed cost of Airbnb hosting may include the psychological cost of renting out property to transient guests, preparing property photos and descriptions, and preparing furnishings and amenities. In the second stage, Airbnb hosts may incur variable costs during the days they list their properties on Airbnb. These costs may include responding to guest inquiries and reservations, checking guests in and out, maintaining the property, and paying utility bills. We discuss how the cost functions are constructed and estimated in the following sections.

4.1 Second Stage: Continuous Decision of Listing on Airbnb

We first describe the model setup for the second-stage decision, as the profits from the second stage are nested into the first-stage decision and property owners need to form expectations about the second stage before they make their first-stage decisions. In the second stage, conditional on choosing Airbnb in the first stage, the owner first receives an availability shock at the beginning of each month. The probability of property i being available in month t is:

$$a_{it} = \frac{\exp\left(\beta^a X_{it}^a\right)}{1 + \exp\left(\beta^a X_{it}^a\right)} \tag{1}$$

where X_{it}^a includes host demographics (age, education, income, marital status, gender), metro area characteristics (population and density), and season fixed effects (spring, summer, fall, winter). Intuitively, a host's availability can be affected by who the host is, where the host lives, and during which time of the year. This setup allows for heterogeneity across hosts by including host-specific factors such as demographics and metro characteristics. It also allows for heterogeneity within the same host by accounting for the time of the year, which is motivated by the time-varying listing pattern even within the same host as discussed in Section 3.2.

Let s_{it} denote the number of days that property i is listed on Airbnb in month t. If the realized shock is unavailable, the owner cannot list the property in that month, so $s_{it} = 0$. If the realized shock is available, the owner chooses $s_{it} \in [0, \bar{s}]$, where the total number of days in each month $\bar{s}$ serves as the upper bound.[13] In the counterfactual analysis, we allow $\bar{s}$ to reflect the maximum listing length imposed by government regulations. In this section, we derive the model by allowing s_{it} to take any value between 0 and $\bar{s}$ for illustrative purposes. We account for the fact that s_{it} is an integer when we estimate the model and detail how we treat the integer issue in the online appendix.

The optimal number of days to list is chosen to maximize the monthly profit from Airbnb for property i in month t:

$$\Pi_{it}^{A}(s_{it}) = p_{it}^{A}\phi_{it}^{A}s_{it} - c_{it}^{Av} \cdot \bar{s}\left(\exp\left(\frac{s_{it}}{\bar{s}}\right) - 1\right) \tag{2}$$

where p_{it}^{A} and ϕ_{it}^{A} are the expected average daily price and occupancy rate of property i. We discuss how the expectations are formed in Section 4.3. c_{it}^{Av} is the heterogeneous *variable* cost of Airbnb hosting per day, to be parameterized later. The first term of the profit function represents the revenue, which is proportional to the number of days booked. The second term represents the cost, which increases with the number of days listed. Note that the profit is zero if the property is not listed, i.e., $\Pi_{it}^{A}(0) = 0$. Taking the derivative with respect to s_{it}, the optimal number of days to list is:

$$s_{it}^{*} = \min\left\{\bar{s} \cdot \ln\left(\frac{p_{it}^{A}\phi_{it}^{A}}{c_{it}^{Av}}\right), \bar{s}\right\} \tag{3}$$

where the min operator accounts for the range of $s_{it} \in [0, \bar{s}]$. The solution suggests that the number of days to list on Airbnb is an endogenous function of the ex ante expected revenue $\left(p_{it}^{A}\phi_{it}^{A}\right)$ and heterogeneous cost of Airbnb hosting $\left(c_{it}^{Av}\right)$. It has the desirable property that the larger the revenue-to-cost ratio is, the longer the property owner chooses to list on Airbnb. Note that we (researchers) only observe the number of days listed in the data, but not hosts' availability. The observation of $s_{it} = 0$ can comes from either the host being not available or the host being available but choosing $s_{it}^{*} = 0$. This will be reflected in the derivation of the second stage probability in Section 5.1.

The variable cost of Airbnb hosting for property i at time t is formulated as follows:

$$c_{it}^{Av} = \bar{c}^{Av} + \beta^{Av}X_{iT}^{Av} + \xi_{mT}^{Av} + \epsilon_{it}^{Av} \tag{4}$$

[13]In practice, there are government regulations that limit the maximum number of days a property can be listed on Airbnb. These regulations were imposed after our sample period ended, so we do not account for them as $\bar{s}$ in our model estimation. The only exception is the Airbnb law in San Francisco, which went into effect on February 1, 2015 and restricts short-term rentals to a maximum of 90 days per year. However, the law was not strictly enforced, as the data show that 25% of the listings were listed for more than 90 days during the 9-month period from February 2015 (when the law went into effect) to October 2015. In fact, the lack of strict law enforcement was also reported during this time period. See https://www.sfchronicle.com/business/article/Airbnb-loses-thousands-of-hosts-in-SF-as-12496624.php.

where $\bar{c}^{Av}$ is the baseline cost, X_{iT}^{Av} are observed characteristics that affect the cost in a continuous way, ξ_{mT}^{Av} is a market-specific time fixed effect that captures any remaining time-varying unobservables, and ϵ_{it}^{Av} is an i.i.d. normally distributed idiosyncratic shock with mean zero and standard deviation σ_2.[14] Specifically, X_{iT}^{Av} includes property characteristics (number of bedrooms/bathrooms/amenities, listing type) and a set of metro area-level characteristics that vary by year and relate to the cost of hosting. The metro area-level characteristics include mortgage affordability index, average wage and employment in the accommodation industry (measured as the percentage of population who work in the accommodation industry), and Airbnb regulation score (measures how friendly city regulations are to short-term rental). Intuitively, a large property may be more costly to maintain and induce a larger variable cost of hosting. Hosts in cities with more employees and lower wages in the accommodations may find it easier to obtain room maintenance services and thus have lower variable cost of hosting. Hosts in cities with more favorable Airbnb regulations may also face lower variable cost of hosting. The level of mortgage pressure can further impact how long the hosts would like to list their property. Finally, the market-specific time fixed effect is specified as $\xi_{mT}^{Av} = \xi_0^{Av} \cdot 1\{T = 2017\} + \xi_1^{Av} \cdot (T - T_m^0)$, where T_m^0 represents the year when Airbnb reached 10% of the total rooms supplied by hotel and Airbnb in a city. The first component captures any yearly unobservables that affect all metro areas (e.g., because of Airbnb's national marketing) relative to those in the baseline year 2015. The second component captures any market-specific time trend that is related to Airbnb history, or how long Airbnb has been present in a city. For instance, Airbnb may be better received in markets where it has been present longer.

We summarize the covariates that enter availability and Airbnb variable cost in Columns 1 and 2 of Table 3. Note that there is no overlap between covariates in availability (a_{it}) and Airbnb variable cost (c_{it}^{Av}). The reason is that availability and Airbnb variable cost are identified form the same host decision (i.e., the second-stage decision of how many days to list). Any overlapping covariates will not be separately identified.

4.2 First Stage: Discrete Decision of Where to List

In the first stage, property owners choose among long-term rentals, Airbnb, and the outside option given the expected yearly profit from the second-stage decision for each option. Let d_{iT} denote the decision of property owner i in year T, and index the alternatives by superscripts A (Airbnb), R (long-term rental),

[14] Note that this specification does not restrict the cost to be non-negative. We keep the specification flexible to accommodate extreme cases, for instance, a host may derive positive utility (i.e., negative cost) from interacting with guests. In the case of a negative cost, the optimal number of days listed is bounded by $\bar{s}$, so that the property owners choose to list for the full month ($s^* = \bar{s}$). We also bound the monthly profit Π_{it}^A by $p_{it}^A \bar{s}$, which is the maximum profit that a listing can possibly generate.

and O (outside option). Property owners solve the following problem

$$\begin{aligned}
&\max_{d\in\{A,R,O\}} \Pi_{iT}^{d} \\
&\Pi_{iT}^{A} = \sum_{t\in T} \left(E\left[\Pi_{it}^{A}(s_{it}^{*})\right] \cdot a_{it} \right) - c_{iT}^{Af} + \epsilon_{iT}^{Af} \\
&\Pi_{iT}^{R} = p_{iT}^{R}\phi_{iT}^{R} - c_{iT}^{R} + \epsilon_{iT}^{R} \\
&\Pi_{iT}^{O} = \epsilon_{iT}^{O}
\end{aligned} \tag{5}$$

where Π_{iT}^{d} represents the yearly profit from each alternative $d \in \{A, R, O\}$. The idiosyncratic error terms $\left\{\epsilon_{iT}^{Af}, \epsilon_{iT}^{R}, \epsilon_{iT}^{O}\right\}$ are assumed to be i.i.d. extreme value type I errors with location parameter 0 and scale parameter σ_1 and independent of the second-stage error terms $\{\epsilon_{it}^{Av}\}$. The deterministic part of the profit of the outside option is normalized to zero. The profit of long-term rental comes from the ex ante expected yearly rent (p_{iT}^{R}) multiplied by the expected occupancy rate (ϕ_{iT}^{R}) minus the cost of long-term rental (c_{iT}^{R}). The profit of Airbnb comes from the sum of the ex ante monthly profit from Airbnb hosting $\left(\sum_{t\in T}\left(E\left[\Pi_{it}^{A}(s_{it}^{*})\right] \cdot a_{it}\right)\right)$ minus the fixed cost of Airbnb hosting (c_{iT}^{Af}). The ex ante monthly profit from Airbnb hosting is obtained by substituting the optimal number of days to list in Equation (3) into Equation (2), taking expectations over the error terms ϵ_{it}^{Av} in c_{it}^{Av}, and multiplying by the probability of being available a_{it}:

$$E\left[\Pi_{it}^{A}(s_{it}^{*})\right] \cdot a_{it} = \left[\int_{-\infty}^{\infty} \Pi_{it}^{A}\left(s_{it}^{*}\right) f\left(\epsilon_{it}^{Av}\right) d\epsilon_{it}^{Av}\right] \cdot a_{it} \tag{6}$$

We show in the online appendix how we compute the ex ante monthly profit, accounting for s_{it} being integers.

Property owners are heterogeneous in their fixed cost of Airbnb hosting and cost of long-term rental. The cost of long-term rental for property i in year T is:

$$c_{iT}^{R} = \bar{c}^{R} \cdot \frac{\exp(\beta^{R} X_{iT}^{R})}{1 + \exp(\beta^{R} X_{iT}^{R})} \tag{7}$$

where X_{iT}^{R} includes host demographics (age, education, income, marital status, gender), metro area characteristics (population, density, mortgage affordability index), and property characteristics (number of bedrooms/bathrooms/amenities, property type).

The fixed cost of Airbnb hosting for property i in year T is:

$$c_{iT}^{Af} = \bar{c}^{Af} + \beta^{Af} X_{iT}^{Af} + \xi_{mT}^{Af} \tag{8}$$

where $\bar{c}^{Af}$ is the baseline cost, ξ_{mT}^{Af} is a market-specific time fixed effect that captures any time-varying

Table 3: Summary of Covariates

	Cost				Revenue			
	Second stage		First stage		Hedonic regression			
	a_{it}	c_{it}^{Av}	c_{iT}^{Af}	c_{iT}^{R}	p_{it}^{A}	ϕ_{it}^{A}	p_{iT}^{R}	ϕ_{iT}^{R}
Property		yes	yes	yes	yes	yes	yes	yes
Demographics	yes		yes	yes	yes	yes	yes	yes
Metro area	yes		yes	yes	metro-month	metro-month	metro-year	metro-year
Season fixed effect	yes				fixed effect	fixed effect	fixed effect	fixed effect
Mortgage		yes	yes	yes				
Wage and employment in accomm. industry		yes	yes					
Airbnb regulation score		yes						
Airbnb history		yes	yes		yes	yes		
Tourism (air passengers)					yes	yes		
Total Airbnb supply					yes	yes		
Airbnb price						yes		
Total rental supply							yes	yes
Rent								yes

Notes: (1) Metro area characteristics include population and density. Property characteristics include number of bedrooms/bathrooms/amenities and property type: house (dummy). Demographics include age 35-65, age over 65, high school education, bachelor's education, 50-100k income, over 100k income, male, marital status never. The baseline demographics group is age below 35, education below high school, income below 50k, female, and married. (2) There is no overlap between covariates in availability (a_{it}) and Airbnb variable cost (c_{it}^{Av}). The reason is that availability and Airbnb variable cost are identified form the same host decision (i.e., the second-stage decision of how many days to list). Any overlapping covariates will not be separately identified.

unobservables, and X_{iT}^{Af} includes host demographics (age, education, income, marital status, gender), metro area characteristics (population, density, mortgage affordability index, average wage and employment in the accommodation industry), and property characteristics (number of bedrooms/bathrooms/amenities, property type). The market-specific time fixed effect is specified as $\xi_{mT}^{Af} = \xi_0^{Af} \cdot 1\{T = 2017\} + \xi_1^{Af} \cdot \left(T - T_m^0\right)$, where T_m^0 represents the year when Airbnb reached 10% of the total rooms supplied by hotel and Airbnb in a city. The first component captures any yearly unobservables that affect all metro areas and the second component captures any market-specific time trend that is related to how long Airbnb has been present in a city.

We summarize the covariates that enter Airbnb fixed cost and long-term rental cost in Columns 3 and 4 of Table 3.

4.3 Expectation on Revenue

Property owners' decisions depend on revenue information, i.e., rent, rental occupancy rate, Airbnb price, and Airbnb occupancy rate. We assume that property owners form expectations over these variables using a typical hedonic approach when making their first-stage decisions. Hedonic regression is a widely used method to estimate property value by decomposing a property's value into its constituent attributes and obtaining contributory values for each attribute (see Sirmans, Macpherson, and Zietz (2005) for a review on using hedonic models to estimate house prices). We use the hedonic approach because it offers the following three advantages. First, the hedonic model incorporates property heterogeneity, which allows us to construct expected revenues for each property in the data. It also parsimoniously capture how hosts set prices and how occupancy rates are determined in practice. Second, this approach allows us to obtain rent, rental occupancy rate, the Airbnb price, and the Airbnb occupancy rate regardless of how the units are utilized. In the data, we observe rent and rental occupancy only for long-term rental properties and observe Airbnb price and occupancy rate only for Airbnb properties. The property attributes, however, are observed for all properties. The hedonic model allows us to construct expected rent and rental occupancy for properties listed on Airbnb and the expected Airbnb price and occupancy rate for properties listed on the long-term rental market. The underlying assumption is that properties with similar attributes will have similar revenues. Third, the hedonic approach allows us to generate counterfactual rent, rental occupancy, Airbnb price, and Airbnb occupancy under counterfactual scenarios, which we discuss in detail in Section 7.2.

The hedonic models of rent and rental occupancy are

$$p_{iT}^{R} = \rho_0 + \rho_1 x_i^P + \rho_2 x_i^D + \rho_3 S_{mT}^R + \psi_{mT}^{Rp} + \varepsilon_{iT}^{Rp} \tag{9}$$

$$\phi_{iT}^{R} = \eta_0 + \eta_1 x_i^P + \eta_2 x_i^D + \eta_3 S_{mT}^R + \psi_{mT}^{Ro} + \eta_4 p_{iT}^R + \varepsilon_{iT}^{Ro} \tag{10}$$

where we regress the rent of property i in year T, p_{iT}^R, on property characteristics x_i^P, household demographics x_i^D, rental supply of comparable units in the metro area S_{mT}^R (measured as the number of comparable units that choose to list on the long-term rental market), and metro-year fixed effects ψ_{mT}^{Rp}.[15] Here, m denotes the metro area to which property i belongs. The hedonic model of the rental occupancy ϕ_{iT}^R uses the same specification except that it also includes rent as an additional regressor because the occupancy rate depends on the price. We run the regression using long-term rental properties with rents over 10th percentile and remove the outliers.

The rental supply of comparable units S_{mT}^R and the rent p_{iT}^R are potentially endogenous variables. We

[15]"Comparable" units are those with the same number of bedrooms. We conduct a robustness check by defining comparable units as those with the same numbers of bedrooms and bathrooms and obtain robust estimates. We keep the original definition because it produces a larger R-squared of the regressions.

address the endogeneity issue using BLP-type instruments. Specifically, we instrument the rental supply S^R_{mT} using the rental supply of non-comparable units in market m in year T (normalized by the total rental supply in market m in year T) and instrument the rent p^R_{iT} using average property characteristics of non-comparable rental properties in market m in year T. The instruments pass the weak IV test.[16] We jointly estimate the rent and rental occupancy equations as a system of equations using three-stage least squares (3SLS) to allow for correlation of the error terms in the two equations.

The hedonic models of Airbnb price and occupancy rate are

$$p^A_{it} = \delta_0 + \delta_1 x^P_i + \delta_2 x^D_i + \delta_3 S^A_{mt} + \delta_4 x^A_{mt} + \psi^{Ap}_{mt} + \varepsilon^{Ap}_{it} \tag{11}$$

$$\phi^A_{it} = \gamma_0 + \gamma_1 x^P_i + \gamma_2 x^D_i + \gamma_3 S^A_{mt} + \gamma_4 x^A_{mt} + \psi^{Ao}_{mt} + \gamma_5 p^A_{it} + \varepsilon^{Ao}_{it} \tag{12}$$

where we regress the monthly logged average nightly price of property i in month t, p^A_{it}, on property characteristics x^P_i, household demographics x^D_i, Airbnb supply of comparable units S^A_{mt} (measured as the number of days listed by all comparable units), Airbnb-related metro area variables x^A_{mt}, and metro area-specific year and month fixed effects ψ^{Ap}_{mt}. The Airbnb-related metro area variables include number of air passengers to the city, which can proxy for the heterogeneous tourism popularity across cities, and Airbnb history (measured as the number of months since Airbnb reached 10% of the total rooms supplied by hotel and Airbnb in a city), which can proxy for unobserved factors that relate to the length of Airbnb presence. The market-specific time fixed effects can capture market-specific seasonality patterns in Airbnb prices. The hedonic model of the occupancy rate uses the same specification except that it also includes the Airbnb price as an additional regressor because the occupancy rate depends on the price.

The Airbnb supply of comparable units S^A_{mt} and the Airbnb price p^A_{it} are potentially endogenous variables. We address the endogeneity issue by instrumenting Airbnb supply S^A_{mt} with the Airbnb supply of non-comparable units in market m in month t and instrumenting the Airbnb price p^A_{it} with average property characteristics of non-comparable Airbnb properties in market m in month t. In addition to these BLP-type instruments, we further include metro area-level Airbnb regulation score, rent-to-own ratio, and unemployment rate as instruments for Airbnb supply. These variables are valid instruments because they serve as cost shifters and affect the hosts' incentive to list their properties, so they are correlated with Airbnb supply; they do not affect tourists' incentives, so they do not directly affect Airbnb demand. The instruments pass the weak IV test.[17] We jointly estimate the Airbnb price and occupancy equations as a system of equations

[16] The instruments pass the weak IV test with the first-stage regression F-statistics of 5915.87 for rental supply in the rent regression, 9568.74 for rental supply in the rental occupancy regression, and 91.52 for rent in the rental occupancy regression.

[17] The instruments pass the weak IV test with the first-stage regression F-statistics of 1,674,186 for Airbnb supply in the Airbnb price regression, 2,489,759 for Airbnb supply in the Airbnb occupancy regression, and 29,684 for Airbnb price in the Airbnb occupancy regression.

using three-stage least squares (3SLS).

We summarize the covariates that enter each hedonic regression in Columns 5-8 of Table 3.

To generate the expected rent, rental occupancy rate, Airbnb price, and Airbnb occupancy rate for all properties, we first estimate the two systems of equations using the observed revenues and property attributes. Specifically, we use the observed long-term rental data from the AHS to estimate the hedonic model of rent and rental occupancy, and use the observed Airbnb data to estimate the hedonic model of Airbnb price and occupancy. Once we obtain the estimates, we can generate the expected rent, rental occupancy rate, Airbnb price, and Airbnb occupancy rate for all properties in the Airbnb and AHS data.[18] These expected revenues are used in the property owners' first-stage decision.

The regression results are provided in the online appendix. The coefficients have the expected signs. For example, both rent and Airbnb price increase with the number of bedrooms and bathrooms. An increase in the aggregate rental supply is associated with a reduction in rent and rental occupancy. A higher rent decreases rental occupancy. Similarly, an increase in the aggregate Airbnb supply is associated with a reduction in price and an increase in the occupancy rate.[19] A higher Airbnb price leads to a lower Airbnb occupancy rate.

5 Estimation Method

5.1 Estimation

We use the maximum likelihood estimation (MLE) method to estimate the model. The likelihood function for individual i is the joint probability of the individual's decision on the use of the property and, if the Airbnb option is chosen, the number of days to list on Airbnb:

$$l_i\left(\Theta|d_{iT}, s_{it}, \mathcal{X}_i\right) = \prod_{T\in\{2015,2017\}} \Pr\left(d_{iT} \mid \mathcal{X}_i\right) \cdot \left[\prod_{t\in T} \left(\Pr\left(s_{it} \mid \mathcal{X}_i\right)\right)\right]^{1(d_{iT}=A)}$$

where $\mathcal{X}_i$ contains all host demographics, metro area and property characteristics that affect the costs and revenues of individual i, $\Pr\left(d_{iT} \mid \mathcal{X}_i\right)$ is the probability of the first-stage decision given the observed characteristics $\mathcal{X}_i$, and $\Pr\left(s_{it} \mid \mathcal{X}_i\right)$ is the probability of the second-stage decision given $\mathcal{X}_i$ (derived below).

[18] One caveat is that the hedonic models for the Airbnb price and occupancy rate contain the variable for listing type (entire place, private room, shared room), which is not available for properties in the AHS data. We assume that they will be listed on Airbnb as the entire place rather than as private or shared rooms, as entire places are the most common type on Airbnb. The results are robust if we allow the properties to be listed as private rooms with a probability that equals the empirical fraction of private room listings in the data.

[19] The Airbnb supply can have two effects on the occupancy rate of a particular listing. First, a larger Airbnb supply can imply more competition and reduce the per-listing occupancy rate. Second, a larger Airbnb supply can attract more guests to the platform and increase the per-listing occupancy rate. The overall effect is an empirical question. We find that the second effect dominates the first effect, so more Airbnb supply leads to a higher occupancy rate.

One caveat is that, as discussed in Section 3.1, we do not observe host demographics for Airbnb properties. We use the zip code-level demographics distribution $f(\mathcal{X}_i)$ from the ACS data. For properties in the Airbnb data, the likelihood for individual i is integrated over the demographic distributions:

$$l_i\left(\Theta|d_{iT}, s_{it}, f(\mathcal{X}_i)\right) = \int_{\mathcal{X}_i} l_i\left(\Theta|d_{iT}, s_{it}, \mathcal{X}_i\right) f\left(\mathcal{X}_i\right) d\mathcal{X}_i$$

The total log-likelihood is $\log \mathcal{L}\left(\Theta\right) = \sum_i \log l_i\left(\Theta|d_{iT}, s_{it}, f(\mathcal{X}_i)\right)$. Given that we conduct a two-step estimation (i.e., estimate the hedonic regressions in the first step and the hosts' decisions in the second step), we correct the standard errors following Murphy and Topel (1985).

Derivation of the second-stage probability. $\Pr(s_{it}\,|\,\mathcal{X}_i)$ is constructed based on the feasible range of the normally distributed error term $\{\epsilon_{it}^{Av}\}$ implied by the optimal choices in Equation (3):

$$\begin{aligned}
Pr(s_{it} = 0\,|\,\mathcal{X}_i) &= a_{it} \cdot \left(\epsilon_{it}^{Av} > p_{it}^{A}\phi_{it}^{A} - (\bar{c}^{Av} + \beta^{Av} X_{iT}^{Av})\right) + (1 - a_{it}) \\
&= a_{it} \cdot \left(1 - \Phi\left(\frac{1}{\sigma_2}\left(p_{it}^{A}\phi_{it}^{A} - (\bar{c}^{Av} + \beta^{Av} X_{iT}^{Av})\right)\right)\right) + (1 - a_{it}) \\
Pr(s_{it} = s, 0 < s < \bar{s}\,|\,\mathcal{X}_i) &= a_{it} \cdot \left(\epsilon_{it}^{Av} = \frac{p_{it}^{A}\phi_{it}^{A}}{\exp(s/\bar{s})} - (\bar{c}^{Av} + \beta^{Av} X_{iT}^{Av})\right) \\
&= a_{it} \cdot \left(\phi\left(\frac{1}{\sigma_2}\left(\frac{p_{it}^{A}\phi_{it}^{A}}{\exp(s/\bar{s})} - (\bar{c}^{Av} + \beta^{Av} X_{iT}^{Av})\right)\right)\right) \\
Pr(s_{it} = \bar{s}\,|\,\mathcal{X}_i) &= a_{it} \cdot \left(\epsilon_{it}^{Av} < \frac{p_{it}^{A}\phi_{it}^{A}}{\exp(1)} - (\bar{c}^{Av} + \beta^{Av} X_{iT}^{Av})\right) \\
&= a_{it} \cdot \left(\Phi\left(\frac{1}{\sigma_2}\left(\frac{p_{it}^{A}\phi_{it}^{A}}{\exp(1)} - (\bar{c}^{Av} + \beta^{Av} X_{iT}^{Av})\right)\right)\right)
\end{aligned}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of the standard normal distribution.

Derivation of the first-stage probability. $\Pr\left(d_{iT}\,|\,\mathcal{X}_i\right)$ is constructed based on the feasible range of the i.i.d. extreme value type I error terms $\left\{\epsilon_{iT}^{Af}, \epsilon_{iT}^{R}, \epsilon_{iT}^{O}\right\}$ implied by the optimal choices in Equation (5). Define $a \equiv \sum_{t\in T}\left(E\left[\Pi_{it}^{A}(s_{it}^{*})\right] \cdot a_{it}\right) - c_{iT}^{Af}$ and $r \equiv p_{iT}^{R}\phi_{iT}^{R} - c_{iT}^{R}$. The probabilities are:

$$\begin{aligned}
\Pr\left(d_{iT} = A\,|\,\mathcal{X}_i\right) &= \frac{\exp\left(a/\sigma_1\right)}{1 + \exp\left(a/\sigma_1\right) + \exp\left(r/\sigma_1\right)} \\
\Pr\left(d_{iT} = R\,|\,\mathcal{X}_i\right) &= \frac{\exp\left(r/\sigma_1\right)}{1 + \exp\left(a/\sigma_1\right) + \exp\left(r/\sigma_1\right)} \\
\Pr\left(d_{iT} = O\,|\,\mathcal{X}_i\right) &= \frac{1}{1 + \exp\left(a/\sigma_1\right) + \exp\left(r/\sigma_1\right)}
\end{aligned}$$

5.2 Identification

We observe three types of information in the data: the revenues, the first-stage decision of the hosts, and the second-stage decision of the hosts. They are used to identify the revenue-side parameters in the hedonic regressions and the cost-side parameters in the hosts' decisions.

The revenue-side parameters are directly identified and obtained by regressing the observed revenues on the observed characteristics.

The cost-side parameters in the second-stage decision include the availability parameters β^a and Airbnb variable cost parameters $\{\bar{c}^{Av}, \beta^{Av}, \xi_0^{Av}, \xi_1^{Av}, \sigma_2\}$, which are identified from the Airbnb listing pattern, namely, the number of days that a property is listed on Airbnb in a given month. In particular, the availability parameters β^a are mainly identified from the bi-modal listing pattern of no listing and full listing and its variation across metro areas, host demographics and seasons. Conditional on availability, the average number of days listed and its variation across properties, metro areas, and seasons identifies the Airbnb variable cost parameters $\left\{\bar{c}^{Av}, \beta^{Av}\right\}$. The time-related parameters $\left\{\xi_0^{Av}, \xi_1^{Av}\right\}$ are identified from the listing pattern differences across years and markets with different lengths of Airbnb history.

The cost-side parameters in the first-stage decision include the long-term rental cost parameters $\{\bar{c}^R, \beta^R\}$, the Airbnb fixed cost parameters $\{\bar{c}^{Af}, \beta^{Af}, \xi_0^{Af}, \xi_1^{Af}\}$, and the standard deviation of the idiosyncratic shocks σ_1. The overall fraction of properties that choose long-term rental and its variation across metro areas, demographics, and properties identify the long-term rental cost parameters $\{\bar{c}^R, \beta^R\}$. Similarly, the fraction of properties that choose Airbnb and its variation across metro areas, demographics, properties, and over time identify the Airbnb fixed cost parameters $\{\bar{c}^{Af}, \beta^{Af}, \xi_0^{Af}, \xi_1^{Af}\}$.

These cost-side parameters are further identified from changes in host decisions over time as Airbnb becomes more available. Table 4 illustrates how these changes can happen. Conditional on the revenues, there are four possible combinations of long-term rental and Airbnb hosting costs, represented by the four cells in the table. Denote the number of hosts in cell i by n_i. Intuitively, without Airbnb presence, hosts with a high long-term rental cost choose the outside option ("O", as in cells 1 and 2) and hosts with a low long-term rental cost choose the long-term rental option ("R", as in cells 3 and 4). Therefore, the ratio of the number of hosts who choose "O" versus "R" without Airbnb presence ($\frac{n_1+n_2}{n_3+n_4}$) identifies the fraction of hosts with high versus low long-term rental costs. In the presence of Airbnb, hosts with a high Airbnb hosting cost will not change their options (O$\rightarrow$O as in cell 1 and R$\rightarrow$R as in cell 3), while hosts with a low Airbnb cost will switch to Airbnb (O$\rightarrow$A as in cell 2 and R$\rightarrow$A as in cell 4). Therefore, among hosts who originally choose "R", the ratio of the number of hosts whose switching behavior exhibits "R$\rightarrow$R" versus "R$\rightarrow$A" with Airbnb presence ($\frac{n_3}{n_4}$) identifies the fraction of hosts with high versus low Airbnb hosting costs. Similarly,

Table 4: Identification

		Airbnb Hosting Cost	
		High	Low
Long-Term	High	(1) O → O	(2) O → A
Rental Cost	Low	(3) R → R	(4) R → A

Note: O, R, A represent the outside option, long-term rental, and Airbnb, respectively.

this fraction can also be identified from the ratio of the number of hosts who originally choose "O" and whose switching behavior exhibits "O→O" versus "O→A" ($\frac{n_1}{n_2}$) with Airbnb presence.

Note that the cost-side parameters entering the first-stage decision can also be identified from the data on the second-stage decision, and vice versa, as the first and second stages are linked. The expected Airbnb profit from the second stage enters the first-stage decision, so the data on the first-stage decision impose over-identifying restrictions on the parameters in the second stage. Similarly, the identification of the parameters in the first stage is also affected by the second-stage parameters.

Exclusion restrictions. Property characteristics, host demographics, metro area characteristics enter both the cost-side components and the revenue-side regressions. Exclusion restrictions come from the non-overlapping variables. As summarized in Table 3, each cost component or revenue regression has exclusive variables that do not enter other components. For instance, aggregate supply of Airbnb and rental units affect only the revenues but not the costs: they affect the prices and occupancy rates through competition in the market, but they do not directly influence the hosting costs of an individual host. Also, tourism (number of air passengers to a city) only affects the revenue side because it captures the demand of tourists, while mortgage only affects the cost side because it influences the incentives of hosts. Finally, Airbnb-related variables (e.g., Airbnb history) only affect Airbnb and not long-term rental market. In general, the exclusion restrictions stem from the fact that renters and hosts face different trade-offs when making their decisions and that Airbnb and long-term rental market serve different consumers (tourists and local renters).

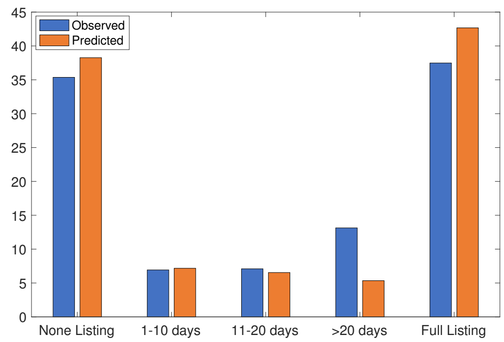
For the overlapping variables that appear in both the cost and the revenue sides, they are separately identified because we observe three types of information (the revenues, the first-stage decision of whether to list, and the second-stage decision of how long to list) and how they vary by the overlapping variables.[20] Consider the number of bedrooms as an example. It enters both the cost-side components and the revenue-side regressions. First, the revenue-side parameters on the number of bedrooms are directly identified from how the observed prices and occupancy rates change with the number of bedrooms. Second, conditional on the revenues, the cost-side parameters are identified from the variation in the observed first- and second-stage

[20]Note that there is no overlap between covariates in availability (a_{it}) and Airbnb variable cost (c_{it}^{Av}). The reason is that availability and Airbnb variable cost are identified form the same host decision (i.e., the second-stage decision of how many days to list). Any overlapping covariates will not be separately identified.

Table 5: Model Fit: First-Stage Decision

[%]	Airbnb	Rental	Outside
Observed	0.86	40.39	58.75
Predicted	0.86	37.82	61.32

Figure 6: Model Fit: Second-Stage Decision



decisions with respect to the number of bedrooms. Specifically, the parameter in the long-term rental cost is identified from how properties with different number of bedrooms differ in the fraction of choosing long-term rental in the first stage. The parameters in the Airbnb fixed cost and variable cost are separately identified if, for instance, properties with more bedrooms are more likely to choose Airbnb in the first stage but list shorter in the second stage; in this case, the coefficient on the number of bedrooms is negative in Airbnb fixed cost and positive in Airbnb variable cost.

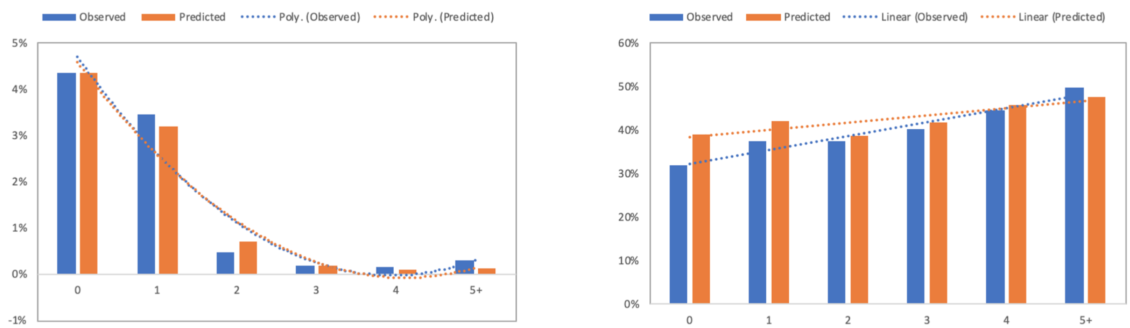
6 Estimation Results

6.1 Model Fit

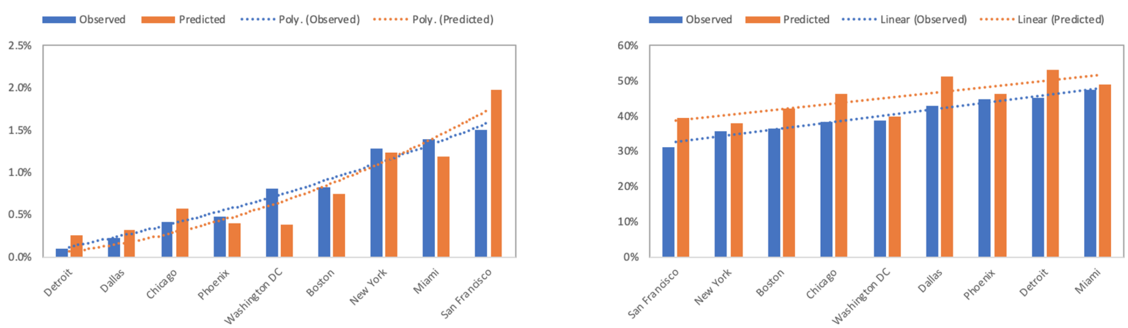
Table 5 shows the observed and predicted percentages of Airbnb, long-term rental, and outside option properties, and Figure 6 shows the observed and predicted Airbnb listing patterns. The model fits the first- and second-stage decisions well. It is also capable of fitting the heterogeneity for both decisions. Figure 7 presents the percentage of Airbnb properties for the first-stage model fit (left) and the percentage of unit-month observations that are listed for full month for the second-stage model fit (right), by property characteristics, metro area, and demographics. The model captures the data pattern: as the number of bedrooms increases, the percentage of Airbnb properties decreases except for from 4-bedroom units to 5+ bedroom units, and the percentage of full-month listing increases. The estimated percentages of the Airbnb properties and full month listings are also comparable to the observed percentages for each metro area and

Figure 7: Model Fit: By Property Characteristics, Metro Area, and Demographics

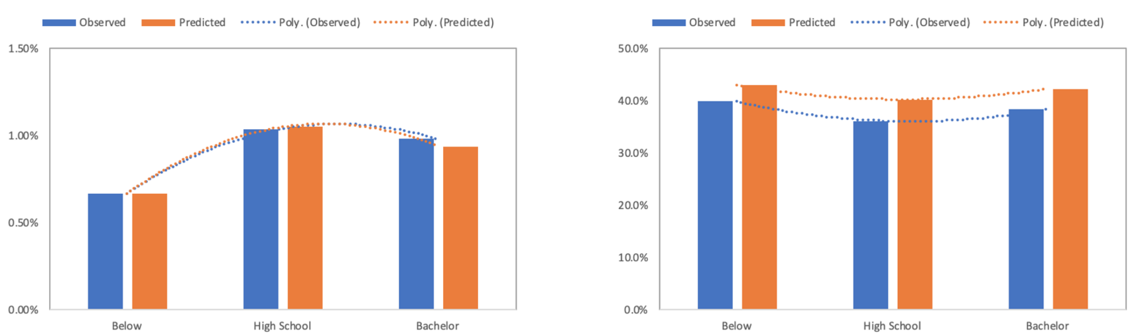
(a) By Property Characteristics (Number of Bedrooms)



(b) By Metro Area



(c) By Demographics (Education)



Note: The y-axis of the plots in the left column represents the percentage of Airbnb properties for the first-stage decision. The y-axis of the plots in the right column represents the percentage of unit-month observations that are listed for the full month for the second-stage decision. In Panel (b), metro areas are sorted by the observed percentages.

demographic group. Overall, these results suggest that the model can recover the heterogeneous costs of long-term rental and Airbnb hosting across property characteristics, metro areas, and demographics.

6.2 Parameter Estimates

Tables 6a and 6b report the parameter estimates for the first-stage and second-stage decisions.

Airbnb variable cost. The estimates of $\bar{c}^{Av}$ and β^{Av} suggest that the average variable cost of Airbnb hosting is \$30.6 per day, with a 25 percentile of \$22.8 and a 75 percentile of \$38.4. The estimates suggest that additional bedrooms, bathrooms, and facilities increase the variable cost of hosting. The daily cost for an entire place listing is \$19.9 larger than that of a private or shared room listing.

Note that these estimates are very comparable to the prices that third-party short-term rental cleaning services charge, which can serve as an out-of-sample validation for our estimates. For example, Tidy charges between \$40 and \$45 for cleaning a one-bedroom unit, which is comparable to our cost estimate of \$37.7 (equals the sum of our estimated baseline cost of \$34.0 and additional bedroom cost of \$3.7).[21]

Metro area characteristics also affect Airbnb variable cost and how long hosts list their properties. The estimated variable cost is lower and hosts list longer in cities with higher mortgage pressure, more employment and lower wage in the accommodations industry, and more favorable Airbnb regulation scores. The estimated variable cost is also lower in 2017 than in 2015.

Availability. The probability of being available varies by host demographics, metro area, and season. The estimates of β^a suggest that availability is higher for hosts who are younger, have high school education, have medium income, male, married and who live in cities with a larger population and a smaller density. Hosts are also more likely to be available in winter and less available in fall. Note that availability and Airbnb variable cost jointly determine how many days to list in a month. Hosts who are more available are more likely to be "part-time" or "full-time" hosts and less likely not to list.

Airbnb fixed cost. The estimates of $\bar{c}^{Af}$ and β^{Af} suggest that the median fixed cost of Airbnb hosting is \$6141 per month, with a 1 percentile of \$1363 and a 5 percentile of \$2052, which explains why only 0.86% of property owners choose Airbnb (note that the average expected revenue is \$1573 per month according to Table 1). The fixed cost can include the psychological cost of renting out property to transient guests and other tangible costs such as preparing property photos and descriptions and furnishings and amenities. The fixed cost can be quite large when, for example, property owners live in the property and need to move out, or they find it uncomfortable to rent their home to complete strangers, or they need to procure more furnishings and amenities to set up their properties as an Airbnb listing.

[21] See https://www.tidy.com/compare-house-cleaning-prices

Table 6: Parameter Estimates

(a) First-Stage Decision

		Est.	Std.			Est.	Std.
σ_1		9.343**	0.0283				
$\bar{c}^{Af}$		27.189**	0.1697	$\bar{c}^{R}$		62.347**	0.5729
	age: 35-65	3.056**	0.0066		age: 35-65	0.658**	0.0030
	age: over 65	6.965**	0.2826		age: over 65	1.959**	0.0022
	edu: high school	-1.813**	0.0013		edu: high school	0.432**	0.0032
	edu: bachelor	-1.249**	0.0056		edu: bachelor	1.124**	0.0018
	income: 50K-100K	-1.583**	0.0013		income: 50K-100K	0.406**	0.0051
	income: over 100K	-9.192**	0.2555		income: over 100K	1.132**	0.0075
	gender: male	-0.899**	0.0054		gender: male	0.133**	0.0113
	marital status: never	5.843**	0.2073		marital status: never	-0.0826**	3.106e-4
	# of bedrooms	17.574**	0.2305		# of bedrooms	0.376**	0.0056
β^{Af}	# of bathrooms	9.669**	0.2384	β^{R}	# of bathrooms	0.254**	0.0026
	# of amenities	-0.0445**	4.807e-4		# of amenities	0.472**	0.0024
	type: house (dummy)	4.950**	0.0155		type: house (dummy)	1.241**	0.0085
	studio (dummy)	13.910**	0.5063				
	population	0.571**	0.0070		population	-0.467**	0.0021
	density	-8.138**	0.3428		density	2.386**	0.0076
	mortgage	-0.223**	0.0083		mortgage	-0.705**	0.0034
	employment in accomm.	-2.688**	0.0047				
	wage in accomm.	0.0148**	5.780e-4				
	year fixed effect	-0.0485**	4.660e-4				
	Airbnb history	-3.574**	0.0049				

(b) Second-Stage Decision

		Est.	Std.			Est.	Std.
	constant	-4.455**	0.0029	σ_2		35.722**	2.730e-4
	age: 35-45	-0.566**	0.0045	$\bar{c}^{Av}$		34.028**	3.861e-4
	age: over 65	-1.865**	0.0047		# of bedrooms	3.681**	0.0011
	edu: high school	5.486**	0.0142		# of bathrooms	0.0051**	6.414e-4
	edu: bachelor	1.192**	0.0033		# of amenities	1.6397**	1.978e-4
	income: 50K-100K	0.341**	0.0052		entire room (dummy)	19.855**	0.0025
β^{a}	income: over 100K	-0.662**	0.0064		mortgage	-0.0221**	1.137e-4
	gender: male	0.221**	0.0037	β^{Av}	employment in accomm.	-4.143**	1.976e-4
	marital status: never	-2.698**	0.0037		wage in accomm.	0.347**	0.0011
	population	0.501**	6.686e-4		Airbnb regulation score	-0.0002**	3.455e-5
	density	-1.816**	0.0020		year fixed effect	-19.454**	0.0024
	summer	-0.603**	0.0033		Airbnb history	8.504**	0.0024
	fall	-1.134**	0.0030				
	winter	0.291**	0.0058				

Note: * and ** represent significance at the 5% and 1% levels. The following variables are in logged form: population, density, mortgage, Airbnb history. Employment in accommodations industry is in percentage and wage in accommodations industry is in $10k. The baseline demographics group is age below 35, education below high school, income below 50k, female, and married.

We find that Airbnb fixed cost is higher for properties with more bedrooms and bathrooms and is higher for a house than an apartment. The fixed cost is lower for property owners who are younger, have higher education level and more income, male, and married. In terms of metro area, the fixed cost is lower and property owners are more likely to choose Airbnb in cities with a smaller population, a larger density, and a longer Airbnb presence.[22] Property owners are also more likely to choose Airbnb in cities where mortgage is expensive, which might be because property owners leverage Airbnb as an additional income source to pay their mortgages. In fact, the primary use of the hosting income is to pay mortgages, according to a survey conducted by Airbnb.[23] Airbnb hosts can even use Airbnb income as a proof when applying for mortgage refinancing.[24] Finally, the Airbnb fixed cost is lower in cities with larger employment and a lower wage in the accommodations industry, as resources in the accommodations industry such as room cleaning can also be used for Airbnb hosting and may facilitate Airbnb hosting.

Long-term rental cost. The estimates of $\bar{c}^R$ and β^R suggest that the median fixed cost of long-term rental is \$2198 per month, with a 25 percentile of \$462 and a 75 percentile of \$4138, which explains why 40.4% of property owners choose long-term rental (note that the average expected revenue is \$1253 per month according to Table 1). Similar to the Airbnb fixed cost, the cost of long-term rental can come from both tangible costs and intangible psychological costs of renting out the unit for a year. In particular, the property owners can have a high cost in our model if they live in the unit and need to move out of the unit and find a new home to live in, or they need the unit for occasional self-use and find it difficult to commit to being available for rent for the full year. Alternatively, property owners can have a low cost in our model if they have multiple units and can easily rent out some of them. The estimates show that properties with more bedrooms, bathrooms, facilities, and being a house are more costly to rent. In terms of demographics and metro area, the long-term rental cost is lower and property owners are more likely to rent if they are younger, have lower education level and less income, female, married and if they live in cities with a larger population, a smaller density, and a higher mortgage pressure.

6.3 Cannibalization and Market Expansion

Airbnb can create both a negative impact of cannibalization and a positive impact of market expansion on the rental housing market. To evaluate the impact of Airbnb, we use the estimates to simulate the property owners' choices if there was no Airbnb. As previously illustrated in Table 4, hosts' decisions can be different

[22]The estimates suggest that density reduces Airbnb fixed cost but decreases availability. It means that hosts in cities with a large density are more likely to choose Airbnb in the first stage but list shorter in the second stage. This may be because hosts in these cities are more likely to use Airbnb to pay their mortgage while they are still living in the properties; although they are willing to list, their availability is limited.

[23]See https://www.airbnbcitizen.com/the-airbnb-community-in-seattle/

[24]See https://www.cnbc.com/2018/02/22/homeowners-are-using-airbnb-rental-income-to-refinance-mortgages.html

with and without Airbnb. Some hosts choose the outside option when Airbnb is not present and choose Airbnb when it becomes available (i.e., O $\rightarrow$ A in Table 4). These hosts represent the market expansion effect of Airbnb: they would not have listed on the long-term rental market and benefit from having Airbnb as an additional income source. Some hosts choose the long-term rental market when Airbnb is not present and choose Airbnb when it becomes available (i.e., R $\rightarrow$ A in Table 4). These hosts are switchers from the long-term rental market and represent the cannibalization effect of Airbnb or the reduction in long-term rental supply due to Airbnb.

Let D^{R0} and D^{R1} denote the number of long-term rental units without and with Airbnb. Let D^A denote the number of Airbnb units when Airbnb is present. Among Airbnb units, the number of switchers or cannibalization units is $D^{R0} - D^{R1}$ and the number of non-switchers or market expansion units is $D^A - (D^{R0} - D^{R1})$. We use two measures to evaluate Airbnb's impact. The first is the percentage change in rental supply due to Airbnb ($\frac{D^{R1}-D^{R0}}{D^{R0}}$), which captures the negative impact of Airbnb on the long-term rental market. The second is the percentage of Airbnb units that come from cannibalization, or the cannibalization rate, $\frac{D^{R0}-D^{R1}}{D^A}$. It represents the percentage of switchers among all (switchers and non-switchers) Airbnb units, which captures the relative sizes of the negative and positive impacts of Airbnb. Both measures are linked to the cost estimates of our model. As illustrated in Table 4, conditional on the revenues, hosts with a low (low) rental cost and a high (low) Airbnb hosting cost are more likely to remain in (leave) the long-term rental market when Airbnb is introduced, which speaks to the first measure. Hosts with a high (low) long-term rental cost and a low (low) Airbnb hosting cost are more likely to be non-switchers (switchers), which speaks to the second measure.

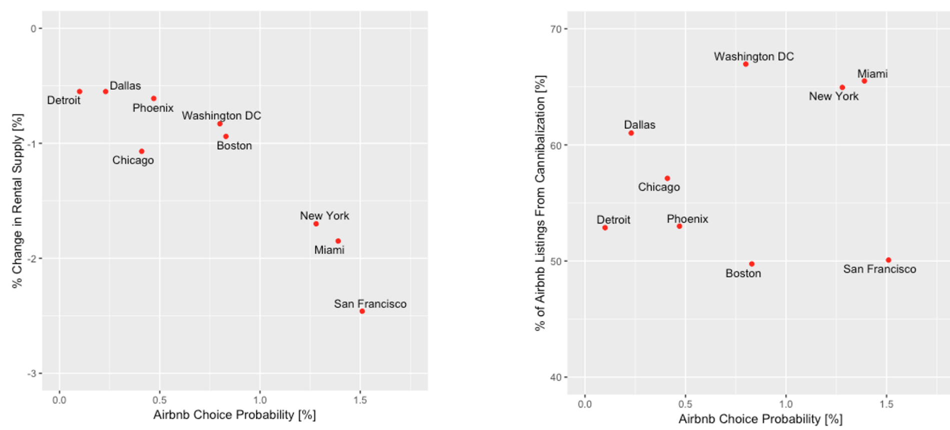
We first plot the percentage change in rental supply across metro areas in Figure 8a. We find that Airbnb causes a mild reduction in the rental supply, ranging from -0.55% in Dallas and Detroit to -2.46% in San Francisco. The reduction in the rental supply tends to be larger in metro areas where Airbnb is a popular choice for property owners.

However, the percentage change in the rental supply alone does not provide a holistic view of Airbnb's impact. We also need to consider the market expansion effect that Airbnb creates. We plot the cannibalization rate, or the percentage of switchers across metro areas in Figure 8b. We find that the percentage of switchers varies significantly, ranging from 49.8% in Boston to 67.0% in Washington D.C.. We find suggestive evidence that the cannibalization rate is higher in metro areas where 1) accommodations resources are abundant, 2) population size is large and density is high, 3) Airbnb has a long history, 4) city policies are friendly to short-term rentals, and 5) mortgage pressure is low.

Interestingly, although the reduction in the rental supply is larger in metro areas where Airbnb is popular, the cannibalization rate is not necessarily larger in these areas. For example, San Francisco has the highest

Figure 8: Cannibalization and Market Expansion: By Metro Area

(a) Change in Rental Supply (b) Cannibalization Rate



Airbnb popularity and suffers the most from rental supply reduction, but the percentage of switchers is among the lowest. It suggests that most of the Airbnb listings are from market expansion rather than cannibalizing the rental supply in San Francisco. City regulators need to thoroughly evaluate both the positive and negative impacts of Airbnb.

We also find suggestive evidence that Airbnb does raise affordable housing concerns as affordable units see a larger reduction in rental supply and a higher fraction of switchers. Table 7 presents the two measures (the percentage change in the rental supply $\frac{D^{R1}-D^{R0}}{D^{R0}}$ and the percentage of Airbnb units from cannibalization $\frac{D^{R0}-D^{R1}}{D^{A}}$) by property characteristics and demographics. In terms of property characteristics, the cannibalization impact is largely concentrated among lower priced, affordable units rather than among higher priced, luxurious ones. A basic studio or one-bedroom apartment is more likely to be taken off the long-term rental market than a house with multiple bedrooms and more amenities. In terms of demographics, the percentage of switchers is higher among young, low-income, never married, and male hosts.

Note that an observed “full-time” (“part-time”) listing does not necessarily imply cannibalization (market expansion). In other words, it is not appealing to assume, without modeling the hosts’ decisions, that all full-time hosts on Airbnb are switchers and should have been listed on the long-term rental market. Therefore, our structural model framework is helpful in recovering the underlying decision-making process of the hosts and identifying the actual potential switchers. Specifically, cannibalization occurs when property owners switch from long-term rental to Airbnb. Even if hosts list their properties on Airbnb full time, it would not be cannibalization if they would have chosen the outside option in the absence of Airbnb. For example, the data show that seniors (age 65+) are more likely to list full time, but the percentage of switchers among

Table 7: Cannibalization and Market Expansion: By Property Characteristics and Demographics

(a) Property Characteristics

[%]		$\frac{D^{R1}-D^{R0}}{D^{R0}}$	$\frac{D^{R0}-D^{R1}}{D^{A}}$	[%]		$\frac{D^{R1}-D^{R0}}{D^{R0}}$	$\frac{D^{R0}-D^{R1}}{D^{A}}$
# of Bedrooms	0	-4.30	78.70	# of Amenities	0	-2.28	79.51
	1	-3.08	70.21		1	-1.83	76.15
	2	-0.59	45.20		2	-1.77	79.05
	3	-0.15	19.40		3	-1.55	76.50
	4	-0.10	13.31		4	-1.54	69.74
	5+	-0.14	13.18		5	-1.26	58.75
# of Bathrooms	1	-1.94	66.87		6	-1.05	40.23
	2	-0.50	32.64		7	-0.49	16.58
	3	-0.27	30.30	Property Type	Apt	-1.67	73.71
	4	-0.13	11.60		House	-0.57	25.68
	5	-0.06	10.03	Listing Type	Entire Place	-1.34	60.25
	6	-0.12	11.96		Private Room	-3.65	59.78

(b) Demographics

[%]		$\frac{D^{R1}-D^{R0}}{D^{R0}}$	$\frac{D^{R0}-D^{R1}}{D^{A}}$	[%]		$\frac{D^{R1}-D^{R0}}{D^{R0}}$	$\frac{D^{R0}-D^{R1}}{D^{A}}$
Age	under 35	-1.35	76.50	Income	under 50K	-1.08	65.74
	35-65	-1.28	58.34		50K-100K	-1.24	61.77
	over 65	-1.55	47.31		over 100K	-2.32	53.42
Education	under	-1.02	60.85	Gender	male	-1.46	61.56
	high school	-1.88	59.13		female	-1.26	58.91
	bachelor's	-1.12	63.46	Marital Status	never married	-1.70	63.65
					other	-1.14	57.38

seniors is relatively low, suggesting that seniors' listings represent a large market expansion impact. On the other hand, part-time listings can be from cannibalization if they would have been in the long-term rental market in the absence of Airbnb. This is possible if the Airbnb profit is large enough to allow property owners to list part time and still earn more than listing in the long-term rental market.

7 Counterfactuals

This section evaluates the impact of a series of policies intended to ensure the supply and affordability of rental housing. First, we consider a set of short-term rental regulations on Airbnb such as imposing taxes or limiting the maximum number of days that a property can be listed. Second, we investigate rent control policy on long-term rental, in particular, how its impact can be affected by the presence of Airbnb.[25]

[25] In practice, Airbnb can affect rental housing affordability both by changing rental supply (i.e., the number of switchers) and rent, both of which are allowed to be endogenous in our counterfactual analysis. We focus on presenting the changes in rental supply in this section because the changes in rent are found to be very small (less than 1%). The reason is that the number of Airbnb properties, compared to owner-occupied and long-term rental properties, is still very small in both the data (0.86%) and the counterfactual analysis. Given the current market landscape, Airbnb's impact on long-term rent is limited; Airbnb mainly affects the long-term rental market by reducing rental supply rather than raising rental prices. The impact on rent could become significant if Airbnb accounts for a larger share of the market in the future.

7.1 Policy Implementation

Short-term rental regulation. We consider three types of short-term rental regulations on Airbnb. The first two regulations impose taxes. Currently, in the U.S., occupancy taxes are levied on facilities that provide transient rental rooms, for example, hotels, motels, and Airbnb. We consider two types of tax, linear and concave. First, the linear tax mimics the current policy that imposes a fixed percentage of the listing price as an occupancy tax. Let p_{it}^{A} denote the listing price paid by consumers and let $p_{it}^{A,host}$ denote the price received by hosts. The price paid by consumers p_{it}^{A} enters the hedonic regressions in Equations 11 and 12, while the price received by hosts $p_{it}^{A,host}$ enters the hosts' decisions in Equations 2 and 5. The prices and occupancy rates are determined such that $p_{it}^{A,host} = p_{it}^{A} - t_1 \cdot p_{it}^{A}$ in equilibrium, where t_1 is the tax rate and $0 < t_1 < 1$. Second, we propose a concave tax that charges a higher tax on affordable units and a lower tax on expensive units, which is motivated by our finding that the cannibalization is largely concentrated among lower priced, affordable units. Similar to the linear tax, we operationalize it as $p_{it}^{A,host} = p_{it}^{A} - t_2 \cdot \log(p_{it}^{A})$ in equilibrium.

The third short-term rental regulation is a listing restriction, which limits the maximum number of days that a property can be listed on short-term rental platforms. San Francisco, for example, allows hosts to rent out their unit as an entire place for up to 90 days per year. We simulate the case in which hosts are able to list up to a certain number of months in a year. The counterfactual analysis is operationalized as follows: in the second stage, we calculate the optimal number of days to be listed per month. We allow the hosts to choose the months that have the highest expected profits, up to the pre-specified maximum number of months. Based on the total ex ante expected profit from the chosen months, they choose among Airbnb, long-term rental, and the outside option in the first stage.

Rent control. Rent control is a system of laws placing a maximum price, or a "rent ceiling," on what landlords may charge tenants. It covers a spectrum of regulations that can vary from setting the absolute amount of rent that can be charged with no allowed increases to placing different limits on the amount that rent can increase. These restrictions may continue between tenancies or may be applied only within the duration of a tenancy. As of March 2019, the states of California, Maryland, New Jersey, New York, and Oregon, and the city of Washington D.C. have some rent control or stabilization policies on the books, while 37 states prohibit or ban rent control outright.[26]

Economists have concluded that rent controls are destructive. According to a 1990 poll of 464 economists, 93% of U.S. respondents agreed, either completely or with provisos, that "a ceiling on rents reduces the quantity and quality of housing available" (Alston, Kearl, and Vaughan 1992). We argue that the negative

[26] See https://www.curbed.com/2019/3/8/18245307/rent-control-oregon-housing-crisis

impact of rent control policy can be exacerbated when another profitable option for hosts, Airbnb, is available. We illustrate how the presence of Airbnb affects the impact of the rent control policy by simulating the policy outcomes with and without Airbnb. To operationalize the rent control policy, we assume that the rent is capped at $r\%$ below the observed rent, where $r\%$ can mimic the type of rent control that limits the maximum percentage of rent increase from the previous year.

7.2 Equilibrium

When implementing the counterfactual policies, it is important to allow the rent, rental occupancy rate, Airbnb price and occupancy rate to endogenously change in the new equilibrium according to the hedonic models outlined in Section 4.3. Specifically, given different counterfactual policies, the number of properties and the types of properties that choose long-term rental and Airbnb can change. The new characteristics and the new aggregate Airbnb and rental supply enter the hedonic models and generate a new set of expectations on rent, rental occupancy rate, Airbnb price and occupancy rate. The equilibrium is defined such that the set of Airbnb price, Airbnb occupancy rate, aggregate Airbnb supply, rent, rental occupancy rate, and aggregate rental supply $\{p_{it}^A, \phi_{it}^A, S_{mt}^A, p_{iT}^R, \phi_{iT}^R, S_{mT}^R\}$ reach a fixed point. The numerical algorithm to solve for the equilibrium is as follows:

1. Let superscript (k) denote the k-th iteration. Begin with the aggregate Airbnb supply $S_{mt}^{A(k)}$ and aggregate rental supply $S_{mT}^{R(k)}$. Given $S_{mt}^{A(k)}$ and $S_{mT}^{R(k)}$, construct the expected rent $p_{iT}^{R(k+1)}$, rental occupancy rate $\phi_{iT}^{R(k+1)}$, Airbnb price $p_{it}^{A(k+1)}$, and Airbnb occupancy rate $\phi_{it}^{A(k+1)}$ for each property using the hedonic models in Equations (9), (10), (11), and (12).

2. Given the updated $p_{it}^{A(k+1)}$, $\phi_{it}^{A(k+1)}$, $p_{iT}^{R(k+1)}$, and $\phi_{iT}^{R(k+1)}$, solve the property owners' problem under each counterfactual policy. Compute the updated aggregate Airbnb supply $S_{mt}^{A(k+1)}$ and aggregate rental supply $S_{mT}^{R(k+1)}$.

3. Check for the convergence of $\left|p_{it}^{A(k+1)} - p_{it}^{A(k)}\right|$, $\left|\phi_{it}^{A(k+1)} - \phi_{it}^{A(k)}\right|$, $\left|S_{mt}^{A(k+1)} - S_{mt}^{A(k)}\right|$, $\left|p_{iT}^{R(k+1)} - p_{iT}^{R(k)}\right|$, $\left|\phi_{iT}^{R(k+1)} - \phi_{iT}^{R(k)}\right|$, and $\left|S_{mT}^{R(k+1)} - S_{mT}^{R(k)}\right|$. If convergence is not achieved, return to Step 1.

We initialize the algorithm using the observed aggregate Airbnb supply and aggregate rental supply. Varying the initialization point produces robust results.

7.3 Result Analysis

7.3.1 Short-Term Rental Regulations

Overall policy impact. Figure 9a shows the effect of the short-term rental regulations by plotting the number of switchers (cannibalization) on the x-axis and the number of non-switchers (market expansion) on the y-axis. Each line represents one type of regulation, and each point on the line represents a particular level of regulation. For example, the level of regulation for the maximum month limit varies from 12 months to 3 months and the level of regulation for the linear tax rate varies from 0% to 90%. Arrow (a) indicates the direction of stricter regulation, for example, a higher tax rate and lower number of months allowed to list. Comparing different levels of regulation within each policy, we find that there is a trade-off in terms of choosing the level of regulation: stricter regulations help reduce the number of switchers (cannibalization), but they also reduce the number of non-switchers (market expansion).

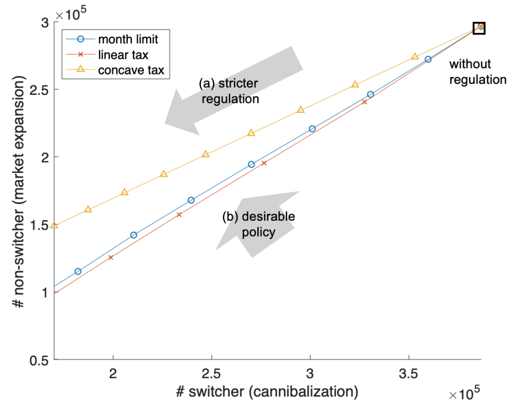
A desirable policy should reduce the negative impact of Airbnb (switcher or cannibalization) while maintain the positive impact of Airbnb (non-switcher or market expansion). Therefore, we examine three measures when comparing across policies: (1) the number of switchers and non-switchers; (2) cannibalization rate, or the fraction of switchers among all (switchers and non-switchers) listings; (3) fraction of affordable units (studio and one-bedrooms) among switchers. The last measure is particularly relevant as in practice the short-term rental regulations were launched to prevent affordable units from switching and causing affordable housing issues for local residents.

We find that our proposed policy of the concave tax is the most desirable among the three short-term rental regulations. The month limit is the second-best policy and the linear tax is the worst. First, a desirable policy should be able to maintain the number of non-switchers while reducing the number of switchers (i.e., measure 1), which is indicated by Arrow (b) in Figure 9a. The concave tax performs better from this perspective as shown in Figure 9a. Second, a desirable policy should reduce the cannibalization rate (i.e., measure 2). As shown in Figure 9b, the concave tax induces a lower cannibalization rate while the other two policies induce a higher cannibalization rate as the policies become stricter. Third, a desirable policy should minimize the fraction of affordable units among switchers (i.e., measure 3). As shown in Figure 9c, the concave tax can keep the fraction of affordable units at a relatively low level, while the other two policies induce a higher fraction of affordable units among switchers.

Differential impact on hosts. Besides the overall policy impact, we also examine how the policies differentially affect heterogeneous host groups. In particular, the positive impact of Airbnb is that it benefits the non-switchers by providing an alternative income source. Those switchers would not have listed on the long-term rental market without Airbnb presence. Imposing the regulations can have a redistribution effect

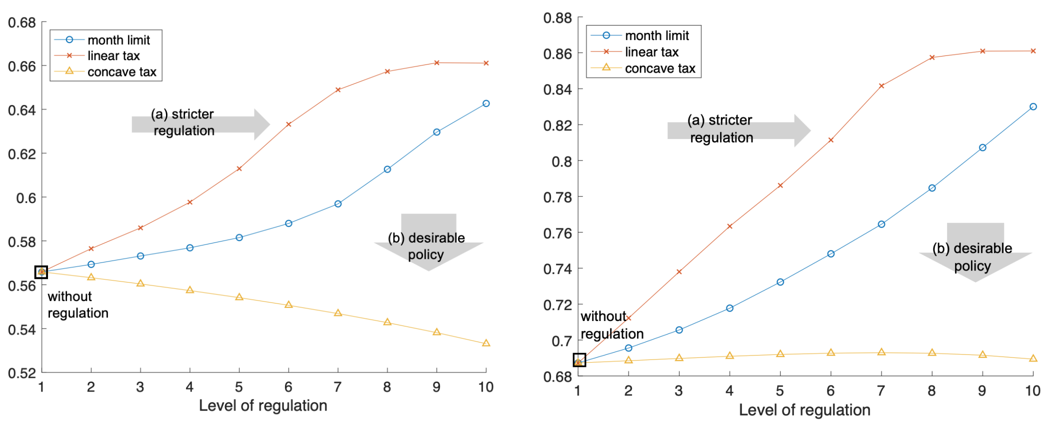
Figure 9: Short-Term Rental Regulations

(a) Number of Switchers and Non-Switchers



(b) Cannibalization Rate

(c) Fraction of Affordable Units Among Non-Switchers

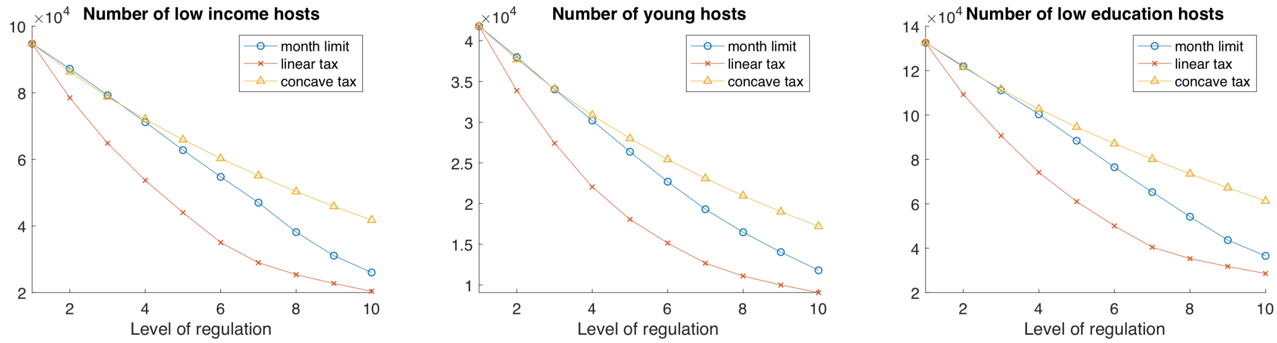


among hosts and impact who can continue to benefit from Airbnb, thereby influencing social inequality.

A desirable policy, besides defined by measures (1)-(3), should also allow economically disadvantaged hosts to continue benefiting from Airbnb. We define an additional measure of the policy desirability: (4) the number of economically disadvantaged hosts among non-switchers. In Figure 10, we plot the numbers of low income hosts (income less than 50k), young hosts (age below 35), and low education hosts (education equals or below high school) among non-switchers. As the level of regulation increases or the regulation becomes stricter, there are fewer non-switchers under all policies, which is consistent with our previous finding in Figure 9a. Comparing across policies, however, we find that the concave tax again performs the best in terms of having the largest number of non-switchers with low income, age, or education. The month limit again performs the second best and the linear tax performs the worst.

Overall, our proposed policy of the concave tax outperforms the other two policies in terms of all four measures: (1) maintaining the number of non-switchers while reducing the number of switchers, (2) reducing

Figure 10: Short-Term Rental Regulations: Differential Impact on Non-Switching Hosts



cannibalization rate, (3) reducing the fraction of affordable units among switchers, and (4) maintaining the number of economically disadvantaged hosts among non-switchers. The month limit appears to perform better than the linear tax. The concave tax performs the best because the cannibalization impact is largely concentrated among lower priced, affordable units rather than among higher priced, luxurious ones. The concave tax discourages lower price properties from being taken off the long-term rental market, which helps to limit cannibalization, but has less influence on higher priced properties, which helps maintain market expansion.

7.3.2 Long-Term Rental Regulations: Rent Control

To examine how Airbnb and rent control policy affect each other, we simulate market outcomes under four scenarios: (a) there is no rent control policy, and Airbnb is not available; (b) rent is controlled, and Airbnb is not available; (c) there is no rent control policy, and Airbnb is available; and (d) rent is controlled, and Airbnb is available. Comparing a (c) and b (d) suggests a negative impact of rent control in the absence (presence) of Airbnb. Comparing a (b) and c (d) suggests a negative impact of Airbnb in the absence (presence) of rent control. Importantly, we find that Airbnb and rent control can exacerbate the negative impact of each other.

First, we find that the presence of Airbnb can amplify the negative impact of rent control. In Table 8, the first column shows the percentage decrease in the rental supply due to rent control in the absence of Airbnb, and the second column shows the percentage when Airbnb is present. Consistent with the near-consensus among economists discussed above, we find that the rent control policy reduces rental supply. Importantly, this negative impact of the rent control policy is exacerbated when Airbnb is available; the reduction in rental supply due to rent control is larger with than without Airbnb. This exacerbating effect is more prominent when rent control policy is stricter. The reason is that Airbnb provides property owners an alternative option in addition to listing on the long-term rental market. When faced with the rent control policy, more

Table 8: Impact of Airbnb on the Negative Effect of Rent Control Policy

	% Change in Rental Supply Due to Rent Control	
Level of Rent Control (r)	without Airbnb	with Airbnb
2.5%	-1.376	-1.425
5.0%	-2.700	-2.795
7.5%	-3.973	-4.113
10.0%	-5.197	-5.380
12.5%	-6.373	-6.597
15.0%	-7.501	-7.764
17.5%	-8.585	-8.886
20.0%	-9.628	-9.964

Table 9: Impact of Rent Control Policy on the Negative Effect of Airbnb

Level of Rent Control (r)	None	2.5%	5.0%	7.5%	10.0%	12.5%	15.0%	17.5%	20.0%
% Change in Rental Supply Due to Airbnb	-2.05	-2.09	-2.14	-2.19	-2.23	-2.28	-2.32	-2.37	-2.41

property owners quit long-term rental and switch to Airbnb.

Second, we find that the presence of rent control can also amplify the negative impact of Airbnb. Table 9 shows the percentage decrease in rental supply induced by Airbnb, under varying strictness of rent control. The percentage reduction in rental supply is larger when a rent control policy is in effect and increases as the rent control policy becomes stricter.

Overall, the presence of Airbnb and rent control policy each can have a negative impact on the long-term rental supply. We find that when they are jointly present, they can exacerbate the negative impact of each other. Policymakers need to take caution when implementing rent control policy in the presence of Airbnb.

8 Conclusion

We investigate how Airbnb affects rental supply and affordability and provide policy implications for short-term rental regulations and long-term rent control. We model property owners' decisions in two stages: (1) the yearly decision on the usage of their properties among Airbnb, long-term rental, and neither and (2) the monthly decision on how many days to list on Airbnb, if they choose to list on Airbnb in the first stage. Given the revenue data on rent, rental occupancy rate, Airbnb price, and Airbnb occupancy rate, we estimate the costs of long-term rental and Airbnb hosting.

We find that Airbnb mildly cannibalizes the rental market but has a market expansion effect. The percentage of switchers varies significantly across cities. The cannibalization impact is largely concentrated among lower priced, affordable units rather than among higher priced, luxurious ones, suggesting that Airbnb can raise concerns about housing affordability. Metro areas where Airbnb is popular (e.g., San Francisco, New York, and Miami) experience a larger reduction in long-term rental supply due to Airbnb, but some of them benefit more from a larger market expansion effect, suggesting that the percentage of switchers is not necessarily larger in these cities.

The counterfactual results suggest that short-term rental regulations help reduce cannibalization, but they also reduce market expansion. We assess commonly used regulations such as limiting the number of days that a property can be listed and a linear tax and propose a new concave tax that charges a higher tax on affordable units. We show that the proposed concave tax outperforms the days limit, which further outperforms the linear tax according to four measures of policy desirability: (1) maintain the number of non-switchers while reduce the number of switchers, (2) reduce cannibalization rate, (3) prevent affordable units from switching, and (4) allow economically disadvantaged hosts (e.g., low income, young, or low education hosts) to benefit from Airbnb. Finally, rent control must be implemented with greater caution when Airbnb is available, as lower profits from long-term rental can cause landlords to switch to Airbnb and exacerbate the side effect of the rent control policy.

There are a few limitations to this study that represent directions for future research. First, we assume away the case in which long-term rental tenants sublet on Airbnb because we cannot observe whether an Airbnb host is a tenant or a property owner. This type of case may be relatively rare, as lease agreements often include clauses that restrain sublets. Services such as SubletAlert.com and SubletSpy also help landlords find tenants who have violated the agreement. Future research may extend the proposed model to incorporate the cases in which tenants sublet on Airbnb if such data are available.

Second, we do not explicitly model the competition between hotels and Airbnb. The hedonic models on Airbnb price and occupancy rate are estimated conditional on the observed competitive landscape between hotels and Airbnb. The implicit assumption is that hotels in the counterfactual analysis follows the same strategy as they do in the observed scenario. The equilibrium we solve for can be regarded as a partial equilibrium without hotel responses. We believe that the trade-off between long-term rental and Airbnb is the first-order effect for property owners, who are the key players in studying the impact of Airbnb on the long-term rental market. In addition, hotels do not appear to have responded to Airbnb in practice, accordingly to Li and Srinivasan (2019), who study hotel and Airbnb competition. In the future, when hotels have systematically responded to Airbnb, researchers can incorporate hotel responses into our proposed framework.

References

[1] Alston, R. M., Kearl, J. R., & Vaughan, M. B. (1992). Is There a Consensus Among Economists in the 1990's? *American Economic Review*, 82(2), 203-209.

[2] Barron, K., Kung, E., & Proserpio, D. (2019). The Sharing Economy and Housing Affordability: Evidence from Airbnb. Forthcoming, *Marketing Science.*

[3] Berger, T., Chen, C., & Frey, C. B. (2018) Drivers of Disruption? Estimating the Uber Effect. *European Economic Review*, 110, 197-210.

[4] Burtch, G., Carnahan, S., & Greenwood, B. N. (2018). Can You Gig It? An Empirical Examination of the Gig Economy and Entrepreneurial Activity. *Management Science*, 64(12), 5461-5959.

[5] Dowling, K., Manchanda, P. and Spann, M. (2019) The Existence and Persistence of the Pay-Per-Use Bias in Car Sharing Services. Available at SSRN: https://ssrn.com/abstract=3204233.

[6] Einav, L., Farronato, C., & Levin, J. (2016). Peer-to-Peer Markets. *Annual Review of Economics*, 8(1), 615-635.

[7] Gong, J., Greenwood, B. N., & Song, Y. (2017). Uber Might Buy Me a Mercedes Benz: An Empirical Investigation of the Sharing Economy and Durable Goods Purchase. Available at SSRN: https://ssrn.com/abstract=2971072.

[8] Greenwood, B. & Wattal, S. (2017). Show Me the Way to Go Home: An Empirical Investigation of Ride-sharing and Alcohol Related Motor Vehicle Fatalities. *MIS Quarterly*, 41(1), 163-187.

[9] Gurran, N. & Phibbs, P. (2017). When Tourists Move In: How Should Urban Planners Respond to Airbnb? *Journal of the American Planning Association*, 83(1), 80-92.

[10] Horn, K. & Merante, M. (2017). Is Home Sharing Driving Up Rents? Evidence From Airbnb in Boston. *Journal of Housing Economics*, 38, 14-24.

[11] Jiang, B. & Tian, L. (2018). Collaborative Consumption: Strategic and Economic Implications of Product Sharing. *Management Science*, 64(3), 983-1476.

[12] Koster, H., van Ommeren, J. and Volkhausen, N., (2019) Short-Term Rentals and the Housing Market: Quasi-Experimental Evidence From Airbnb in Los Angeles. CEPR Discussion Paper No. DP13094. Available at SSRN: https://ssrn.com/abstract=3226869

[13] Lee, D. (2016). How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendations. *Harvard Law and Policy Review*, 10(1), 229-253.

[14] Li, J., Moreno, A., & Zhang, D. (2016). Pros vs Joes: Agent Pricing Behavior in the Sharing Economy (August 28, 2016). Ross School of Business Paper No. 1298. Available at SSRN: https://ssrn.com/abstract=2708279.

[15] Li, H. & Srinivasan, K. (2019). Competitive Dynamics in the Sharing Economy: An Analysis in the Context of Airbnb and Hotels. *Marketing Science*, 38(3), 2019; 365-391

[16] Murphy, K. and R. Topel (1985). Estimation and Inference in Two Step Econometric Models. *Journal of Business and Economic Statistics* 3, 370–9.

[17] Sheppard, S. & Udell, A. (2018). Do Airbnb Properties Affect House Prices? Working paper.

[18] Sirmans, G. S., Macpherson, D. A., & Zietz, E. N. (2005). The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature*, 13(1), 1-44.

[19] Tian, L. & Jiang, B. (2017). Effects of Consumer-to-Consumer Product Sharing on Distribution Channel. P*roduction and Operations Management*, 27(2), 350-367.

[20] 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.

[21] Zhang, S., Mehta, N., Singh, P., & Srinivasan, K. (2018). Do Lower-Quality Images Lead to Greater Demand on Airbnb? Working paper.

Appendix

A. Restricting the Number of Days to be Integers

Let $u(s) \equiv \frac{p_{it}^{A}\phi_{it}^{A}}{\bar{s}(\exp(s/\bar{s})-\exp((s-1)/\bar{s}))} - (\bar{c}^{Av} + \beta^{Av} X_{iT}^{Av})$ and $l(s) \equiv \frac{p_{it}^{A}\phi_{it}^{A}}{\bar{s}(\exp((s+1)/\bar{s})-\exp(s/\bar{s}))} - (\bar{c}^{Av} + \beta^{Av} X_{iT}^{Av})$. Taking into account the fact that the number of days to list the property on Airbnb is integer, the second-stage probabilities are

$$\begin{aligned}
\Pr(s_{it} = 0) &= \Pr\left(\Pi_{it}^{A}(0) > \Pi_{it}^{A}(1)\right) = \Pr\left(\epsilon_{it}^{Av} > l(0)\right) = 1 - \Phi\left(\frac{l(0)}{\sigma_2}\right) \\
\Pr(s_{it} = s\ (s = 1, 2, ..., \bar{s})) &= \Pr\left(\Pi_{it}^{A}(s) > \Pi_{it}^{A}(s-1) \text{ and } \Pi_{it}^{A}(s) > \Pi_{it}^{A}(s+1)\right) \\
&= \Pr\left(l(s) < \epsilon_{it}^{Av} < u(s)\right) = \Phi\left(\frac{u(s)}{\sigma_2}\right) - \Phi\left(\frac{l(s)}{\sigma_2}\right) \\
\Pr(s_{it} = \bar{s}) &= \Pr\left(\Pi_{it}^{A}(\bar{s}) > \Pi_{it}^{A}(\bar{s}-1)\right) = \Pr\left(\epsilon_{it}^{Av} < u(\bar{s})\right) = \Phi\left(\frac{u(\bar{s})}{\sigma_2}\right)
\end{aligned}$$

Given the optimal number of days to list the property on Airbnb, the ex ante monthly profit from Airbnb hosting is

$$E\left[\Pi_{it}^{A}(s_{it}^{*})\right] \cdot a_{it} = \left[\int_{-\infty}^{\infty} \Pi_{it}^{A}\left(s_{it}^{*}\right) f\left(\epsilon_{it}^{Av}\right) d\epsilon_{it}^{Av}\right] \cdot a_{it}$$

where the integral is expanded as

$$\begin{aligned}
\int_{-\infty}^{\infty} \Pi_{it}^{A}\left(s_{it}^{*}\right) f\left(\epsilon_{it}^{Av}\right) d\epsilon_{it}^{Av} = &\int_{l(0)}^{\infty} \Pi_{it}^{A}(0) f(\epsilon_{it}^{Av}) d\epsilon_{it}^{Av} \\
&+ \int_{l(1)}^{u(1)} \Pi_{it}^{A}(1) f(\epsilon_{it}^{Av}) d\epsilon_{it}^{Av} \\
&+ ... \\
&+ \int_{l(\bar{s}-1)}^{u(\bar{s}-1)} \Pi_{it}^{A}((\bar{s}-1)) f(\epsilon_{it}^{Av}) d\epsilon_{it}^{Av} \\
&+ \int_{-\infty}^{u(\bar{s})} \Pi_{it}^{A}(\bar{s}) f(\epsilon_{it}^{Av}) d\epsilon_{it}^{Av}
\end{aligned}$$

Here, the first term for the interval with $s_{it}^{*} = 0$ is zero:

$$\int_{l(0)}^{\infty} \Pi^{A}(0) f(\epsilon_{it}^{Av}) d\epsilon_{it}^{Av} = \int_{l(0)}^{\infty} 0 \cdot f(\epsilon_{it}^{Av}) d\epsilon_{it}^{Av} = 0$$

The terms for the intervals with $s_{it}^{*} = s\ (s = 1, 2, ..., \bar{s})$ is computed as:

$$
\begin{aligned}
&\int_{l(s)}^{u(s)} \left[p_{it}^{A}\phi_{it}^{A}s - c_{it}^{Av}\bar{s}\left(\exp\left(\frac{s}{\bar{s}}\right)-1\right)\right] f(\epsilon_{it}^{Av})d\epsilon_{it} \\
&= \left[p_{it}^{A}\phi_{it}^{A}s - \bar{s}\left(\bar{c}^{Av}+\beta^{Av}X_{iT}^{Av}\right)\left(\exp\left(\frac{s}{\bar{s}}\right)-1\right)\right]\int_{l(s)}^{u(s)} f(\epsilon_{it}^{Av})d\epsilon_{it}^{Av} \\
&\quad - \left[\bar{s}\left(\exp\left(\frac{s}{\bar{s}}\right)-1\right)\right]\int_{l(s)}^{u(s)} \epsilon_{it}^{Av} f(\epsilon_{it}^{Av})d\epsilon_{it}^{Av} \\
&= \left[p_{it}^{A}\phi_{it}^{A}s - \bar{s}\left(\bar{c}^{Av}+\beta^{Av}X_{iT}^{Av}\right)\left(\exp\left(\frac{s}{\bar{s}}\right)-1\right)\right]\left[\Phi\left(\frac{u(s)}{\sigma_2}\right)-\Phi\left(\frac{l(s)}{\sigma_2}\right)\right] \\
&\quad - \left[\bar{s}\left(\exp\left(\frac{s}{\bar{s}}\right)-1\right)\right]\left[-\frac{\sigma_2}{\sqrt{2\pi}}\left(\exp\left(-\frac{u(s)^2}{2\sigma_2^2}\right)-\exp\left(-\frac{l(s)^2}{2\sigma_2^2}\right)\right)\right]
\end{aligned}
$$

For the last term for the interval with $s^* = \bar{s}$, recall that $\Pi_{it}^{A}(\bar{s})$ is bounded by the maximum possible profit, $p_{it}^{A}\bar{s}$.

$$
\begin{aligned}
&\int_{-\infty}^{u(\bar{s})} \Pi_{it}^{A}(\bar{s}) f(\epsilon_{it}^{Av})d\epsilon_{it}^{Av} \\
&= \int_{-\infty}^{\frac{p_{it}^{A}(\phi_{it}^{A}-1)}{\exp(1)-1}-(\bar{c}^{Av}+\beta^{Av}X_{iT}^{Av})} \left[p^{A}\bar{s}\right] f(\epsilon_{it}^{Av})d\epsilon_{it}^{Av} \\
&\quad + \int_{\frac{p_{it}^{A}(\phi_{it}^{A}-1)}{\exp(1)-1}-(\bar{c}^{Av}+\beta^{Av}X_{iT}^{Av})}^{u(\bar{s})} \left[p_{it}^{A}\phi_{it}^{A}\bar{s} - c_{it}^{Av}\bar{s}\left(\exp(1)-1\right)\right] f(\epsilon_{it}^{Av})d\epsilon_{it}^{Av}
\end{aligned}
$$

where the integrals are computed similarly as in the other terms for the intervals with $s_{it}^* = s$ ($s = 1, 2, ..., \bar{s}$).

B. Hedonic Regression Results

Tables 10 and 11 show the hedonic regression results for rent and rental occupancy rate. Tables 12 and 13 show the hedonic regression results for Airbnb price and occupancy rate. The R-squared value for Airbnb occupancy rate is relatively low, which may be due to a large variation in occupancy rate over time even within the same property. In fact, Airbnb occupancy rate seems to be quite random at the individual property level. Analysis of within- and across-property variation shows a large within-property variation across months, and including market-specific month fixed effects in the regression does not explain the large within-property variation. However, the model-predicted Airbnb occupancy rate is consistent with average occupancy rate for each property. In other words, hosts are on average correct in predicting their occupancy rates, which is more important when they make the first-stage decisions.

Table 10: Hedonic Regression: Rent

DV: Rent			
Constant	1068.44** (100.14)	Rental Supply	-164.8** (36.97)
Metro Area - Year FE	Yes		
Demographics		**Property Characteristics**	
Age		# of Bedrooms	
35-65	-109.54** (21.73)	1	280.82** (75.89)
Over 65	-27.31 (35.47)	2	453.23** (83.17)
Education		3	578.68** (82.02)
Bachelor's	52.81* (27.82)	4	528.42** (99.85)
High School Grad	403.41** (31.41)	5+	931.32** (132.82)
Marital Status		# of Bathrooms	97.73** (9.51)
Never Married	33.02 (25.49)	# of Rooms	5.72 (14.15)
Married Now	28.05 (25.72)	# of Amenities	26.85** (8.34)
Gender		Property Type	
Male	40.78** (19.45)	House	-321.4** (29.7)
Race		Other	-613.31** (118.6)
Asian	98.94** (42.06)	# of Units in the Structure	
Black	-226.82** (44.71)	2	-286.89** (37.7)
White	-172.4** (51.87)	3-4	-256.14** (34.12)
Origin		5-9	-186.35** (31.62)
Hispanic	-188.39** (24.44)	10+	-262.43** (31.39)
Household Income		Unit Age	-11.8** (1.45)
50K-100K	53.31** (23.07)	Unit Age Squared	0.08** (0.01)
Over 100K	489.24** (27.94)		
N	15,670	R^2	0.2278

Note: * $p < 0.05$, ** $p < 0.01$. Standard errors in parentheses.

Table 11: Hedonic Regression: Rental Occupancy

DV: Rental Occupancy			
Constant	0.402** (0.029)	Rental Supply	-0.016** (0.006)
Metro Area - Year FE	Yes	Rent	-0.00012** (0.00002)
Demographics		**Property Characteristics**	
Age		# of Bedrooms	
35-65	0.049** (0.004)	1	0.035** (0.012)
Over 65	0.081** (0.004)	2	0.061** (0.015)
Education		3	0.094** (0.018)
Bachelor's	0.128** (0.004)	4	0.099** (0.018)
High School Grad	0.204** (0.011)	5+	0.144** (0.028)
Marital Status		# of Bathrooms	0.009** (0.003)
Never Married	0.087** (0.003)	# of Rooms	-0.0003 (0.0017)
Married Now	0.072** (0.003)	# of Amenities	0.009** (0.001)
Gender		Property Type	
Male	0.031** (0.003)	House	-0.041** (0.009)
Race		Other	-0.059** (0.021)
Asian	0.463** (0.008)	# of Units in the Structure	
Black	0.459** (0.008)	2	-0.035** (0.008)
White	0.494** (0.006)	3-4	-0.028** (0.008)
Origin		5-9	-0.016** (0.006)
Hispanic	0.026** (0.006)	10+	-0.026** (0.008)
Household Income		Unit Age	-0.0007** (0.0003)
50K-100K	-0.027** (0.003)	Unit Age Squared	0.000003 (0.000003)
Over 100K	-0.069** (0.013)		
N	15,670	R^2	0.6075

Note: * $p < 0.05$, ** $p < 0.01$. Standard errors in parentheses.

Table 12: Hedonic Regression: Airbnb Price

DV: Logged Airbnb Price			
Constant	3.576** (0.017)	Metro Area - Year FE	Yes
Airbnb Supply	-0.069** (0.002)	Metro Area - Month FE	Yes
Demographics		**Property Characteristics**	
Age	0.016** (0.0001)	# of Bedrooms	
Education		1	0.151** (0.002)
Bachelor's	-0.004 (0.007)	2	0.402** (0.001)
High School Grad	-0.284** (0.008)	3	0.673** (0.002)
Marital Status		4	0.956** (0.002)
Never Married	0.662** (0.013)	5+	1.161** (0.003)
Married Now	-0.664** (0.012)	# of Bathrooms	0.169** (0.0005)
Gender		# of Amenities	0.010** (0.0002)
Male	0.070** (0.008)	Property Type	
Race		House	0.008** (0.0009)
Asian	0.143** (0.007)	Other	0.062** (0.0008)
Black	-0.177** (0.007)	Room Type	
White	0.225** (0.006)	Private/Shared	-0.604** (0.0008)
Origin			
Hispanic	-0.068** (0.003)	**Airbnb-related metro variables**	
Household Income	0.002** (0.00001)	Airbnb history	0.002** (0.0008)
		Air passengers (in millions)	0.006 (0.0004)
N	3,219,447	R^2	0.5700

Note: * $p < 0.05$, ** $p < 0.01$. Standard errors in parentheses.

Table 13: Hedonic Regression: Airbnb Occupancy

DV: Airbnb Occupancy			
Constant	0.735** (0.045)	Metro Area - Year FE	Yes
Airbnb Supply	0.016** (0.001)	Metro Area - Month FE	Yes
Logged Airbnb Price	-0.095** (0.012)		
Demographics		**Property Characteristics**	
Age	0.001** (0.0002)	# of Bedrooms	
Education		1	-0.031** (0.002)
Bachelor's	-0.190** (0.004)	2	-0.014** (0.005)
High School Grad	-0.430** (0.007)	3	0.0004 (0.008)
Marital Status		4	0.017 (0.012)
Never Married	0.247** (0.012)	5+	0.039** (0.014)
Married Now	-0.043** (0.011)	# of Bathrooms	-0.020** (0.002)
Gender		# of Amenities	0.023** (0.0002)
Male	-0.063** (0.005)	Property Type	
Race		House	0.024** (0.024)
Asian	-0.020** (0.005)	Other	0.059** (0.059)
Black	0.016** (0.005)	Room Type	
White	0.011** (0.005)	Private/Shared	-0.121** (0.007)
Origin			
Hispanic	0.010** (0.002)	**Airbnb-related metro variables**	
Household Income	1.18e-06 (0.00003)	Airbnb history	0.006** (0.006)
		Air passengers (in millions)	0.025** (0.025)
N	3,219,447	R^2	0.1172

Note: * $p < 0.05$, ** $p < 0.01$. Standard errors in parentheses.

Chapter II

The Impact of Airbnb on the Residential Real Estate Market: Aggregate and Micro Level Analyses

1 Introduction

The global sharing economy market is rapidly growing and is projected to generate about $335 billion by 2025 (PwC report 2015). In particular, Airbnb, which enables homeowners to rent out their properties as short-term lodging for travelers, has obtained a valuation of over $30 billion[27]. As such, the impact of Airbnb on the real estate market has been a topic of heated debate, as we discuss later. In this research, by leveraging the entire Airbnb data during the 2014-2015 period, which spans nearly 350,000 listings, and the zip code level and property level real estate data, we study the impact of Airbnb on the residential real estate market. First, we investigate the aggregate impact of Airbnb on the transaction volume of residential real estate properties across the US. Second, we perform a micro-level (individual housing units across several focal cities) analysis to examine the impact of Airbnb on the real estate price. Lastly, we explore the underlying mechanism of how home sharing platforms affect the real estate market.

The impact of home sharing platforms on the real estate market has been a topic of heated debate, because there are different forces at play. On the one hand, having a home sharing option may lead to an increase in the demand for residential real estate properties and raise prices. Airbnb serves as an extra income source and helps hosts pay mortgage, which increases demand for houses; more people can and are willing to afford to buy homes. According to a survey done by Airbnb, the number one use of the hosting income is to pay mortgage. Real estate investors are also beginning to see Airbnb hosting as a lucrative investment opportunity. Individual hosts might consider buying a second house as an investment to rent out on Airbnb. There are even startups, such as Mashvisor and Revestor, that give real estate investment recommendations for the short-term rental business.

On the other hand, the presence of home sharing platforms may lead to a decrease in real estate demand and price because of the negative externality issues. Popular examples are noise, nuisance, safety, and traffic

[27] http://www.wsj.com/articles/airbnb-valued-at-31-billion-after-new-funding-round-1489086240

issues stemming from an increase in transient travelers. For example, in 2016, an Airbnb host was taken to court by neighbors after renting out his unit on Airbnb, where guests threw a wild party for 150 people. These Airbnb issues potentially affect home buying and selling, as sellers are required to disclose any disputes or complaints about their properties or properties nearby. Being involved in such disputes or complaints can make the property less desirable for buyers, making selling harder for sellers.

The impact of home sharing platforms on the residential real estate market has emerged as a major interest, as it directly relates to housing availability and affordability. Affordable housing has been a big issue, especially for cities with tight housing markets. Housing advocates have raised concerns that home sharing platforms adversely impact the real estate market and claim that they should be regulated. For example, in San Francisco, a ballot initiative was raised to increase regulations on short-term rentals to protect affordable housing.

In contrast to the significance of the problem and the size of press coverage, the impact of home sharing platforms on the real estate market is surprisingly understudied in the academic literature. The existing studies are descriptive (Gurran and Phibbs (2017); Lee (2016)), limited to a particular city (Gurran and Phibbs (2017); Horn and Merante (2017); Lee (2016); Sheppard and Udell (2016); van der Bijl (2016)), or their main focus is on the long-term rental (Barron, Kung, and Proserpio (2017); Gurran and Phibbs (2017); Horn and Merante (2017); Kim, Li, and Srinivasan (2017); Lee (2016)). We extend and build on the existing studies by conducting econometric analyses about the impact on real estate transaction volume and price. We also contribute to the literature by combining various datasets from different sources, which cover comprehensive area across the US. Moreover, we explore the underlying mechanisms of this impact.

This paper investigates the impact of Airbnb on the residential real estate market. In particular, we seek to answer the following three questions: (1) Does Airbnb affect real estate transaction volume and price? If so, in which direction? (2) What are the differential impacts of Airbnb on markets with different housing market conditions, such as real estate supply and owner occupancy? (3) What is the driving force of the impact? To answer these questions, we conduct two analyses. First, we conduct aggregate level analysis to investigate the impact of Airbnb on real estate transaction volume, using Airbnb and zip code level real estate data. Then, we study the impact on real estate price and explore the underlying mechanisms, in the micro-level analysis using Airbnb and property level real estate data.

In the aggregate level analysis, we estimate the impact of Airbnb on real estate transaction volume using a difference in differences strategy. We regress the number of real estate transactions on Airbnb supply in a specific zip code. The Airbnb treatment impact is identified by comparing differences in the number of transactions for zip codes affected by Airbnb before and after the entry of Airbnb, against a baseline of difference in the number of transactions for zip codes unaffected by Airbnb over the same period of time.

Since the real estate market is a complex market, we incorporate various controls such as real estate supply and price, demographics, mortgage and rent affordability indices, and tourism. We further control for zip code and time- fixed effects. We first perform the analysis for all residential transactions, and repeat it for four different property types of houses: condo/co-ops, multi-family residences, single-family residences, and townhouses. Among different types of houses, we find that Airbnb has a positive and significant impact on two out of the four types: condo and multi family residences. A 10% increase in Airbnb listings is associated with a statistically significant 0.175% and 0.135% increase in the number of condo/co-op and multi family residence transactions, respectively. The reason why Airbnb affects certain types of houses may be that these two types are more suitable for Airbnb hosting. For instance, condos are easier to maintain; multi-family houses may have multiple entrances or are easier to separate the space. To further illustrate this point and examine the differential impact on different properties and markets, we conduct the analysis at the property level.

In the micro-level analysis, we study whether properties with a higher suitability of Airbnb have higher prices in the real estate market. To do so, we construct a measure of how suitable a property is to be listed on Airbnb, which we call "Airbnb index". A property may be suitable for Airbnb due to specific property characteristics such as separate entrance or location characteristics. We look at the properties on Airbnb and their performance and train a machine learning algorithm to map their characteristics to their performance on Airbnb. Using this algorithm, we can map any property with a specific set of characteristics to a value of Airbnb suitability, in other words, the Airbnb index. Interestingly, we find that the property and location characteristics that make a property suitable for Airbnb differ by the type of listings. Some characteristics make a property suitable for listing as entire places, while other characteristics make a property suitable for listing as private rooms. For example, properties with high Airbnb index for entire place listing are usually a two-bedroom or smaller apartment with good amenities and location. For private room listing, both apartment and houses are suitable, and amenities and location matter less. We thus construct two Airbnb indices for each property in the real estate dataset.

To see whether Airbnb drives up the prices of properties that are more suitable for Airbnb, we regress log sale price on Airbnb index. We remain agnostic about whether the property would be listed as entire places or private rooms and run separate regressions by listing type. Recall that there were two forces that impact real estate price – suitability for listing as entire places or private rooms on Airbnb drive up prices, while negative externality decreases price. The estimated coefficients for Airbnb indices in the regression will show a net effect of the two forces. Besides the main effect, to analyze how their impacts differ by housing market conditions, we extend the regression model to include the interaction effects with real estate supply and owner occupancy. The results show that suitability for entire place listings raises real estate price, while the direction of the main effect from suitability for private room listings depends on housing market conditions. We find that the price-increasing impact of Airbnb is larger in tight real estate markets with low real estate supply. The impact from suitability for private room listings is larger in markets with high owner occupancy, as it is easier to list on Airbnb as a homeowner than as a tenant.

The findings of this study have a number of policy implications, as our results can be used as a basis for policy and regulations on home sharing platforms to alleviate housing availability and affordability issues. First, the "Airbnb index" algorithm that maps property and location characteristics to Airbnb suitability can be used to find out which properties are highly affected by Airbnb. Second, we find that suitability for entire place listings raises real estate price, and suitability for private room listings can have significantly positive or negative impact on real estate price depending on the housing market conditions. The existing studies often assume that private room listings do not have an adverse impact on housing markets and Airbnb "helps" the hosts by offering an extra income for paying mortgage. However, our findings indicate that private room listings can lead to an increase in real estate price. Third, the impacts differ by housing market conditions. The impact of Airbnb is larger in tight real estate markets with low real estate supply. The impact from private room listings is larger in markets with high owner occupancy. These findings can be used to determine where Airbnb should be and which Airbnb properties should be regulated.

The rest of the paper is organized as follows. Section 2 provides a brief overview of relevant literature and Section 3 introduces the data. Section 4 and 5 explain the empirical strategy and discuss the key findings for the aggregate-level analysis on the real estate transaction volume and for the micro-level analysis on the real estate price, respectively. We conclude and provide future research plans in Section 6.

2 Literature Review

This paper is most closely related to the literature on the impact of home sharing platforms on the housing market. Despite the importance of the issue, there remains a paucity of evidence on how home sharing platforms affect housing market. Sheppard and Udell (2016) and van der Bijl (2016) studied the impact of Airbnb on residential real estate price using hedonic regression of house price on Airbnb activity in New York and Amsterdam, respectively. However, their analyses do not take account of the endogeneity problem that house price and Airbnb activity influence each other simultaneously. Barron et al. (2017) explain that Airbnb increases rents, which in turn raises house price because rent is capitalized into house prices. They find that a 10% increase in Airbnb listings lead to a 0.42% in rents and a 0.76% increase in house prices. Other studies have mainly focused on the impact of Airbnb on the rental housing market. Lee (2016) and Gurran and Phibbs (2017) provide policy and urban planning recommendations, respectively, based on the descriptive analyses of the impact of Airbnb. Horn and Merante finds that a one standard deviation increase in Airbnb listings is associated with a 0.4% increase in rents in Boston. Kim et al. (2017) studies how Airbnb affects rental housing supply in a structural modeling approach.

The previous studies about the impact of Airbnb on the housing market are either descriptive (Gurran and Phibbs 2017; Lee 2016), limited to a particular city (Gurran and Phibbs 2017; Horn and Merante 2017; Lee 2016; Sheppard and Udell 2016; van der Bijl 2016), or their main focus is on the long-term rental (Barron et al. 2017; Gurran and Phibbs 2017; Horn and Merante 2017; Kim et al. 2017; Lee 2016). We extend and build on the existing studies by conducting econometric analyses about the impact on real estate transaction volume and price. We also contribute to the literature by combining various datasets of Airbnb and the real estate market from different sources, which cover comprehensive area across the US. Moreover, we explore the underlying mechanisms of this impact.

This paper contributes to the more general literature on the impact of sharing economy. The recent literature focuses on the impact of sharing economy and online peer-to-peer platforms such as Airbnb (accommodation), Uber (transportation), Taskrabbit (domestic tasks), and Craigslist (local goods and services). Though peer-to-peer sharing of a product is not a new concept, recent technological advances have facilitated the process by reducing search costs and frictions. For example, Airbnb provides a platform where both property owners and tourists can easily find each other, reducing search costs and frictions for both sides. Kroft and Pope (2014) show that Craigslist significantly reduces market frictions by providing a platform for online searches instead of traditional searches (e.g., in print) and improves matching efficiency. Although the online peer-to-peer platforms greatly reduce search costs and frictions, several studies suggest that these platforms are also inherently frictional, and that appropriate market design such as search and matching algorithms can lead to even more increase in matching efficiencies. (Einav et al. 2016; Fradkin 2017; and Cullen and Farronato 2015)

The growth of sharing economy and online peer-to-peer platforms has changed consumers' usage and buying behavior, which in turn have a impact on various firm behaviors including pricing and product strategies. In the sharing economy, the value of product ownership comes from not only the self-usage of the product but also from the income from renting out the product when it is not in use. Benjaafar et al. (2015) build an analytical framework to study how peer-to-peer sharing affects product ownership and usage. Fraiberger and Sundararajan (2015) also develop a dynamic model taking secondary market into account. Razeghian and Weber (2016) study product owners' decision to share the product or not. Gong et al. (2017) empirically analyze the impact of Uber entry on new vehicle ownership in China and find that the entry of Uber is associated with a considerable increase in vehicle purchase. The sharing economy also expands consumer choice set by providing alternative options. Zervas et al. (2017) find that the entry of

Airbnb decreases hotel revenue, because Airbnb stays substitute hotels, especially for budget travelers. Li and Srinivasan (2017) also study the impact of Airbnb on the hotel industry, and find that the impact largely depends on consumer composition (i.e., business and leisure travelers) and market-specific seasonality. Kroft and Pope (2014) focus on Craigslist and estimate that the introduction of this website brought about a 7% reduction in classified job ads in newspapers. Seamans and Zhu (2014) also show that the introduction of Craigslist led to a drop in classified-ad rates. Firm strategies change responding to changes in buyer behavior under sharing economy. Jiang and Tian (2016) show that the firm chooses different level of quality in the presence of sharing market. Razeghian and Weber (2017) study how sharing markets change the pricing and product design strategy, particularly in terms of durability of the product. Abhishek et al. (2016) analyze the impact of peer-to-peer rental markets on the behavior of original equipment manufacturer (OEM).

3 Data

For our study, we collect and combine data from various sources. The three main datasets used in this study are Airbnb listings data, aggregate-level housing market data provided by Redfin, a national real estate firm, and micro-level real estate listings data collected from a leading MLS platform. We supplement the three main datasets with data on demographics from American Community Survey (ACS), real estate metrics published by real estate companies such as Zillow and Redfin, air traffic data from the T-100 Market (All Carriers) database, and the measures of walkability, public transit, and bikeability from Walk Score.

The first main dataset is listing-level Airbnb data from August 2014 to October 2015 in the United States. It contains the information on every listings listed on Airbnb in the sample period. During the sample period, 342,873 properties were listed on Airbnb. Each listing includes over 20 property characteristics such as the number of bedrooms and bathrooms, facilities (e.g., pool, gym, etc.), location (zip code, latitude/longitude), property type (e.g., house, apartment, etc.), home type (entire place, private room, or shared room). It also provides the monthly performance details such as availability (i.e., open for booking), booking, and average daily price. We also web scrape some information that is not available from this dataset, including the date that hosts became Airbnb members.

The second main dataset is the aggregate level housing market data published by Redfin, one of the leading online real estate platforms. Redfin publishes monthly time series housing market data for metropolitan areas, cities, neighborhoods, and zip codes across the nation. The available data include prices (e.g., median sale price, median price per square foot), sales (e.g., the number of homes sold, median days on the market), and inventory (e.g., the number of homes on the market, the number of new listings).

For the third main dataset, we web scraped a random sample of real estate listings in major US cities from a leading MLS platform. We have data for 39,275 transactions that mostly occurred between 2011 and 2017, with property details (e.g., the number of bedrooms and bathrooms, year built, floor size, school district, tax history, etc.) and transaction history (listing price, price change, selling price, time on the market).

We also collect data from various data sources for auxiliary purposes of assembling a set of control variables and supplementing the parts that are unobserved from the main dataset. We collect yearly zip code level demographics such as population, population density, age, income, and unemployment rate from American Community Survey dataset for the 2012-2015 period. We use various real estate industry metrics that help undersand the market and the industry, the metrics which are constructed and published by Zillow and Redfin. For instance, we collect quarterly metro-level housing affordability indices from Zillow. We also obtain the monthly number of domestic and international airline passengers by destination airport from the T-100 Market (All Carriers) database and use it as a proxy for the tourism demand. Lastly, we collect Walk Score data which consists of Walk Score (the walkability of a given location), Transit Score (how well a location is served by public transit), and Bike Score (whether a location is good for biking) for each address. The scores are a number between 0 and 100.

4 Aggregate-level Analysis on Real Estate Transaction Volume

We estimate the aggregate impact of Airbnb on the residential real estate transaction volume using difference in differences strategy. The difference in differences methodology enables us to identify the Airbnb treatment by comparing differences in the number of transactions for zip codes affected by Airbnb before and after the entry of Airbnb, against a baseline of difference in the number of transactions for zip codes unaffected by Airbnb over the same period of time. The entry and growth of Airbnb show significant variation across geographical regions. Table 1 and Figure 1 show the number of Airbnb properties that were available for booking on Airbnb for at least a day during the period from August 2014 to October 2015. Specifically, Table 1 shows the number for the top 20 cities with the largest number of properties, and Figure 1 shows the number by zip code for the top 6 cities. They show that the presence of Airbnb greatly varies by city and zip codes. In our Airbnb listings data, there are many zip code areas without the presence of Airbnb, whereas the zip code with the largest number has 3,658 listings (zip code 33139, located in Miami Beach area).

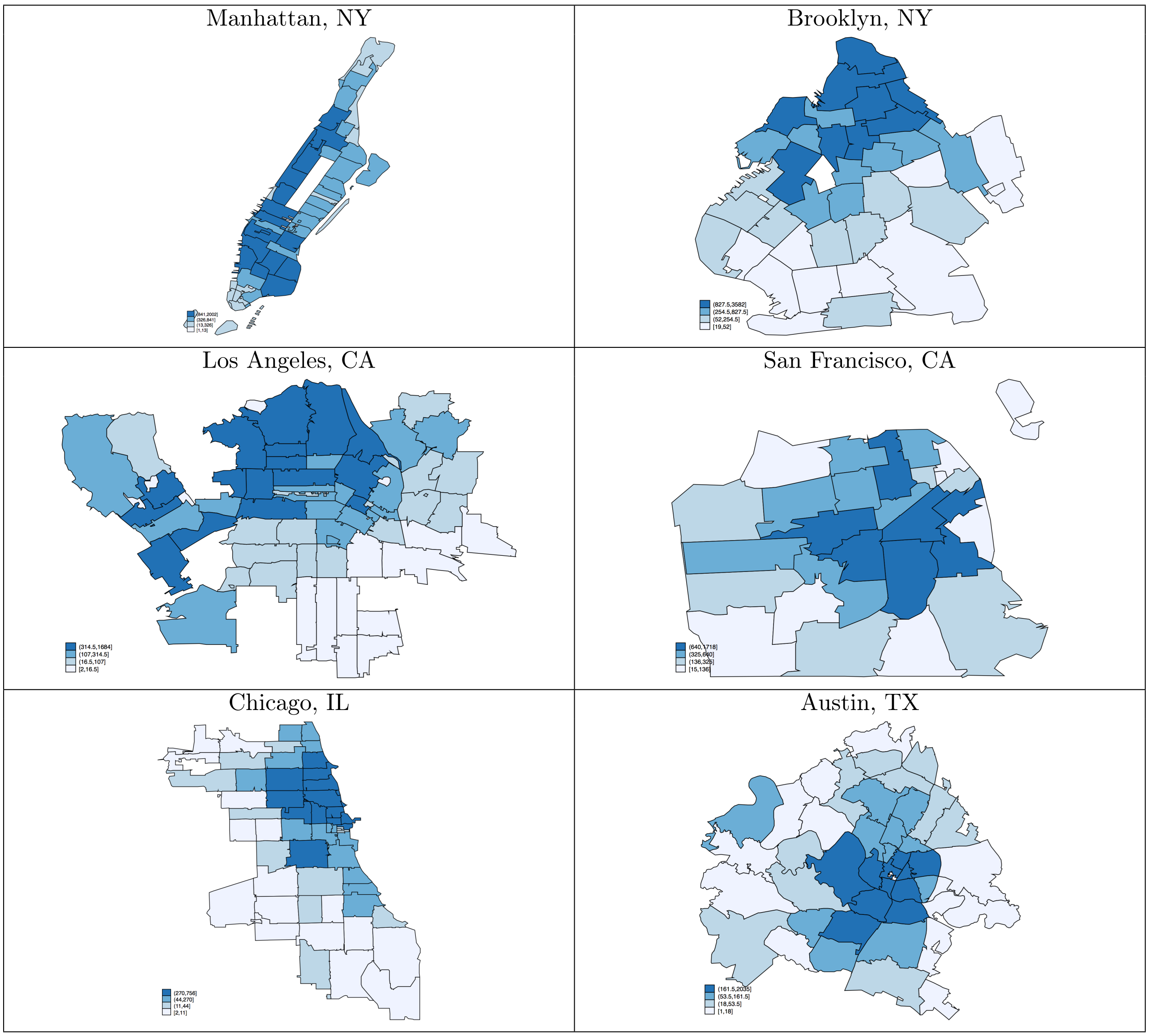
We use the following specification for our difference in differences estimation:

$$\text{log Transactions}_{it} = \beta \text{log Airbnb Supply}_{it} + Z_{it}'\gamma + \iota_i + \tau_t + \epsilon_{it} \tag{1}$$

Table 1: The number of Airbnb properties that were available for booking for at least one day during August 2014 to October 2015

	State	City	Airbnb Properties
1	NY	Manhattan	31,149
2	NY	Brooklyn	19,824
3	CA	Los Angeles	14,986
4	CA	San Francisco	12,299
5	IL	Chicago	8,820
6	TX	Austin	7,983
7	DC	Washington	5,855
8	FL	Miami Beach	5,690
9	CA	San Diego	5,621
10	WA	Seattle	5,105
11	OR	Portland	4,033
12	FL	Miami	4,012
13	LA	New Orleans	3,905
14	PA	Philadelphia	3,554
15	CO	Denver	2,968
16	TN	Nashville	2,619
17	MA	Boston	2,557
18	NY	Queens	2,358
19	NV	Las Vegas	2,252
20	CA	Venice	2,201

Figure 1: The number of Airbnb properties that were available for booking for at least one day during August 2014 to October 2015



The *dependent variable* is the log of the number of residential real estate transactions that occurred in zip code i at time (year-month) t. We use the zip code level time-series real estate transaction data from January 2012 to October 2015, provided by Redfin. For the *independent variable*, we measure the Airbnb supply by the log of the cumulative number of all properties listed on Airbnb platform prior to a given month in a given zip code. Since we do not observe the date each listing became available, we approximate the entry date of a listing by the date that its host became an Airbnb member, following Zervas et al. (2017). This measure of Airbnb supply assumes that there is no exit. The coefficient of interest is β, which indicates and quantifies the impact of Airbnb on the number of transactions.

Since real estate market is a complex market, we incorporate various *controls* (Z_{it}) such as the real estate supply, price, and demographics. The way that Airbnb can affect the real estate market is that it provides additional income sources for hosts to pay for new houses. We thus further control for the factors that would impact the incentives of the hosts. For instance, we control for mortgage and rent affordability indices because they directly affect the incentives of the potential property buyers in the real estate market. Mortgage affordability index is the proportion of the monthly mortgage payment for the median-valued home to the monthly median income, and rent affordability index is the proportion of the monthly median rent to the monthly median income. We control for tourism as it affects how profitable listing on Airbnb is as an income source. Lastly, we include zip code fixed effects (ι_i) and time fixed effects (τ_t).

We estimate equation (1) for all residential transactions and for different types of properties: condo/co-ops, multi family residences (2-4 unit), single family residences, and townhouses. The results are reported in Table 2. Among different types of properties, Airbnb has positive and significant impacts on two out of the four types: condo/co-ops ($\beta = 0.0175$) and multi family residences ($\beta = 0.0135$). A 10% increase in Airbnb listings is associated with a statistically significant 0.175% and 0.135% increase in the number of condo/co-op and multi family residence transactions, respectively. The impact on single family residences is positive ($\beta = 0.0051$) and the impact on townhouses ($\beta = -0.0042$) is negative, but they are both insignificant. The estimated coefficients for the controls have the expected signs and magnitudes (e.g., lower mortgage affordability index, the proportion of mortgage payment to income, is associated with increase in real estate transaction volume).

The reason for the differential impact on different property types may come from suitability for Airbnb hosting. The two property types with positive and significant impacts may be more suitable for Airbnb hosting. For instance, condo/co-ops are easier to manage; multi family residences may have multiple entrances or are easier to separate the space. To further illustrate this point and examine the differential impact on different properties and markets, we conduct the analysis at the property level in Section 5.

Table 2: Difference-in-differences estimates of the impact of Airbnb on the number of residential real estate transactions

log Transactions	(1) All Residential	(2) Condo/Co-op	(3) Multi Family	(4) Single Family	(5) Townhouse
log Airbnb Supply	0.0051	0.0175***	0.0135**	0.0033	-0.0042
	(1.60)	(3.20)	(2.55)	(1.01)	(-0.75)
log Real Estate Supply	0.0426***	0.0031	-0.0306***	0.0362***	0.0509***
	(5.01)	(0.77)	(-8.70)	(4.67)	(8.84)
Real Estate Price	0.0253	-0.0165	-0.0536**	-0.0239**	-0.0280*
	(0.93)	(-0.59)	(-2.52)	(-2.00)	(-1.80)
Population	0.0080***	0.0147***	-0.0039	0.0074***	0.0065
	(2.94)	(3.70)	(-1.21)	(2.70)	(1.55)
Median Age	-0.0008	-0.0064**	-0.0053**	0.0007	0.0027
	(-0.50)	(-2.31)	(-2.06)	(0.44)	(0.90)
Median Income	-0.0005	0.0020***	0.0004	-0.0006	-0.0001
	(-1.14)	(2.63)	(0.45)	(-1.45)	(-0.10)
Unemployment Rate	-0.0028**	-0.0010	-0.0005	-0.0034***	-0.0081***
	(-2.33)	(-0.45)	(-0.24)	(-2.86)	(-3.29)
Mortgage Index	-4.3965***	-3.8170***	-1.5096***	-4.3409***	-1.0489***
	(-38.51)	(-22.70)	(-9.33)	(-39.59)	(-4.33)
Rent Index	1.3795***	1.5144***	0.3129	1.2679***	0.5163*
	(10.51)	(6.66)	(1.25)	(9.67)	(1.72)
Air Traffic	-0.0052	-0.0378***	-0.0179	0.0028	0.0275**
	(-0.76)	(-3.39)	(-1.56)	(0.38)	(2.15)
N	313,312	194,815	119,536	309,826	167,009
Within R^2	0.3595	0.1341	0.0242	0.3354	0.1262

Note: Cluster-robust t-statistics (at the zip code level) are shown in parentheses. All specifications include zip code fixed effects and year-month fixed effects.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01

5 Micro-level Analysis on Real Estate Price

In the previous section, we have discussed the differential impact of Airbnb on the different property types and pointed out that the impact of Airbnb potentially depends on how suitable the properties are for listing on Airbnb. In Section 5, we study whether properties with a higher suitability of Airbnb have higher prices in the real estate market. To do so, we construct a measure of how suitable a property is to be listed on Airbnb, in Section 5.1. Section 5.2 describes the empirical strategy to study the impact of suitability for Airbnb on real estate price. We discuss the results and findings in Section 5.3.

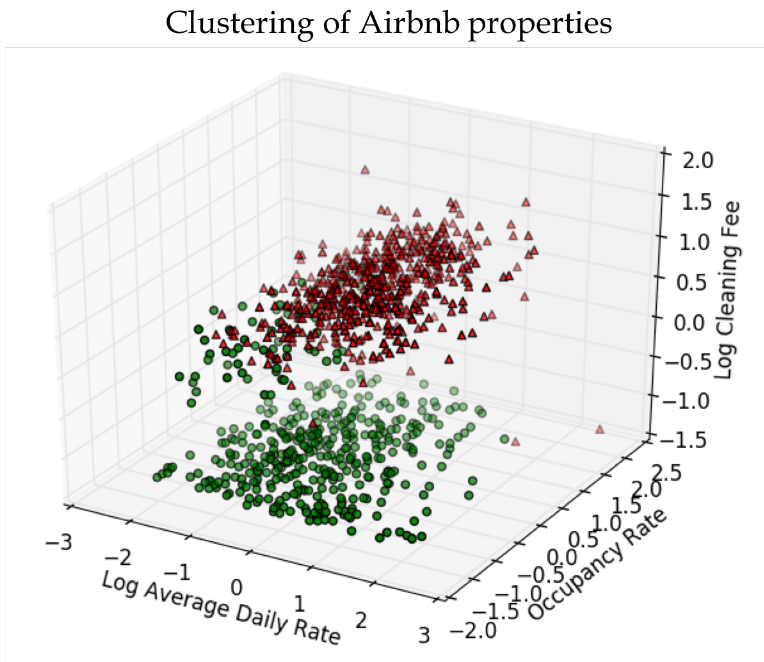
5.1 Constructing Airbnb Index

To study whether properties with a higher suitability of Airbnb have higher prices in the real estate market, we need a measure of how suitable a property is to be listed on Airbnb, which we call "Airbnb index". A property may be suitable for Airbnb due to specific property characteristics such as separate entrance or location characteristics. We look at the properties on Airbnb and their performance and train a machine learning algorithm to map their characteristics to their performance on Airbnb. Using this algorithm, we can map any property with a specific set of characteristics to a value of Airbnb suitability, in other words, the Airbnb index. To develop the algorithm, we follow the two steps: (1) obtain suitability labels for Airbnb properties, and (2) train a machine learning algorithm to predict the labels using property and location characteristics as features.

As the first step, we obtain suitability labels for Airbnb properties. We choose the labels based on how homeowners make their decisions about whether to list on Airbnb. To decide whether to list on Airbnb or not, a homeowner will consider potential profit from Airbnb, which comes from both revenue and cost. We choose three dimensions that potentially affect revenue and cost. For revenue, we select average daily price and occupancy rate. We use cleaning fee as a proxy for the cost for hosting Airbnb. Cleaning fee is a one-time fee that hosts impose to guests regardless of the length of their stay, and would reflect the opportunity cost for the actual time that hosts spend to maintain the house. Then, we cluster the Airbnb properties into two groups, using the chosen three dimensions of average daily price, occupancy rate, and cleaning fee. Figure 2 shows the result of the cluster analysis.

Note that the properties are cleanly separated into two groups. We also find that the two groups are different in terms of the performance on Airbnb. Using the cleaning fee as a proxy for the cost of housing Airbnb, the average ratio of revenue to cost for the green group (= 67.52) is about 1.5 times higher than the ratio of the red group (= 45.23). Thus, we label the green group as the high Airbnb index group (Airbnb index = 1) and the red group as the low Airbnb index group (Airbnb index = 0).

Figure 2: Clustering of Airbnb properties



In the second step, we train the algorithm and obtain a mapping between characteristics and Airbnb suitability. We use property and location characteristics (e.g., the number of bedrooms and bathrooms, property type, Walk Score, etc.) as features, because the two main factors that determine whether a property is a good Airbnb are property and location. We use extreme gradient boosting for training the classifier, which is a popular off-the-shelf algorithm that has been favored by many Kaggle winners. Parameters are tuned by 5-fold cross validation, and cross validation and test accuracies are both 80% or higher.

This algorithm predicts a property's suitability for Airbnb listing given its characteristics. Looking at the post-hoc interpretation of the algorithm, we find that the property and location characteristics that make a property suitable for Airbnb differ by the type of listings. Some characteristics make a property suitable for listing as an entire place, while other characteristics make a property suitable for listing as a private room. For instance, properties with high Airbnb index for entire place listing are usually a two-bedroom or smaller apartment with good amenities and location. For private room listing, both apartment and houses are suitable, and amenities and location matter less. We thus construct two Airbnb indices for each property in the real estate dataset.

5.2 Empirical Strategy

Our empirical goal is to investigate the impact of suitability for Airbnb on real estate price. We regress log sale price on Airbnb index to see whether Airbnb drives up the prices of properties that are more suitable for Airbnb. We remain agnostic about whether the property would be listed as entire places or private rooms and run separate regressions by listing type. We employ the following specification:

$$\log \text{Sale Price}_i = \beta_0 + \beta_1 X_i + \gamma(\text{Controls}_i) + \epsilon_i, \quad X_i \in \{X_i^E, X_i^P\}$$

where X_i denotes Airbnb index for property i: X_i^E for an entire place listing and X_i^P for a private room listing. There are two forces that impact real estate price – suitability for listing as entire places or private rooms on Airbnb drive up prices, while negative externality decreases price. The estimated coefficient β_1 for Airbnb index in the regression will show a net effect of the two forces.

Because real estate markets are very complex, we incorporate various controls that potentially affect real estate prices. Property and location are the two primary factors that determine real estate price. We control for various property characteristics such as the number of bedrooms and bathrooms, property types, floor size, age, appliances included, and building amenities. For location characteristics, we include zip code fixed effects, Walk Score, and nearby school ratings (high, low, or not available). The real estate literature has studied that real estate price also depends on transaction characteristics such as time on the market and the number of pictures in the listing. For instance, a prospective home buyer can make an inference about the quality of a house from the amount of time it spends on the market, which indicates that the longer a house has been on the market, the lower the selling price will be (Taylor 1999). With the widespread of the online real estate platforms such as Zillow and Redfin, the number of pictures in the listing also influences selling price (Carrillo 2008; Benefield et al. 2011; Benefield et al. 2012; Allen et al. 2015). We thus control for the transaction characteristics including time on the market, the number of pictures in the listing, whether it has an agent or not, and the transaction year- and month- fixed effects.

Besides the main effect, we also examine the differential impact on different markets. More specifically, we investigate how the impacts differ by markets with different housing market conditions such as real estate supply and owner occupancy. To do so, we extend the regression model by including the interaction effects with real estate supply (ForSale) and owner occupancy (OwnerOcc):

$$\log \text{Sale Price}_i = \beta_0 + X_i(\beta_1 + \beta_2 \text{ForSale}_i + \beta_3 \text{OwnerOcc}_i) + \gamma(\text{Controls}_i) + \epsilon_i, \quad X_i \in \{X_i^E, X_i^P\} \tag{2}$$

where the variable ForSale is the percentage of for-sale real estate units and OwnerOcc is the percentage of owner-occupied units in the zip code in which property i is located.

We study the following two interaction effects. First, we investigate how the impact differs by real estate supply. The percentage of for-sale units (ForSale) indicates real estate supply, and it is often used as a benchmark to understand whether real estate market is tight or not. We hypothesize that the impact of Airbnb will be larger in tight or seller's real estate markets or reduced in buyer's market (i.e., $\beta_2 < 0$ for both X^E and X^P). Intuition is that if the owner can extract profit from Airbnb, the owners' reservation price will be higher.

Second, we analyze the differential impact on markets with different owner occupancy. Owner occupancy will be potential supply for private room listings, because it would be easier to rent out extra space at home as a homeowner rather than as a renter or a tenant. Thus, our hypothesis is that the impact from suitability for private room listings is larger in markets with high owner occupancy (i.e., $\beta_3 > 0$ for X^P). On the other hand, higher owner occupancy means lower number of renter occupied units, which indicates there are less absentee landlords who potentially switch from long-term rental to Airbnb. We thus hypothesize that the impact from suitability for entire place listings is smaller in markets with high owner occupancy (i.e., $\beta_3 < 0$ for X^E).

5.3 Results

Table 3 reports the regression results for entire place Airbnb index (Column 1) and for private room Airbnb index (Column 2).

Table 3: Regression results for real estate price

Variable	Entire Place	Private Room
Airbnb Index	0.1764***	-0.0532***
Airbnb Index X OwnerOcc	-0.1641*	0.1052***
Airbnb Index X ForSale	-3.1557**	-1.0059**
Bedroom1	0.0172	-0.0032
Bedroom2	0.2662***	0.2386***
Bedroom3	0.3248***	0.2945***
Bedroom4	0.2634***	0.2311***
Bedroom5	0.1469***	0.1131***
Bedroom6	0.0882***	0.0533**
Bedroom7	0.0501	0.017
bathroom	0.0301***	0.0291***

Table 3: Regression results for real estate price

Variable	Entire Place	Private Room
Type: Single Family	0.3079***	0.3057***
Type: Multi Family	0.2676***	0.2703***
Type: Townhose	0.1722***	0.1701***
Type: Condo	0.1056**	0.1033**
Type: Co-op	0.0285	0.0283
Type: Apartment	0.2640***	0.2654***
Type: Multiple Residence	0.2068**	0.2063**
Age	-0.0025***	-0.0025***
Age Squared	0.0000***	0.0000***
Floor Size	0.0003***	0.0003***
Flooring: Hardwood	0.0450***	0.0452***
Flooring: Carpet	-0.0281***	-0.0282***
Flooring: Tile	-0.0125**	-0.0121**
Cooling: Central	0.0282***	0.0266***
Heating: Gas	0.0109*	0.0094
Heating: Air	0.0058	0.0038
Parking: Garage	0.0417***	0.0415***
Parking: Street	-0.0376***	-0.0383***
Parking: Driveway	0.0167	0.0178
Appliance: Dishwasher	0.0191***	0.0191***
Appliance: Dryer	0.0526***	0.0523***
Appliance: Freezer	0.0136*	0.0138*
Appliance: Garbage Disposal	-0.0041	-0.0043
Appliance: Microwave	0.0057	0.0055
Appliance: Range/Oven	-0.0330***	-0.0322***
Appliance: Refrigerator	-0.0316***	-0.0319***
Appliance: Trash Compactor	-0.0022	-0.0039
Appliance: Washer	0.0293***	0.0290***
Amenities: Pool	0.0380***	0.0375***
Amenities: Gym	0.0095	0.0087

Table 3: Regression results for real estate price

Variable	Entire Place	Private Room
Amenities: Hot Tub	-0.0187*	-0.0192*
Amenities: Fireplace	0.0513***	0.0483***
Amenities: Doorman	0.0352**	0.0364**
Amenities: Elevator	0.0100	0.0061
Walk Score	-0.0005***	-0.0006***
School: High	-0.0237**	-0.0247**
School: Low	-0.0304***	-0.0304***
Time on the Market	-0.0001***	-0.0001***
Number of Pictures	0.0017***	0.0017***
Has Agent	-0.0401***	-0.0398***
N	26,304	26,304
R^2	0.8395	0.8394

Note: All specifications include fixed effects for zip code, year sold, and month sold.

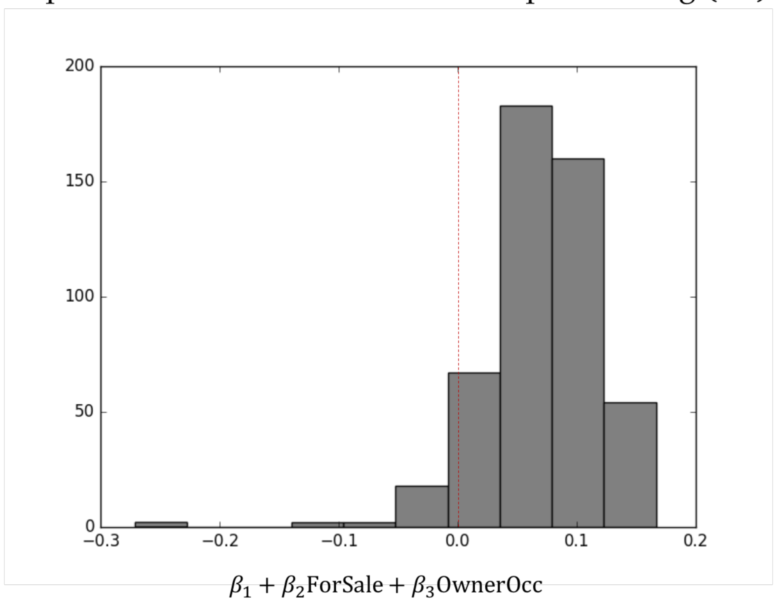
Significance levels: * p<0.1, ** p<0.05, *** p<0.01

Main effects. To assess the direction and the magnitude of the impact of Airbnb suitability, we present the histograms for the main effects of Airbnb indices ($\beta_1 + \beta_2 \text{ForSale}_i + \beta_3 \text{OwnerOcc}_i$) for the total of 498 zip codes in the sample, in Figure 3. The results show that suitability for entire listings raises real estate price (Figure 3, top panel). The main effects are positive for the 94% of the zip codes, and none of the negative ones are significantly negative. On the other hand, the direction of the main effect for suitability for private room listings depends on housing market conditions (Figure 3, bottom panel). For the 71% of the zip codes, we find a negative effect, which means that if a property is suitable for listing as private rooms, the real estate price is lower. In other words, the negative externality impact seems to dominate the suitability for private room listings. However, there are a few zip codes where the impact is significantly positive. For example, the impact of Airbnb suitability for private room listings is higher in zipped 78739, a zip code located in Austin, Texas with low real estate supply and high owner occupancy. It is important that housing market conditions are taken account for, in order to identify the direction of the impact of Airbnb suitability for private room listings.

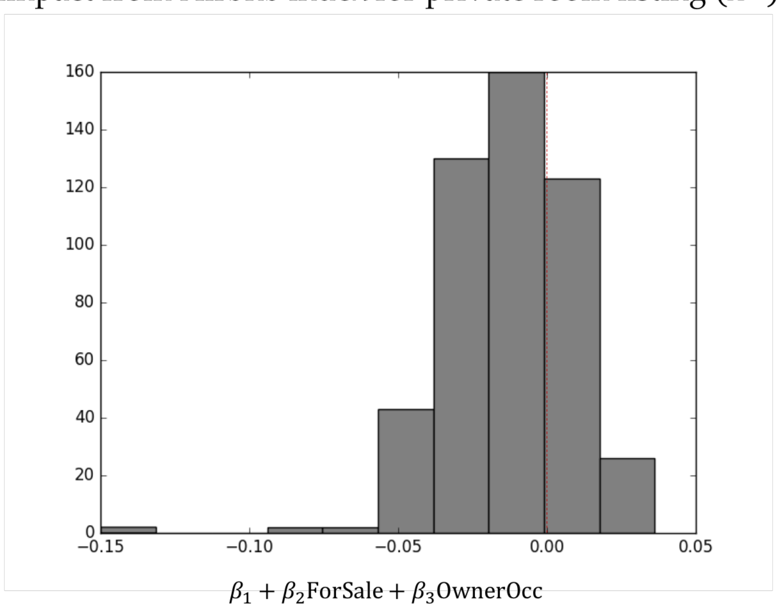
Interaction effects. In terms of the interaction effects, we find that the impact of Airbnb is larger in tight

Figure 3: Histograms for the main effects of Airbnb indices

Impact from Airbnb index for entire place listing (X^E)



Impact from Airbnb index for private room listing (X^P)



real estate markets with low real estate supply, as expected. In other words, β_2 is negative and statistically significant for both entire place ($\beta_2 = -3.156$ for X^E) and private room listings ($\beta_2 = -1.006$ for X^P). The coefficient of owner occupancy, β_3, for entire place listings shows the impact of absentee landlords who potentially switch from long-term rental to Airbnb. It has the expected negative sign, but the impact is only marginally significant ($\beta_3 = -0.164$ for X^E; p-value < 0.1). For private room listings, the coefficient of owner occupancy is positive and statistically significant ($\beta_3 = 0.105$ for X^P). The results confirm our hypothesis that the impact from suitability for private room listings is larger in markets with high owner occupancy, because it is easier to list on Airbnb as a homeowner than as a tenant.

6 Conclusion

This paper studies the impact of Airbnb on the residential real estate market, focusing on the two aspects of the real estate transaction volume and price. The results indicate that an increase in Airbnb listings is associated with the increase in the number of housing transactions for condo/co-ops and multi-family residences. The price-increasing impact of Airbnb from suitability for an entire place listing is larger than that from suitability for a private room listing. Moreover, the impact differs by housing market conditions such as real estate supply and owner occupancy. This paper has policy implications for regulating home sharing platforms to alleviate housing availability and affordability issues. Our findings can be used to determine where Airbnb should be and which Airbnb properties should be regulated.

This paper is a work in progress and future work will improve the paper by providing various robustness checks, refining the Airbnb index algorithm, and further exploring the underlying mechanisms of the impact. For instance, we plan to develop a continuous measure of Airbnb suitability, in addition to the current binary Airbnb index. We are also working on scraping more properties to include in the micro-level analysis. A potential avenue to proceed is to match the Airbnb listings data and the property level real estate data, as the two data both include location information. This will enable us to further explore the impact of Airbnb on the real estate market.

References

[1] Abhishek, V., Guajardo, J. A., & Zhang, Z. (2016). Business Models in the Sharing Economy: Manufacturing Durable Goods in the Presence of Peer-To-Peer Rental Markets. Working Paper.

[2] Allen, M. T., Cadena, A., & Rutherford, J. (2015). Effects of Real Estate Brokers' Marketing Strategies: Public Open Houses, Broker Open Houses, MLS Virtual Tours, and MLS Photographs. *Journal of Real Estate Research*, 37(3), 343–369.

[3] Barron, K., Kung, E., & Proserpio, D. (2017). The Sharing Economy and Housing Affordability: Evidence from Airbnb. Working Paper.

[4] Benefield, J. D., Cain, C. L., & Johnson, K. H. (2011). On the Relationship Between Property Price, Time-on-Market, and Photo Depictions in a Multiple Listing Service. *The Journal of Real Estate Finance and Economics*, 43(3), 401–422.

[5] Benefield, J., Cain, C., & Gleason, A. (2012). Visual display of residential property information: Some additional evidence. *Journal of Real Estate Practice and Education*, 15(2), 89–100.

[6] Benjaafar, S., Kong, G., Li, X., & Courcoubetis, C. (2015). Peer-to-Peer Product Sharing: Implications for Ownership, Usage and Social Welfare in the Sharing Economy. Working Paper.

[7] Carrillo, P. E. (2008). Information and real estate transactions: the effects of pictures and virtual tours on home sales. Working Paper.

[8] Cullen, Z., & Farronato, C. (2015). Outsourcing Tasks Online: Matching Supply and Demand on Peer-to-Peer Internet Platforms. Working Paper.

[9] Einav, L., Farronato, C., & Levin, J. (2016). Peer-to-Peer Markets. *Annual Review of Economics*, 8(1), 615–635.

[10] Fradkin, A. (2017). Search, Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb. Working Paper.

[11] Fraiberger, S. P., & Sundararajan, A. (2015). Peer-to-Peer Rental Markets in the Sharing Economy. Working Paper.

[12] Gong, J., Greenwood, B. N., & Song, Y. (2017). Uber Might Buy Me a Mercedes Benz: An Empirical Investigation of the Sharing Economy and Durable Goods Purchase. Working Paper.

[13] Gurran, N., & Phibbs, P. (2017). When Tourists Move In: How Should Urban Planners Respond to Airbnb? *Journal of the American Planning Association*, 83(1), 80–92.

[14] Horn, K., & Merante, M. (2017). Is Home Sharing Driving Up Rents? Evidence from Airbnb in Boston. *Journal of Housing Economics*, 38, 14–24.

[15] Jiang, B., & Tian, L. (2016). Collaborative Consumption: Strategic and Economic Implications of Product Sharing. *Management Science*.

[16] Kim, Y., Li, H., & Srinivasan, K. (2017). How Airbnb Affects Local Rental Market. Working Paper.

[17] Kroft, K., & Pope, D. G. (2014). Does online search crowd out traditional search and improve matching efficiency? Evidence from Craigslist. *Journal of Labor Economics*, 32(2), 259–303.

[18] Lee, D. (2016). How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendations. *Harvard Law and Policy Review*, 10(1), 229–253.

[19] Li, H., & Srinivasan, K. (2017). Competitive Dynamics in the Sharing Economy: An Analysis in the Context of Airbnb. Working Paper.

[20] PwC Report. (2015). Consumer Intelligence Series: The Sharing Economy.

[21] Razeghian, M., & Weber, T. A. (2016). To Share or Not to Share: Adjustment Dynamics in Sharing Markets. Working Paper.

[22] Razeghian, M., & Weber, T. A. (2017). The Impact of Sharing Markets on Product Durability. Working Paper.

[23] Seamans, R., & Zhu, F. (2014). Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers. *Management Science*, 60(2), 476–493.

[24] Sheppard, S., & Udell, A. (2016). Do Airbnb properties affect house prices? Working Paper.

[25] Taylor, C. R. (1999). Time-on-the-Market as a Sign of Quality. *The Review of Economic Studies*, 66(3), 555–578.

[26] van der Bijl, V. (2016). The effect of Airbnb on house prices in Amsterdam. University of Amsterdam.

[27] 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*.

Chapter III

Societal Impact of Machine Learning Deployment: The Case of Airbnb's Smart Pricing Algorithm

1 Introduction

Sharing economy platforms leverage technologies such as machine learning and artificial intelligence to facilitate trades in the sharing economy. For example, Airbnb controls search results using their search algorithm to facilitate matching between hosts and guests.[28] LendingClub and Prosper use credit score which is calculated by applying a mathematical algorithm to credit report dat. Uber relies extensively on machine learning (ML) to establish a robust and reliable dynamic pricing system that balances demand and supply, and thus, facilitates rides between riders and drivers.[29]

Airbnb's Smart Pricing algorithm is another example of leveraging technologies to facilitate trades. Deciding on the right prices to charge for Airbnb listings can be a challenging task for the hosts, especially because they are uncertain about demand. For example, lodging demand are subject to quite strong seasonal variation. While hosts may be able to take account for some general trends, the uncertainty about the demand still remains, which makes it difficult for hosts to price their listings. Launched in November 2015, Airbnb's Smart Pricing algorithm uses machine learning to provide price recommendations to hosts. The algorithm is based on Airbnb's rich data on lodging demand, the data which is not readily available to hosts. For example, Airbnb's Smart Pricing algorithm uses the data such as how many people are searching for and booking homes in the area. Hosts may opt in to the algorithm by setting a price range in which they would like their prices to fall. Then the price suggestions within the range are automatically adopted for all available nights.

Despite the potential benefits, only 21.8% of the listings used the algorithm at least once during the 21-month period from the launch of the algorithm to August 2017. In fact, many hosts argue that they

[28]See https://www.airbnb.com/resources/hosting-homes/a/how-airbnb-search-works-44.

[29]See https://www.forbes.com/sites/nicolemartin1/2019/03/30/uber-charges-more-if-they-think-youre-willing-to-pay-more/.

do not adopt the algorithm because the algorithm is biased towards Airbnb's interest, not hosts' interest. In particular, majority of the posts about Smart Pricing algorithm in Airbnb Hosts Forum speculate the presence of algorithmic bias, claiming that the algorithm suggests way low prices than they expect.[30] Thus, hosts face a trade-off coming from incentive misalignment when deciding whether to adopt the algorithm or not. On the one hand, the algorithm may benefit hosts as it is based on rich data that is not available to hosts. On the other hand, the algorithm may be designed to work for Airbnb not for hosts.

In this paper, we seek to answer the following two questions. First, is the algorithm biased, i.e., is it really favorable to Airbnb rather than hosts? If so, how much is the algorithm biased? Second, how does hosts' belief about algorithmic bias affect their adoption behavior, and thus, their profit? To answer the questions, we estimate a structural model of hosts' pricing and algorithm adoption using the Airbnb listings data. In order to model hosts' trade-off coming from the incentive misalignment, it is fundamental to include the two components in the model: (1) learning about demand from the algorithm's price suggestions and (2) learning about algorithmic bias by comparing the demand inferred from the price suggestions and the realized demand. On the one side, Airbnb has more rich data than hosts. Hosts can infer demand from the algorithmic suggested prices. On the other hand, the algorithm may be biased towards Airbnb's profit rather than hosts' profit. Hosts form beliefs about algorithmic bias and make decisions on whether to use the algorithm or not based on their beliefs.

The results show the presence of algorithmic bias in the Smart Pricing algorithm. Using the price suggestions from the algorithm does not always guarantee an increase in profit. For example, more than 40% of host-month observations are actually worse off when they use the price suggestions instead of their base prices. The profit from using the suggested prices are 18% less than the maximum profit that hosts can earn by using optimal prices for each month. In the counterfactual analysis, we vary two parameters related to hosts' belief about algorithmic bias and examine how they change the algorithm adoption, and thus, hosts' profit. We find that more correct belief about algorithmic bias and more precise signal from the price suggestions reduce the algorithm adoption, thereby increasing hosts' profit.

2 Literature Review

This paper is related to three streams of literature: (1) learning about demand, (2) algorithmic bias, and (3) pricing on the Airbnb platform. First, we contribute to the literature on learning about demand. Sellers are often uncertain about demand and they form and revise their beliefs using available information. The

[30] See, for example, https://airhostsforum.com/t/smart-pricing-needs-to-be-smarter/27926, https://airhostsforum.com/t/smart-pricing-have-you-used-it-have-it-improved-your-income/8846, and https://community.withairbnb.com/t5/Hosting/Why-you-should-never-use-Smart-Pricing/td-p/547981/page/3.

closest related paper is Pavlov and Berman (2019), who study the question of whether a sharing economy platform should set prices for the products on the platform (e.g., Uber), or let sellers set their own prices while providing price recommendations (e.g., Airbnb with Smart Pricing). In their analytical model, sellers infer demand from the platform's price recommendations and incorporate the information into their pricing decisions. Doraszelski, Lewis, and Pakes (2018) and Huang, Ellickson, and Lovett (2019) focus on demand uncertainty in new markets. Doraszelski, Lewis, and Pakes (2018) show that after the market for frequency response opens (in the UK electricity system), prices appear random at first and evolve to patterns that can be rationalized as a stationary equilibrium. Huang, Ellickson, and Lovett (2019) find that entrants to the new market of the privatized Washington State liquor market learn about demand over time and adapt towards fully-informed pricing. Jeon (2019) studies firm investment decisions with demand uncertainty due to changing conditions in the container shipping industry and finds that firms learn more from recent realizations of demand shocks than from those in the distant past. Lastly, Huang, Luo, and Xia (2019) study the seller learning under item-specific demand uncertainty in a used-car market.

Second, this paper is related to the literature of algorithmic bias. Algorithmic bias describes systematic and repeatable errors in a computer system that create unfair outcomes, such as privileging one arbitrary group of users over others. The study of algorithmic bias is most concerned with algorithms that reflect systematic and unfair discrimination. For example, studies in the context of online advertising show that ads for high-income jobs are presented to men much more often than to women and ads for arrest records are significantly more likely to show up on searches for distinctively black names (Datta et al. 2015; Sweeney 2013). Lambrecht and Tucker (2019) study why an advertisement algorithm delivered ads promoting job opportunities in the science, technology, engineering and math fields to more men than women. Rather than focusing on discrimination, our paper studies whether an algorithm works in one group's interest over another group's interest in the context of Airbnb's Smart Pricing algorithm.

Lastly, this paper contributes to the recent literature on the sharing economy (see Einav, Farronato, and Levin 2016 for a review of the sharing economy). In particular, our work relates to the stream of literature on pricing on the Airbnb platform. Li, Moreno and Zhang (2016) study the pricing decisions of Airbnb hosts and compare the performances of professional hosts who list multiple properties and nonprofessional hosts who only list one property. They find that professional hosts perform better than nonprofessional hosts because nonprofessional hosts are less likely to offer different rates across stay dates based on the underlying demand patterns. Focusing on the fact that hosts do not change their prices often, Pan (2019) and Olimov (2019) develop a structural model of dynamic pricing behavior on Airbnb and provide revenue and welfare implications of automated pricing. We build on the previous studies and differentiate from them by studying the pricing decisions when the Smart Pricing is available.

3 Data

3.1 Research Context

The Smart Pricing algorithm was first announced in November 2015 and launched in December 2015.[31] Before the launch of the algorithm, hosts had two options for pricing: base and custom. When hosts list their properties on Airbnb, they were required to set the "base" price, which will be the default price for all future periods. They could override the base price with "custom" price. Using custom price required hosts to manually visit the calendar settings webpage and update price for a specific period.

After the launch of the algorithm, hosts have three options for pricing: base, algorithm, and custom. Figure 1 presents screenshot of the pricing webpage for each pricing method. First, hosts can turn off Smart Pricing and use "base" price. When Smart Pricing is turned off, the default price for all future periods is the base price. Second, hosts can turn on Smart Pricing and use "algorithmic" price. Again, turning Smart Pricing on makes the algorithmic price as the default price for all future periods. Note that turning on Smart Pricing requires hosts to set the price range that hosts would like their prices to fall. Third, hosts can override either base or smart price with "custom" price. Because Airbnb shows price suggestions on the calendar settings webpage, hosts must review the suggested prices when using custom price. Using custom price does not change the default price method.

3.2 Data Description

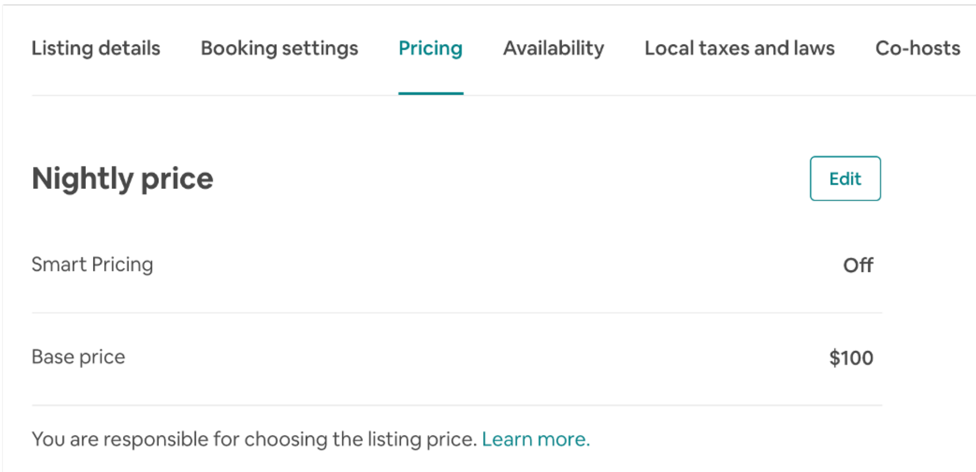
Our data span from October 2014 to August 2017 and consist of 275 randomly selected Airbnb properties located in two zip codes (78702 and 78722) of East Austin area of Austin, TX. We observe detailed information on characteristics, such as number of bedrooms and bathrooms, listing type (entire home, private room, or shared room), property type. The data also includes the price and the pricing method (base, custom, or algorithm), availability and booking information for each day.

The data come from two sources, which is summarized in Table 1. The first source is a third-party company, Airdna, that specializes in data collection and analysis. An advantage of this data set is that it contains not only consumer-facing information such as property characteristics, but also backstage information such as availability (i.e., open for booking) and booking information. The latter is crucial for accurately measuring Airbnb demand. A property may be unavailable to consumers because the host chooses to block the calendar or the property has been booked. By observing both availability and booking information, we can effectively distinguish between the two cases. Second, we collected consumer-facing information includ-

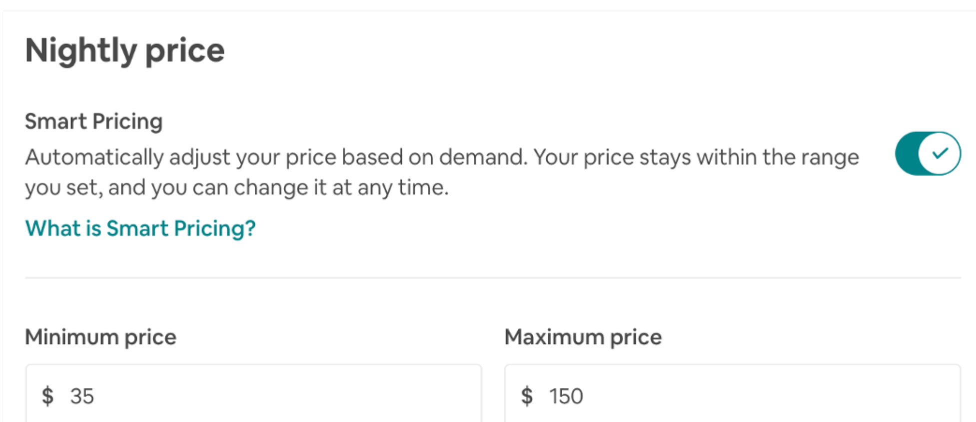
[31] See https://blog.atairbnb.com/new-features-revealed-at-ao2015/ and https://www.cnbc.com/2015/11/12/airbnb-launches-smart-pricing-for-hosts.html.

Figure 1: Pricing Methods

(a) Base



(b) Algorithm



(c) Custom

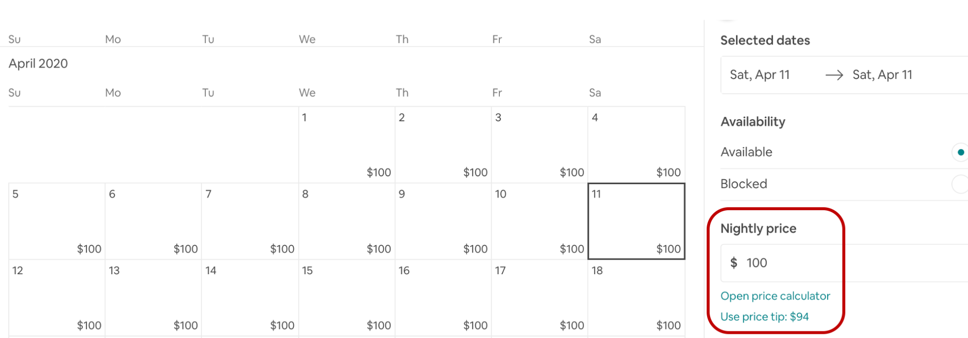


Table 1: Data Source

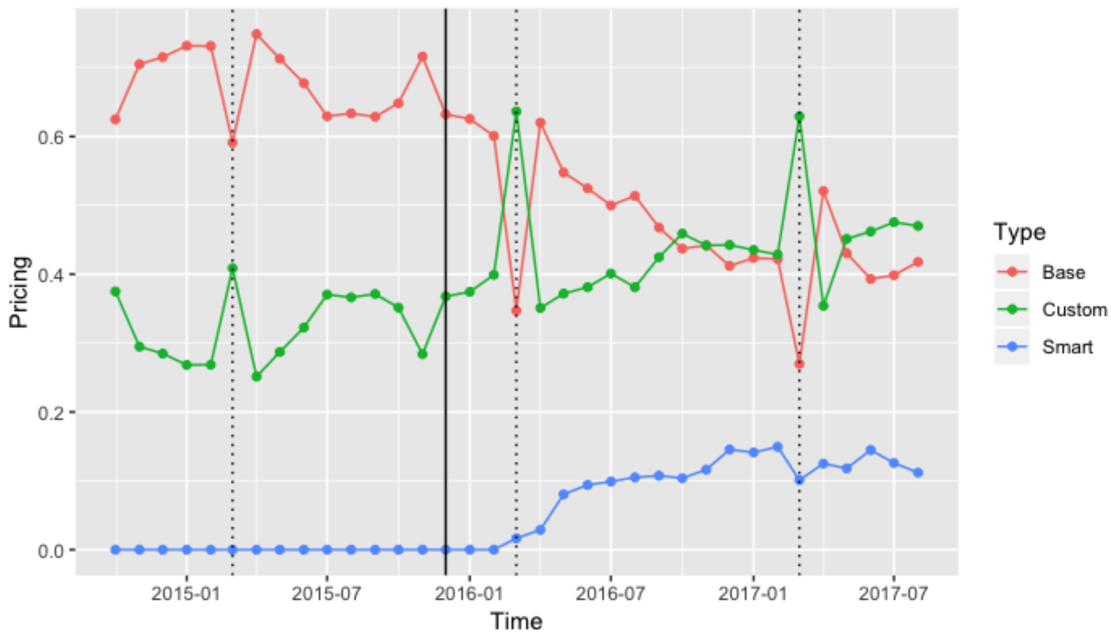
	Airdna	Web-scraped
Time Frame	Oct 2014 - Aug 2017	Dec 2015 - Aug 2017
Property Characteristics	Yes	Yes
Price	Yes	Yes
Pricing Method	No	Yes
Availability and Booking	Yes	No

ing availability calendars from the property webpages. The calendar information includes pricing method (base, custom, or algorithm) for each day. The data collection was conducted between December 2015 and August 2017, which covers the time period right after the launch of the Smart Pricing algorithm.

Note that our data lack two kinds of information that is essential for the model. First, the pricing method data between October 2014 and November 2015 is not available as the web-scraping data starts from December 2015. Based on the fact that 61.7% of the listings do not change their base price during the 21-month time frame of web-scraped data, we impute the base price before December 2015 using the first base price in the web-scraped data. Second, prices suggested by the Smart Pricing algorithm are available only for the days that properties use the algorithm and only for the properties that adopted the algorithm. For the days and properties that suggested prices data are not available, we replicated the algorithm following Ye et al. (2018) and generated the suggested prices. Ye et al. (2018) is a paper written by the current/former members of the Pricing Modeling team at Airbnb and describes the Smart Pricing algorithm in detail. The mean absolute percentage error (MAPE) for the replicated data is 7%. We also impute the minimum price for the Smart Pricing algorithm using the lowest price used by each host.

We aggregate the daily price, pricing method, occupancy data at the monthly level. For each listing-month observation, we select the most used pricing method as the representative pricing method and average the price and occupancy over the days when hosts set prices using the representative pricing method. The main reason for the data aggregation is because the differentiated product demand model (explained in Section 4) cannot be estimated by the standard estimation procedure (Berry, Levinsohn, and Pakes (1995), BLP for short) when data includes zero-share observations. However, the data aggregation naturally leads to some information loss; in our case, 64% of listing-month observations use the same pricing method in a given month. A future version of the paper may reduce the information loss by aggregating at a more granular level and using new approaches to estimating differentiated product demand models in the presence of zero empirical shares (e.g., Gandhi, Lu, and Shi 2019). For example, 93% of the listing-time observations use the same pricing method in a given time period when aggregated at the week-weekday/weekend level.

Figure 2: Pricing Method Choice



Note: Vertical line represents the launch of the Smart Pricing algorithm.
Dotted vertical lines represent March of 2015, 2016, and 2017.

3.3 Model-Free and Reduced-Form Insights

Figure 2 presents the proportion of hosts choosing the three pricing methods: base, custom, and algorithm. The vertical line represents the launch of the algorithm in December 2015, and the dotted vertical lines represent March of each year. The proportion of the base pricing method decreases over time, while the proportion of the custom pricing method increases over time. The proportion of the algorithmic pricing method increases right after the launch of the algorithm, but flattens in the 10-15% range.

Custom Pricing. As seasonality is a fundamental feature of the travel and tourism industry, it is crucial for hosts to respond to seasonal demand and charge different prices in order to maximize their profit. Table 2 presents the results of the regression of revenue on the use of custom pricing method.[32] The first column shows that using custom pricing method is associated with the revenue increase of $22.3 per day.

However, setting custom price requires hosts to come up with an appropriate price for the night, which potentially incurs costs such as hassle cost and menu cost. Hosts will therefore use custom price only when the revenue increase compensates the cost. This explains why the proportion of the custom pricing method

[32]Note that the regression may suffer from an endogeneity issue because the use of custom pricing method is not random across the hosts. This selection issue can be addressed by using Propensity Score Matching (PSM) to create a matched sample of treatment and control groups, i.e., users and non-users custom pricing. We find that the results presented in Table 2 are robust; the results from using the Propensity Score Matching do not differ qualitatively from the results presented in Table 2.

Table 2: Regression of Revenue on the Use of Custom Pricing Method

DV: Revenue	(1)	(2)	(3)
Custom Pricing	22.3***	6.55	11.1*
	(5.20)	(5.47)	(5.98)
Custom Pricing ×After Launch		24.8***	16.7**
		(6.66)	(7.60)
Months on Airbnb			0.52***
			(0.19)
Intercept	91.3***	90.8***	83.1***
	(2.58)	(2.63)	(4.07)
Listing FE	Yes	Yes	Yes
N	4,351	4,351	4,351

Note: Standard errors clustered at the listing level are shown in parentheses.
Significance levels: * p<0.1, ** p<0.05, *** p<0.01

soars in March of each year in Figure 2. In mid-March, South by Southwest (SXSW), which is a collection of music, film and interactive conferences and festivals, takes place in Austin, Texas. SXSW is the single most profitable event for the City of Austin's hospitality industry, with its economic impact estimated to be $348.6 million in 2017.[33] As the demand for lodging peaks in the SXSW season, hosts can charge much higher prices compared to slow season prices. Hosts are more likely to use custom price in the SXSW season because the revenue increase from charging high prices outweighs the cost of setting custom prices.

Recall that Airbnb displays price suggestions on the calendar settings webpage where hosts must visit to set custom prices. Since the price recommendations leverage the data that are only available to Airbnb but not hosts, the price suggestions may serve as a signal for the demand, and thus, help hosts set better prices. To examine whether reviewing the price suggestions helps hosts set better prices, we regress revenue on the use of custom pricing method and its interaction with whether it is after the launch of the Smart Pricing algorithm. The second column of Table 2 shows that the revenue-increasing effect of the custom pricing mostly comes from the time periods after the algorithm launch, which indicates that reviewing price suggestions helps hosts set better prices. The result is consistent when we control for the potential improvement of hosts' pricing ability, i.e., the number of months listed on Airbnb since October 2014, as shown in the third column of Table 2.

Algorithmic Pricing. The adoption rate for the Smart Pricing algorithm is quite low; only 21.8% of the listings used the algorithmic pricing method at least once during the sample period. There may be two reasons for the low adoption rate. First, hosts tend to maintain the default pricing method from the previous period. In other words, they rarely turn on or off the Smart Pricing algorithm. Table 3 shows

[33] See https://www.sxsw.com/wp-content/uploads/2016/05/2017-SXSW-Economic-Impact-Press-Release.pdf

Table 3: Turning Smart Pricing On or Off

			Month t		
			Off	On	Total
Month $t-1$	Off	(N)	2,168	44	2,212
		(%)	98.01	1.99	100
	On	(N)	6	234	240
		(%)	2.50	97.50	100

Table 4: Difference Between Actual and Suggested Prices

Variable	Mean	Std. Err.	Std. Dev	[95% Conf.	Interval]
Suggested Price	262.89	4.22	250.77	254.62	271.15
Actual price	277.74	4.76	283.36	268.40	287.08
Difference	-14.85	1.62	96.41	-18.03	-11.67

that among all host-month observations, only 2.04% $(= (44+6)/(2,212+240))$ of them change the default pricing method. When their default pricing method in the previous time period is base, 1.99% turn on the Smart Pricing algorithm and change their pricing method to algorithm. Similarly, when their default pricing method in the previous time period is algorithm, 2.50% turn off the Smart Pricing algorithm and change their pricing method to base. Second, the algorithm may work in Airbnb's interest rather than in hosts' interests.[34] In particular, many hosts argue that the algorithm suggests prices lower than they expect, which is also confirmed from the data. Table 4 presents the results of the paired t-test for testing the equality of means of the hosts' own and algorithm-suggested prices. The results indicate potential algorithmic bias in the Smart Pricing algorithm. The algorithm-suggested prices are, on average, \$14.85 lower than the hosts' own prices, the difference which is significantly different from zero.

We also examine whether hosts' decision to keep using the algorithm depends on their belief of the algorithmic bias. In particular, we analyze whether hosts who experienced less algorithmic bias from their previous use of the algorithm are more likely to keep using the algorithm. Table 5 presents the results of the regression of the use of the algorithmic pricing method on average revenue from past algorithmic prices and whether they had used the algorithm before. We find that the larger the average revenue from past algorithmic prices is, the more likely hosts are to use the algorithm. In addition, hosts are less likely to use the algorithm when they have had never used the algorithm before. The results remain consistent when we further control for the time and month fixed effects.

[34]See, for example, https://airhostsforum.com/t/turned-on-smart-pricing-but-didnt-like-it/6674/6, or https://airhostsforum.com/t/smart-pricing-needs-to-be-smarter/27926.

Table 5: Regression of Smart Pricing Use on Past Smart Pricing Revenue

DV: Use Algorithmic Pricing	(1)	(2)	(3)
Past Smart Pricing Revenue [10^3]	0.300*	0.311**	0.316**
	(0.155)	(0.156)	(0.156)
Never Used Smart Pricing Before	-0.656***	-0.658***	-0.657***
	(0.016)	(0.016)	(0.016)
Months on Airbnb		-0.0006	-0.0005
		(0.0004)	(0.0005)
Intercept	0.656***	0.670***	0.672***
	(0.015)	(0.018)	(0.020)
Month FE			Yes
Adj R-squared	0.6507	0.6508	0.6512
N	2,821	2,821	2,821

Note: Standard errors are shown in parentheses.
Significance levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$

4 Model Setup

4.1 Model Overview

Figure 3 illustrates the timing/sequence of the events for each month. First, at the beginning of each month, hosts are aware of the guest demand, except that they are uncertain about the demand shock from seasonality. They have prior beliefs on seasonality shock, which is estimated from the guest demand model. In addition, they have beliefs about the bias of the algorithm. At this time, their Smart Pricing is either on or off. They also observe the states of their peers.

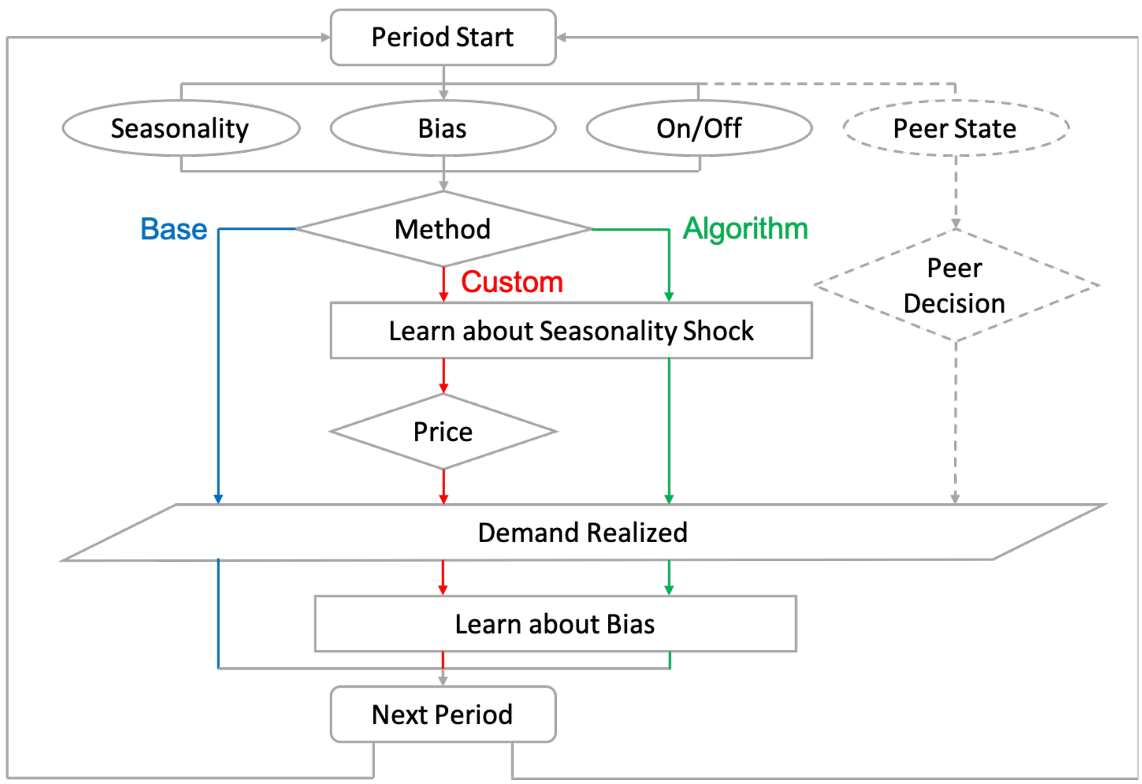
Second, hosts decide pricing method among base, custom, and algorithm based on their prior beliefs about the seasonality shock and the bias of the algorithm. Note that the hosts are not yet to observe the prices suggested by the algorithm, and thus, they compute the expected profit from the algorithmic prices based on their beliefs about bias of the algorithm, instead of the actual prices suggested by the algorithm.

Third, if hosts chose either custom or algorithm in Step 2, they observe the prices suggested by the algorithm. The algorithmic prices serve as a signal about the seasonality shock, because the seasonality shock is the only missing piece of information about the guest demand. Hosts update their belief about the seasonality shock using the signal. Note that the seasonality shock is assumed to be i.i.d over time. There is learning within a given period, but no learning over different time periods.

Fourth, if hosts chose custom in Step 2, they set prices that maximizes the current period profit. The pricing decision is based on hosts' posterior information on the seasonality shock.

Fifth, guest demand is realized. Observing the realized demand and observing the realized seasonality shock are equivalent because the seasonality shock is the only missing piece of information about the guest

Figure 3: Timing of Events



demand.

Lastly, if hosts chose custom or algorithm in Step 2 and observed the algorithmic suggested prices in Step 3, they observe the difference between the realized seasonality shock and the shock inferred from the algorithmic prices, the difference which serves as a signal for bias of the algorithm. Hosts update their beliefs about the bias of the algorithm using the signal. Note that unlike the seasonality shock, bias is learned over different time periods and the posterior belief about the bias at the current time period is the prior at the next time period.

4.2 Hosts' Beliefs

4.2.1 Seasonality Shock

Suppose there are J_t Airbnb listings in the market at time t. In each time period t, guest i chooses at most 1 property from the J_t alternatives. Guests are also allowed to choose an outside option (denoted as $j = 0$, e.g., staying at a hotel or at a friend's house). The utility from choosing property j is:

$$u_{ijt} = \begin{cases} x_{jt}\beta - \alpha p_{jt} + \tau_{it} + \xi_{jt} + \varepsilon_{ijt} & \text{if } j \neq 0 \\ \varepsilon_{ij0} & \text{if } j = 0 \end{cases}$$

where x_{jt} represents property and host characteristics (e.g., the number of bedrooms and bathrooms, listing type (entire place or private/shared room), location, response rate/time, and superhost), p_{jt} is average daily rate, τ_{it} is seasonality at time t, ξ_{jt} is characteristics unobserved to researchers, and ε_{jt} is type-1 extreme value utility shocks.

Note that the seasonality term, τ_{it}, is heterogeneous across guests. We can decompose the seasonality term as the sum of mean seasonality τ_t and the heterogeneous shock $\psi_{it} \sim N(0, \sigma^2_{\psi_t})$, i.e., $\tau_{it} = \tau_t + \psi_{it}$. Thus, the utility function can be written as:

$$u_{ijt} = \begin{cases} x_{jt}\beta - \alpha p_{jt} + \tau_t + \psi_{it} + \xi_{jt} + \varepsilon_{ijt} & \text{if } j \neq 0 \\ \varepsilon_{ij0} & \text{if } j = 0 \end{cases} \tag{1}$$

Hosts are fully aware of the guest demand (including the unobserved characteristics term ξ_{jt}), except that they are uncertain about the seasonality shock ψ_{it}. Though they do not know the exact value of the seasonality shock, they have rational beliefs about this shock. In other words, their beliefs are the same as the true empirical distribution of ψ_{it} in the market, which is $\psi_{it} \sim N(0, \sigma^2_{\psi_t})$.

The Smart Pricing algorithm potentially has more information about the seasonality shock than hosts do. For example, Airbnb is able to observe how many people are searching for listings in the market, while individual hosts are not able to do so. The Smart Pricing algorithm reflects the information that is not available to hosts when providing price recommendations to hosts.

4.2.2 Bias of the Smart Pricing Algorithm

We assume that hosts' belief about the algorithmic suggested price takes the following form:

$$p_{jt}^{A\text{tip}} = p_{jt}^{*}(1 + \Delta_{jt} + \nu_{jt}) \tag{2}$$

where $p_{jt}^{A\text{tip}}$ is the algorithm suggested price that is uncensored by minimum and maximum range set by the host; p_{jt}^{*} is the optimal price under complete knowledge of the seasonality shock, ψ_{it}, which is known to the algorithm but not to hosts; Δ_{jt} is host's belief about the bias of the algorithm; and $\nu_{jt} \sim N(0, \sigma_\nu^2)$ is the signal noise of the algorithmic suggested prices. In other words, the algorithmic suggested price is the product of the optimal price under complete knowledge of the seasonality shock, p_{jt}^{*}, and the bias term, $(1 + \Delta_{jt} + \nu_{jt})$.

Hosts form and update beliefs about the bias Δ_{jt}, which represents how far the algorithmic suggested prices are from the optimal price. Note that we take the algorithm as given and remain agnostic about why and how the algorithm is biased, as the purpose of the paper is to study the algorithm adoption behavior given the algorithm.

When needed for computational reasons, we approximate Equation (2) using:

$$p_{jt}^{A\text{tip}} = p_{jt}^{*} + p_{jt}^{o*}(\Delta_{jt} + \nu_{jt}) \tag{3}$$

where p_{jt}^{o*} is the optimal price under zero seasonality shock, i.e., $\psi_{it} = 0$.

4.3 Pricing Method Choice

Hosts decide pricing method among base, custom, and algorithm, $d_{jt} \in \{B \text{ (Base)}, C \text{ (Custom)}, A \text{ (Algorithm)}\}$. The choice is a dynamic problem because of the following two inter-temporal trade-offs. First, turning the algorithm on or off, which incurs cost of doing so, changes the default pricing method for all future periods. For example, hosts may not choose algorithm as their pricing method (i.e., turn on the Smart Pricing algorithm) even though it gives the highest payoff in the current period, if they expect the future algorithmic prices to be significantly suboptimal. It is because turning on the algorithm sets the default pricing method for the future periods will be the algorithm – unless they turn off Smart Pricing which incurs the cost of turning it on/off. Second, choosing custom or algorithm pricing method allows hosts to learn about the bias of the algorithm, which will help them make better decisions in the future periods.

4.3.1 Profit Function and Optimal Price

The profit from hosting on Airbnb for property j at time t is:

$$\Pi_{jt} = (p_{jt} - c_j) \cdot ms_{jt} \cdot M$$

where p_{jt} is daily price, c_j is marginal cost, ms_{jt} is market share, and M is market size. Market share ms_{jt} is defined as the number of property-nights sold in each time period, divided by the market size. A property's market share is derived from the guests' utility function explained in Equation (1). Considering that a property's market share cannot exceed the total number of days in each month divided by the market size, the market share is:

$$ms_{jt} = \min\left\{\frac{\exp\left(x_{jt}\beta - \alpha p_{jt} + \tau_t + \psi_{it} + \xi_{jt}\right)}{1 + \sum_{j' \in J_t} \exp\left(x_{j't}\beta - \alpha p_{j't} + \tau_t + \psi_{it} + \xi_{j't}\right)}, \frac{T_t}{M}\right\} \tag{4}$$

where T_t is the total number of days in month t. We denote $\bar{ms}_{jt} \equiv T_t/M$.

Market size measures the total number of nights (including Airbnb properties, hotels, and other alternatives) that could possibly be sold to the travelers. We approximate the market size using the tourism statistics and the distribution of hotel supply across the regions in the Austin MSA (Metropolitan Statistical Area). Specifically, market size is defined as the product of average monthly visitors in Austin MSA, average length of stay, and the percentage of hotels in the sample area of Austin MSA. We use a constant market size of 36,800 (= 25.6M/12 * 1.96 * 0.0088).[35]

Assuming the existence of a pure-strategy Bertrand-Nash equilibrium in prices, and that the prices that support it are strictly positive, the prices must satisfy the first order condition without considering the capacity constraint $\bar{ms}_{jt}$:

$$\frac{\exp\left(x_{jt}\beta - \alpha p_{jt} + \tau_t + \psi_{it} + \xi_{jt}\right)}{1 + \sum_{j' \in J_t} \exp\left(x_{j't}\beta - \alpha p_{j't} + \tau_t + \psi_{it} + \xi_{j't}\right)} + (p_{jt} - c_j)\frac{\partial}{\partial p_{jt}}\left(\frac{\exp\left(x_{jt}\beta - \alpha p_{jt} + \tau_t + \psi_{it} + \xi_{jt}\right)}{1 + \sum_{j' \in J_t} \exp\left(x_{j't}\beta - \alpha p_{j't} + \tau_t + \psi_{it} + \xi_{j't}\right)}\right) = 0 \tag{5}$$

If the market share with the price implied by the first order condition exceeds the maximum capacity $\bar{ms}_{jt}$,

[35] 25.6 million travelers visited Austin MSA in 2016. The average length of stay was 1.96 days, including overnight and day trips. Though the MSA-level tourism data is the most granular data available, we can approximate the market size for East Austin area of the MSA from the distribution of hotel supply (number of hotel rooms). As hotels choose locations with high demand, the distribution of hotels can capture the distribution of travel demand across areas.
See http://www.austintexas.gov/edims/document.cfm?id=305298 for the number of travelers and http://www.austintexas.gov/edims/document.cfm?id=276727 for the average length of stay.

then the optimal price can be computed from the fact that the profit from the logit demand and the profit from the maximum capacity are equivalent:

$$(p_{jt}^{*}-c_j)\cdot\frac{\exp\left(x_{jt}\beta-\alpha p_{jt}+\tau_t+\psi_{it}+\xi_{jt}\right)}{1+\sum_{j'\in J_t}\exp\left(x_{j't}\beta-\alpha p_{j't}+\tau_t+\psi_{it}+\xi_{j't}\right)}=(p_{jt}^{*}-c_j)\cdot\frac{T_t}{M}\tag{6}$$

Denoting $b_{jt}\equiv x_{jt}\beta+\tau_t+\xi_{jt}$, the optimal price can be expressed as:

$$p_{jt}^{*}=\begin{cases}c_j+\frac{1}{\alpha}+\frac{1}{\alpha}W\left(\frac{\exp(b_{jt}-\alpha c_j-1)}{\exp(-\psi_{it})+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)}\right) & \text{if } ms_{jt}\neq \bar{ms}_{jt}\\ \frac{1}{\alpha}\left[b_{jt}-\log\left(\frac{T_t}{M-T_t}\right)-\log\left(\exp\left(-\psi_{it}\right)+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)\right)\right] & \text{if } ms_{jt}=\bar{ms}_{jt}\end{cases}\tag{7}$$

where $W(.)$ is the Lambert W function. The derivation is provided in Appendix A.

4.3.2 Per Period Payoff

Denote Π_{jt}^{d} as the profit from decision $d\in\{B\text{ (Base)},C\text{ (Custom)},A\text{ (Algorithm)}\}$. Define Π_{jt}^{0} as the profit from the default pricing method, i.e., profit from the algorithmic price if the Smart Pricing algorithm is on and profit from the base price if the algorithm is off.

$$\Pi_{jt}^{0}\equiv ON_{jt}\cdot\Pi_{jt}^{A}+(1-ON_{jt})\cdot\Pi_{jt}^{B}$$

Per period payoff from choosing each pricing method is:

$$\begin{aligned}U_{jt}^{B}&=E\left[\Pi_{jt}^{B}\right]-E\left[\Pi_{jt}^{0}\right]-ON_{jt}\cdot c^{\text{on/off}}+\epsilon_{jt}^{B}\\&=ON_{jt}\cdot\left(E\left[\Pi_{jt}^{B}\right]-E\left[\Pi_{jt}^{A}\right]-c^{\text{on/off}}\right)+\epsilon_{jt}^{B}\\U_{jt}^{C}&=E\left[\Pi_{jt}^{C}\right]-E\left[\Pi_{jt}^{0}\right]-c^{\text{menu}}+\epsilon_{jt}^{C}\\U_{jt}^{A}&=E\left[\Pi_{jt}^{A}\right]-E\left[\Pi_{jt}^{0}\right]-(1-ON_{jt})\cdot c^{\text{on/off}}+\epsilon_{jt}^{A}\\&=(1-ON_{jt})\cdot\left(E\left[\Pi_{jt}^{A}\right]-E\left[\Pi_{jt}^{B}\right]-c^{\text{on/off}}\right)+\epsilon_{jt}^{A}\end{aligned}\tag{8}$$

where c^{menu} is the cost of setting custom price and $c^{\text{on/off}}$ is the cost of turning the Smart Pricing algorithm on or off. The idiosyncratic error terms $\left\{\epsilon_{jt}^{B},\epsilon_{jt}^{C},\epsilon_{jt}^{A}\right\}$ are assumed to be i.i.d. extreme value type I errors. The payoff of each pricing method $d\in\{B,C,A\}$ comes from the difference between the expected profit from the pricing method d and the expected profit from the default pricing method ($E\left[\Pi_{jt}^{d}\right]-E\left[\Pi_{jt}^{0}\right]$) minus the costs if applicable. Because the payoff considers the expected profit *difference*, not the expected profit from each pricing method, the deterministic part of the payoff of the default pricing method is automatically normalized to zero. In addition to the expected profit difference, hosts incur the cost of setting custom price c^{menu} if they choose custom pricing method, and the cost of turning the algorithm on or off $c^{\text{on/off}}$ if they choose base when Smart Pricing is on and if they choose algorithm when Smart Pricing is off.

Note that hosts use *expected* profit because they only know the distribution, not the exact value, of the seasonality shock. In addition, algorithmic price is also a random variable because hosts are yet to observe the algorithmic suggested price at the time of the pricing method choice. The expected profit from each pricing method can be written as:

Base.

$$\begin{aligned} E\left[\Pi_{jt}^{B}\right] &= E_{\psi_{it}}\left[(p_{jt}^{B}-c_j)\cdot ms_{jt}(p^{B},\psi_{it})\cdot M\right] \\ &= (p_{jt}^{B}-c_j)\cdot E_{\psi_{it}}\left[ms_{jt}(p^{B},\psi_{it})\right]\cdot M \end{aligned}$$

Custom.

$$\begin{aligned} E\left[\Pi_{jt}^{C}\right] &= E_{\psi_{it}}\left[\left(p_{jt}^{*}(\psi_{it})-c_j\right)\cdot ms_{jt}\left(p_{jt}^{*}(\psi_{it}),\psi_{it}\right)\cdot M\right] \\ &= E_{\psi_{it}}\left[\left(p_{jt}^{*}(\psi_{it})-c_j\right)\cdot ms_{jt}\left(p_{jt}^{*}(\psi_{it}),\psi_{it}\right)\right]\cdot M \end{aligned}$$

Note that not only the market share ms_{jt} but also the optimal price p_{jt}^{*} depend on the seasonality shock.

Algorithm. Denote $p_{jt}^{A\text{tip}}$ as the algorithm suggested price that is uncensored by minimum and maximum range set by the host. At the time of the pricing method choice, the algorithmic suggested price is a random variable, as Equation (3) suggests: $p_{jt}^{A\text{tip}} = p_{jt}^{*}(\psi_{it})+p_{jt}^{o*}(\Delta_{jt}+\nu_{jt}) \sim N\left(p_{jt}^{*}(\psi_{it})+p_{jt}^{o*}\bar{\Delta}_{jt},(p_{jt}^{o*})^2(\sigma_{\Delta jt}^2+\sigma_{\nu}^2)\right)$.

$$E\left[\Pi_{jt}^{A\text{tip}}\right] = E_{\psi_{it}}E_{p_{jt}^{A\text{tip}}}\left[\left(p_{jt}^{A\text{tip}}-c_j\right)\cdot ms_{jt}(p^{A\text{tip}},\psi_{it})\right]$$

We also need to consider the minimum and maximum range set by the host. Denote $p_{jt}^{A\text{min}}$ as the minimum Smart Pricing price set by the host. Because there is no reason that hosts would limit the maximum profit, we do not consider the maximum range. The expected profit from the minimum Smart Pricing price is:

$$\begin{aligned} E\left[\Pi_{jt}^{A\min}\right] &= E_{\psi_{it}}\left[\left(p_{jt}^{A\min}-c_j\right)\cdot ms(p_{jt}^{A\text{min}},\psi_{it})\cdot M\right] \\ &= \left(p_{jt}^{A\min}-c_j\right)\cdot E_{\psi_{it}}\left[ms(p_{jt}^{A\text{min}},\psi_{it})\right]\cdot M \end{aligned}$$

Thus, the expected profit from the algorithmic pricing method is:

$$E\left[\Pi_{jt}^{A}\right] = \max\left\{E\left[\Pi_{jt}^{A\text{tip}}\right],E\left[\Pi_{jt}^{A\min}\right]\right\}$$

The expectations over ψ_{it} and p_{jt}^{Atip} are calculated using finite discrete approximation to the normal distribution.[36]

4.3.3 State Variables and State Transitions

This section defines the state variables that affect an individual host's payoff over time and discusses the dynamics in the transition of individual states driven by hosts' decisions. The set of state variables for individual j at time t is $s_{jt} = \{ON_{jt}, \Delta_{jt}\}$, where ON_{jt} indicates whether the Smart Pricing algorithm is on or off, Δ_{jt} is the host's belief about the bias of the algorithm. Similarly, denote the set of state variables for host j's peers at time t as s_{-jt}. Hosts' decision will depend not only on their own state, but also on the states of their peers. We describe the transition of each state variable below.

First, the transition of ON_{jt} is governed by individual j's choice of the pricing method in every period. For example, if individual j chose algorithm (base) as their pricing method at time t, meaning they turn on (off) the Smart Pricing algorithm, the state at time $t+1$ will be $ON_{j,t+1} = 1$ ($ON_{j,t+1} = 0$). If a host chose custom as his/her pricing method, then the state of $ON_{j,t+1}$ remains the same. That is:

$$ON_{j,t+1} = 1(d_{jt} \neq C) \cdot 1(d_{jt} = A) + 1(d_{jt} = C) \cdot ON_{jt}$$

Second, hosts expect that as they learn about the bias of the algorithm, their beliefs about the algorithmic bias become more precise, i.e., the variance $\sigma^2_{\Delta_{jt}}$ decreases. The evolution of the variance of the bias is explained in detail in Section 4.6. However, there is no way that hosts expect the evolution of the mean of the bias. This is because the mean of the bias is updated based on the difference between the optimal price given the realized seasonality shock and the price suggested by the algorithm. At the time of the pricing method choice, hosts are not aware of neither the seasonality shock realized at the end of the time period nor the algorithmic suggested price as they are not yet to observe the algorithmic suggested prices. Thus, we assume that hosts use the current mean of the bias for all future periods, i.e.,

$$\bar{\Delta}_{j,t' \geq t} = \bar{\Delta}_{jt}$$

Note that our assumption is different from the rational expectation assumption in typical dynamic learning models, where individuals are assumed to be aware of the true or empirical state transitions.

Third, hosts also base their decisions on the state of their peers. Given the current states and their own private shock, hosts make decisions simultaneously and compete with each other in each period. A proper equilibrium concept for this dynamic game is Markov Perfect Equilibrium (Ericson and Pakes 1995, hereafter

[36] See https://www.pivoters.com/math/normal-mass-function/

MPE). However, solving for MPE is computationally intractable in our context with a large number of hosts. To address the problem of curse of dimensionality, Weintraub et al. (2008, 2010) proposed an equilibrium concept called Oblivious Equilibrium, which approximates the MPE. The key notion is that in a market with large number of players, the simultaneous changes in each individual's moves can be averaged out. As a result, the average industry state either remains stationary over time or can be tracked as a deterministic trajectory changing with a stationary (steady) pace. Thus, each player does not need to track everyone's state over time; instead, it is sufficient for one to make a near-optimal decision by considering only her own state and the average industry state.

Following the idea of Oblivious Equilibrium and the implementation of Pan (2019), we assume that each host's decision is based on her own state and a set of system states which evolve deterministically. Recall from Equation (4) that market share of listing j at time t is:

$$ms_{jt} = \min\left\{\frac{\exp\left(b_{jt} - \alpha p_{jt}\right)}{\exp\left(-\psi_t\right) + \exp\left(b_{jt} - \alpha p_{jt}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)}, \bar{ms}_{jt}\right\}$$

Denote $ss_{jt} \equiv \log\left(\sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)\right)$. In our context, the number of peers is large, and thus, ss_{jt} would be almost across j, i.e., $ss_{jt} \approx ss_t$. Hence, we can rewrite Equation (4) as

$$ms_{jt} = \min\left\{\frac{\exp\left(b_{jt} - \alpha p_{jt}\right)}{\exp\left(-\psi_t\right) + \exp\left(b_{jt} - \alpha p_{jt}\right) + \exp\left(ss_t\right)}, \bar{ms}_{jt}\right\}$$

where ss_t is defined as the system state. Hosts only consider their own states and the system states ss_t, which summarizes the number of the competitors and their price decisions. The advantage of this simplification is that while the model becomes tractable, competition is not entirely ignored. We assume that the system states evolve deterministically and the path of ss_t is perceived by all hosts.

4.3.4 Dynamic Optimization Problem

On an infinite horizon, each host j chooses an infinite sequence of actions of pricing method $\{d_{jt}\}_{t=t'}^{\infty}$ to maximize the sum of expected lifetime payoff:

$$\max_{d_{jt}} \sum_{t=t'}^{\infty} \rho^{t-t'} E_{\{s_{jt}\}}\left[U_{jt}\left(d_{jt} | s_{jt}\right)\right]$$

where ρ is discount factor, $s_{jt} = \{ON_{jt}, \Delta_{jt}, ss_t\}$ denote the transitioned individual state (including in the next period for host j, and U_{jt} is the per-period payoff.

Let $V(s_{jt})$ denote the value function for host j at time t:

$$V(s_{jt}) = \max_{d_{jt}} \sum_{t=t'}^{\infty} \rho^{t-t'} E_{\{s_{jt}\}} \left[U_{jt} \left(d_{jt} | s_{jt} \right) \right]$$

In an infinite horizon optimization problem, the above equation can be solved through equation (Bellman 1957).

$$V_j(s_{jt}) = \int \max_{d_{jt}} \left\{ U_{jt}(s_{jt}) + \rho \int V_j(s'_{jt}) f(s'_{jt} | s_{jt}, d_{jt}) ds'_{jt} \right\} g(\epsilon_{jt}) d\epsilon_{jt}$$

4.4 Price Setting

At each time period t, hosts who use custom pricing set their own price for their listings. They choose the price that maximizes the expected profit at time period t. Thus, pricing is a static decision, unlike the pricing method choice which is a dynamic decision. Given hosts' belief on the seasonality shock $F(\psi_{it})$, the expected profit is:

$$E_{\psi_{it}} \left[\Pi_{jt} \right] = E_{\psi_{it}} \left[(p_{jt} - c_j) \cdot ms_{jt} \cdot M \right]$$

where

$$ms_{jt} = \min \left\{ \frac{\exp\left(b_{jt} - \alpha p_{jt}\right)}{\exp\left(-\psi_{it}\right) + \exp\left(b_{jt} - \alpha p_{jt}\right) + \exp\left(ss_t\right)}, \bar{ms}_{jt} \right\}$$

Thus, the optimal price satisfies:

$$\begin{cases} p^*_{jt} - c_j = \dfrac{\int \frac{\exp\left(b_{jt} - \alpha p^*_{jt}\right)}{\exp(-\psi_{it}) + \exp\left(b_{jt} - \alpha p^*_{jt}\right) + \exp(ss_t)} dF(\psi_{it})}{\int \frac{\alpha \cdot \left[\exp(-\psi_{it}) + \exp(ss_t)\right] \cdot \left[\exp\left(b_{jt} - \alpha p^*_{jt}\right)\right]}{\left[\exp(-\psi_{it}) + \exp\left(b_{jt} - \alpha p^*_{jt}\right) + \exp(ss_t)\right]^2} dF(\psi_{it})} & \text{if } ms_{jt} \neq \bar{ms}_{jt} \\ \int \frac{\exp\left(b_{jt} - \alpha p^*_{jt}\right)}{\exp(-\psi_{it}) + \exp\left(b_{jt} - \alpha p^*_{jt}\right) + \exp(ss_t)} dF(\psi_{it}) = \frac{T_t}{M} & \text{if } ms_{jt} = \bar{ms}_{jt} \end{cases}$$

We make an approximation similar to Doraszelski, Lewis, and Pakes (2018) in which we simply set the mean value of ψ_{it} instead of integrating over ψ_{it}. Thus, the optimal price obtained by substituting $E\left[\psi_{it}\right]$ for ψ in Equation (7).

$$p^*_{jt} = \begin{cases} c_j + \frac{1}{\alpha} + \frac{1}{\alpha} W\left(\frac{\exp(b_{jt} - \alpha c_j - 1)}{\exp(-E[\psi_{it}]) + \exp(ss_t)} \right) & \text{if } ms_{jt} \neq \bar{ms}_{jt} \\ \frac{1}{\alpha} \left[b_{jt} - \log\left(\frac{T_t}{M - T_t} \right) - \log\left(\exp\left(-E\left[\psi_{it}\right]\right) + \exp\left(ss_t\right) \right) \right] & \text{if } ms_{jt} = \bar{ms}_{jt} \end{cases}$$

4.5 Learning about Seasonality Shock

If hosts chose either custom or algorithm as their pricing method, they observe the prices suggested by the algorithm. The algorithmic prices serve as a signal about the seasonality shock, because the seasonality shock is the only missing piece of information about the guest demand.

Recall that optimal price can be written as a function of the seasonality shock given the demand estimates $\theta_{D1} = \{\alpha, \beta, \tau_t, \xi_{jt}\}$ and marginal cost c_j:

$$p^*_{jt} = f\left(\psi_{it}; \theta_D, c_j\right)$$

Substituting Equation (3) into the above equation,

$$p^{\text{Atip}}_{jt} - p^{o*}_{jt}(\Delta_{jt} + \nu_{jt}) = f\left(\psi_{it}; \theta_D, c_j\right)$$

Substituting $\Delta_{jt} = \bar{\Delta}_{jt} + \eta_{jt}$ where $\eta_{jt} \sim N(0, \sigma^2_{\Delta_{jt}})$,

$$p^{Atip}_{jt} - p^{o*}_{jt}\bar{\Delta}_{jt} - p^{o*}_{jt}(\eta_{jt} + \nu_{jt}) = f\left(\psi_{it}; \theta_D, c_j\right)$$

In other words, we can express ψ_{it} as a function of the bias-adjusted algorithmic price, $(p^{\text{Atip}}_{jt} - p^{o*}_{jt}\bar{\Delta}_{jt})$, and the error term, $(\eta_{jt} + \nu_{jt})$.

$$\psi_{it} = f^{-1}\left(p^{Atip}_{jt} - p^{o*}_{jt}\bar{\Delta}_{jt} - p^{o*}_{jt}(\eta_{jt} + \nu_{jt}); \theta_D, c_j\right) \tag{9}$$

Using the Equation (7), we obtain

$$\psi_{it} + \sigma_{\Lambda_{jt}}\varepsilon_{\Lambda_{jt}} = \Lambda_{jt} \tag{10}$$

where

$$\Lambda_{jt} = \lambda_{1,jt}\left[\log\left(\lambda_{1,jt}\lambda_{2,jt}\right) - \frac{\alpha p^{o*}_{jt}}{\lambda_{2,jt}} - b_{jt} + \alpha\left(\frac{1}{\lambda_{2,jt}} + 1\right)\left(p^{\text{Atip}}_{jt} - p^{o*}_{jt}\bar{\Delta}_{jt}\right)\right]$$

$$\sigma_{\Lambda_{jt}} = \alpha p^{o*}_{jt}\lambda_1\left(\frac{1}{\lambda_2} + 1\right)\sqrt{\sigma^2_{\Delta_{jt}} + \sigma^2_{\nu}}$$

$$\varepsilon_{\Lambda_{jt}} \sim N(0,1)$$

with $\lambda_1 = 1 + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)$ and $\lambda_2 = \alpha p^{o*}_{jt} - \alpha c_j - 1$. The derivation is provided in Appendix A.

Given the prior distribution $\psi^{prior}_{it} \sim N(0, \sigma^2_{\psi_t})$ and the signal in Equation (10), the hosts' posterior beliefs will be distributed as $\psi^{post}_{it} \sim N(\tilde{\psi}_t, \tilde{\sigma}^2_{\psi_t})$, where

$$\tilde{\psi}_t = \frac{\Lambda_{jt}/\sigma^2_{\Lambda_{jt}}}{1/\sigma^2_{\psi_t} + 1/\sigma^2_{\Lambda_{jt}}}$$

$$\frac{1}{\tilde{\sigma}^2_{\psi_t}} = \frac{1}{\sigma^2_{\psi_t}} + \frac{1}{\sigma^2_{\Lambda_{jt}}}$$

4.6 Learning about Bias of the Algorithm

After the demand is realized and the true seasonality shock ψ^T_{it} is realized, hosts can compute the optimal price under the true seasonality shock. The optimal price is a function of the true seasonality shock ψ^T_{it}, demand estimates $\theta_D = \{\alpha, \beta, \tau_t, \xi_{jt}\}$, and marginal cost c_j:

$$p^*_{jt} = f\left(\psi^T_{it}, \theta_D, c_j\right)$$

Substituting Equation (3) into the above equation,

$$p^{\text{Atip}}_{jt} - p^{o*}_{jt}(\Delta_{jt} + \nu_{jt}) = f\left(\psi^T_{it}, \theta_D, c_j\right)$$

In other words, we can express Δ_{jt} as a function of the difference between the optimal price given the realized seasonality shock and the price suggested by the algorithm, $(p^{\text{Atip}}_{jt} - f\left(\psi^T_{it}, \theta_D, c_j\right))$, and the error term, ν_{jt}.

$$\Delta_{jt} + \nu_{jt} = \frac{1}{p^{o*}_{jt}}\left[p^{\text{Atip}}_{jt} - f\left(\psi^T_{it}, \theta_D, c_j\right)\right] \tag{11}$$

Using the Equation (7), we obtain

$$\Delta_{jt} + \nu_{jt} = \kappa_{jt} \tag{12}$$

where

$$\kappa_{jt} = \frac{1}{\alpha}\frac{1}{p^{o*}_{jt}}\frac{\lambda_2}{1+\lambda_2}\left[\log\left(\lambda_1\lambda_2\right) - \frac{\psi^T_{it}}{\lambda_1} - \frac{\alpha p^{o*}}{\lambda_2} - b + \alpha\left(1 + \frac{1}{\lambda_2}\right)p^{\text{Atip}}_{jt}\right]$$

$$\nu_{jt} \sim N(0, \sigma^2_\nu)$$

with $\lambda_1 = 1 + \sum_{j' \neq j}\exp\left(b_{j't} - \alpha p_{j't}\right)$ and $\lambda_2 = \alpha p^{o*}_{jt} - \alpha c_j - 1$. The derivation is provided in Appendix A.

Given the prior distribution $\Delta_{jt} \sim N(\bar{\Delta}_{jt}, \sigma^2_{\Delta_{jt}})$ and the signal in Equation (12), the hosts' posterior beliefs will be distributed as $\Delta_{j,t+1} \sim N(\bar{\Delta}_{j,t+1}, \sigma^2_{\Delta_{j,t+1}})$, where

$$\bar{\Delta}_{j,t+1} = \frac{\bar{\Delta}_{jt}/\sigma^2_{\Delta_{jt}} + \kappa_{jt} \cdot 1(d_{jt} \neq B)/\sigma^2_{\nu}}{1/\sigma^2_{\Delta_{jt}} + 1(d_{jt} \neq B)/\sigma^2_{\nu}} = \frac{\bar{\Delta}_0/\sigma^2_{\Delta_0} + \sum_{s=1}^{t} \left[\kappa_{js} \cdot 1(d_{js} \neq B)\right]/\sigma^2_{\nu}}{1/\sigma^2_{\Delta_0} + \sum_{s=1}^{t} 1(d_{js} \neq B)/\sigma^2_{\nu}}$$

$$\frac{1}{\sigma^2_{\Delta_{j,t+1}}} = \frac{1}{\sigma^2_{\Delta_{jt}}} + \frac{1(d_{jt} \neq B)}{\sigma^2_{\nu}} = \frac{1}{\sigma^2_{\Delta_0}} + \frac{\sum_{s=1}^{t} 1(d_{js} \neq B)}{\sigma^2_{\nu}}$$

To complete the specification of evolution of the belief, we represent hosts' prior beliefs at the beginning of the period by $\Delta_{j0} \sim N(\bar{\Delta}_0, \sigma^2_{\Delta_0})$.

5 Estimation and Identification

The demand side parameters from the property market share model include price coefficient α, coefficients for property and host characteristics β, seasonality fixed effects τ_t, variance of seasonality shock $\sigma^2_{\psi_t}$, and property- and time-specific unobserved characteristics ξ_{jt}. The supply side parameters from the model of hosts' pricing method choice and pricing include marginal cost c_j, cost of setting custom price c^{menu}, cost of turning the Smart Pricing algorithm on or off $c^{\text{on/off}}$, hosts' prior beliefs at the beginning of the period $\left\{\bar{\Delta}_0, \sigma^2_{\Delta_0}\right\}$, and variance of the signal noise of the algorithm suggested prices σ^2_{ν}. We first estimate the demand side and then estimate the supply side given the demand estimates.

5.1 Demand side

We estimate the demand side parameters using the algorithm proposed by BLP. Recall that the indirect utility function for guests is:

$$u_{ijt} = x_{jt}\beta - \alpha p_{jt} + \xi_{jt} + \tau_{it} + \varepsilon_{ijt}$$

We split the indirect utility into two parts: mean utility across guests, $\delta_{jt} = x_{jt}\beta - \alpha p_{jt} + \tau_t + \xi_{jt}$, and individual deviations from the mean, $\mu_{ijt} = \psi_{it}$. Denote the parameters related to the mean utility as $\theta_{D1} = \{\alpha, \beta, \tau_t, \xi_{jt}\}$ and the parameters related to the individual deviations as $\theta_{D2} = \{\sigma^2_{\psi_t}\}$. The algorithm is composed of the following five steps.

1. Draw individual draws from a standard normal distribution for a set of guests. Compute initial mean utility based on homogeneous logit: $\delta^{(0)}_{jt} = \log(ms_{jt}) - \log(ms_{0t})$.

2. For given θ_{D2}, compute the individual deviations from mean utility. Using the computed individual deviations and the given mean utility, compute the predicted shares ps_{jt}.

3. Given nonlinear parameters θ_{D2}, search for δ_{jt} such that the observed shares ms_{jt} are equivalent to the predicted shares ps_{jt} using the following contraction mapping.

$$\delta_{jt}^{(h+1)} = \delta_{jt}^{(h)} + \log(ms_{jt}) - \log(ms_{0t})$$

4. From δ_{jt}, estimate the linear parameters θ_{D1} using an analytical formula. Form the GMM objective function $Q(\theta_{D2})$ using a set of moments enforcing that demand shocks ξ_{jt} are conditional mean zero, given instruments z_{jt}: $E[\xi_{jt}|z_{jt}] = 0$. We use the own product characteristics and the mean of competitor product characteristics as instruments.

$$\theta_{D1} = (X'ZWZ'X)^{-1}(X'ZWZ'\delta)$$

$$Q(\theta_{D2}) = \xi(\theta_{D2})'ZWZ'\xi(\theta_{D2})$$

5. Minimize $Q(\theta_{D2})$ over θ_2 with Steps 2-4 nested for every θ_{D2} trial.

5.2 Supply Side

Given the estimated demand system, we estimate the supply side in three steps. In the first step, we estimate the parameters of the marginal cost function γ assuming the absence of the algorithm ($\psi_{it} = 0$).[37] Specifically, we calculate the implied marginal cost c_j using the optimality condition in Equation 5, which we impose on the observations that choose custom as their pricing method. To estimate marginal cost for the observations that choose base or algorithm as their pricing method, we assume a linear marginal cost function. For an individual host j with a vector of characteristics w_j, the marginal cost is given by:

$$\log(c_j) = \gamma w_j + \omega_j$$

where ω_j is the unobserved idiosyncratic cost associated with host j. w_j may include the same characteristics that affect demand x_{jt} or different characteristics. $\gamma$is identified from the inter-temporal and cross sectional variation in demand and the impact of this variation on the prices set by custom pricing method, p^*_{jt}.

In the second step, we estimate the cost of setting custom price c^{menu} using the observations before the launch of the algorithm. Hosts make static decisions between base and custom pricing methods by comparing

[37] Note that we abstract away from learning about seasonality shock when hosts set their own custom price. γ should be estimated assuming $\psi_{it} = 0$ for pre-algorithm period and $\psi_{it} = \tilde{\psi}_t$ for post-algorithm period. This issue will be addressed in the next version of the paper.

the per period payoffs in Equation 8. The likelihood function is:

$$L = \prod_{jt} \left(\frac{\exp\left(U_{jt}^{B}\right)}{\exp\left(U_{jt}^{B}\right) + \exp\left(U_{jt}^{C}\right)} \right)^{1(d_{jt}=B)} \left(\frac{\exp\left(U_{jt}^{B}\right)}{\exp\left(U_{jt}^{B}\right) + \exp\left(U_{jt}^{C}\right)} \right)^{1(d_{jt}=C)}$$

c^{menu} is identified from the pricing method choice (base vs. custom) before the algorithm was launched. Specifically, c^{menu} is identified from the inter-temporal variation in $E\left[\Pi^{C} - \Pi^{B}\right]$ and the impact of this variation on the choice of pricing method.

In the third step, we estimate the algorithm-related parameters – the cost of turning the Smart Pricing algorithm on or off $c^{\text{on/off}}$, hosts' prior beliefs at the beginning of the period $\left\{\bar{\Delta}_0, \sigma^2_{\Delta_0}\right\}$, and variance of the signal noise of the algorithm suggested prices σ^2_{ν} – using the observations after the launch of the algorithm. Hosts make dynamic decisions among base, custom, and algorithmic pricing methods by comparing the choice-specific value functions in Section 4.3.4.

$$L = \prod_{jt} \prod_{d \in \{B,C,A\}} \left(\frac{\exp\left(v_{jt}^{d}\right)}{\exp\left(v_{jt}^{B}\right) + \exp\left(v_{jt}^{C}\right) + \exp\left(v_{jt}^{A}\right)} \right)^{1(d_{jt}=d)}$$

σ^2_{ν} is identified by the impact of the number of past usage of custom/algorithm pricing method on the host's current propensity to choose the algorithm as the pricing method. The greater the amount of learning (as captured by the inverse of the noise in bias signal, σ_{ν}), the greater the impact of number of past usage of custom/algorithm pricing method on the current propensity to choose the algorithm as the pricing method. Given σ^2_{ν}, the proportion of hosts choosing algorithm as their pricing method depends on $\bar{\Delta}_0$ and $\sigma^2_{\Delta 0}$. In addition, given σ^2_{ν}, the prices set by custom pricing method, p^{*}_{jt}, also depends on $\bar{\Delta}_0$ and $\sigma^2_{\Delta 0}$. Hosts' choice of algorithm as their pricing method and the prices set by custom pricing method jointly identify $\bar{\Delta}_0$ and $\sigma^2_{\Delta 0}$. Lastly, given $\bar{\Delta}_0$ and $\sigma^2_{\Delta 0}$, $c^{\text{on/off}}$ is identified by the variation in $E\left[\Pi^{A}(\bar{\Delta}_0, \sigma^2_{\Delta 0}) - \Pi^{B}\right]$ and and its impact on the pricing method choice.

6 Estimation Results

6.1 Parameter Estimates

Demand side. Table 6 and Figure 4 present the estimation results of the property market share model. Most coefficients have the expected signs. For example, the price coefficient $\alpha$is negative. Listings managed by superhost and listings with more reviews and photos are more likely to be booked. The seasonality fixed effect is smaller in 2015 compared to 2016 and 2017, which is consistent with the exponential growth of

Airbnb in this period. In addition, the size of the seasonality shock is larger in 2015 compared to 2016 and 2017, which may be attributed to hosts with more experience in Airbnb hosting, especially commercially operating hosts who rent out multiple units.[38] Lastly, the estimates capture the monthly seasonality pattern in which lodging demand spikes during SXSW in March and during Austin City Limits Music Festival in October.

Supply side. Table 7 presents the estimation results of the supply side model. First, Figure 5 shows the histogram of marginal cost estimates on the left and the histogram of price-cost margins on the right. The estimates γ suggest that the median marginal cost of hosting is \$49.9 per day, with a 25 percentile of \$31.9 and a 75 percentile of \$104.4. The estimates suggest that additional bedrooms and bathrooms increase the marginal cost of hosting. The daily cost for an entire place listing is 36.5% ($= e^{0.311} - 1$) larger than that of a private or shared room listing. As hosts gain more experience hosting, as indicated by their superhost status or the number of listings managed, their marginal cost of hosting decreases. The median price-cost margin is 64.0%, with a 25 percentile of 52.1% and a 75 percentile of 74.2%. Note that these estimates are comparable to the price-cost margin reported by hosts. For example, a host reports her revenue and cost during her 1-year experience as a host and reports 72.3% price-cost margin.[39] The cost of setting custom price is \$2.23 per day, and the cost of turning Smart Pricing on or off is \$84.10, which explains why only 2.04% of the observations change the default pricing method. At the launch of the Smart Pricing algorithm, hosts believe the algorithm will suggests prices that are close to optimal prices — just 2.5% higher. Lastly, the ratio of the signal noise of the algorithmic suggested price to the initial uncertainty is 15.97%, which indicates that the suggested prices are sufficiently precise to serve as a signal.

6.2 Hosts' Profit from Each Pricing Method

Given the demand side estimates and the marginal cost estimates, we compute hosts' profit using the base, custom, and algorithmic prices. We first examine how much more hosts can earn by responding to seasonality and setting optimal prices every time period. Figure 6(a) shows the histogram of percentage gains of hosts' profit by changing the pricing method from base to custom. Changing the pricing method from base to custom increases the hosts' profit by 33.4% on average, with a 25 percentile of 6.8% and a 75 percentile of 52.9%. In dollar terms, average increase is \$8.78 per day and median increase is \$1.80 per day. Note that these numbers are comparable to the amount of revenue increase of which third-party pricing services promise. For example, Beyond Pricing promises 10-40% increase in revenue.[40]

Next, we examine the size of the algorithmic bias. Figure 6 shows the histogram of percentage gains of

[38] See https://skift.com/2017/03/10/airbnbs-growth-is-being-driven-by-commercial-operators-report-says/.
[39] See https://affordanything.com/10-lessons-ive-learned-as-an-airbnb-host/.
[40] See https://techcrunch.com/2015/03/26/best-airbnb-price/.

Table 6: Demand Side Parameter Estimates

		Est.	Std.
α	price coeff.	-0.0091	
β	bedrooms	1.2641	
	bathrooms	0.3625	
	entire place	0.2014	
	type: house	0.3245	
	type: apt	0.5146	
	# reviews	0.0159	
	# photos	0.0458	
	superhost	0.2043	
	respond always	2.7008	
	respond fast	-0.0761	
τ_t	year 2015	-0.9204	
	year 2016	0.6940	
	year 2017	0.4227	
σ_{ψ_t}	year 2015	2.0163	
	year 2016	0.0811	
	year 2017	0.0904	

Figure 4: Demand Side Parameter Estimates

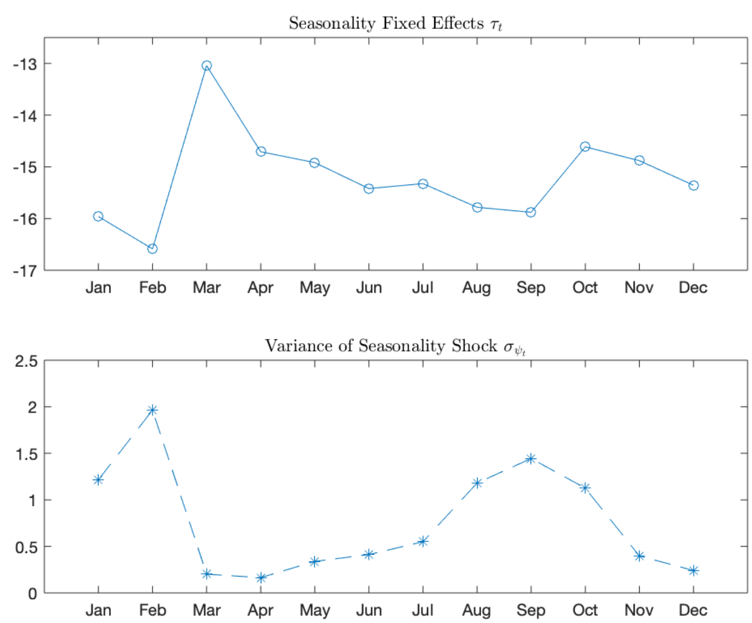
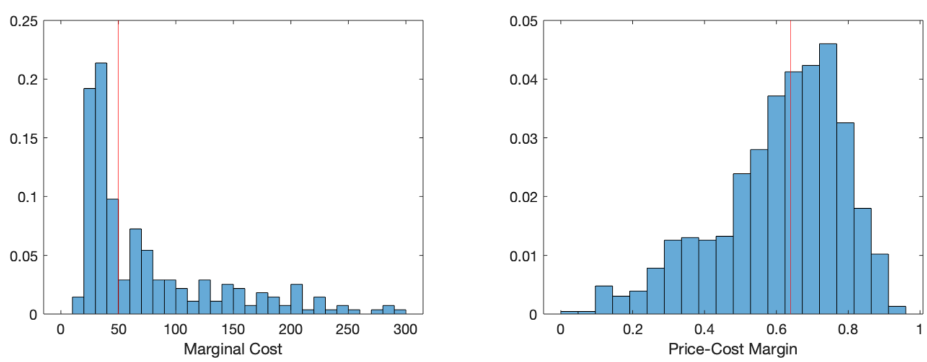


Table 7: Supply Side Parameter Estimates

		Est.	Std.
γ	constant	2.716	
	bedrooms	0.382	
	bathrooms	0.355	
	entire place	0.311	
	type: house	0.171	
	type: apt	-0.213	
	superhost	-0.313	
	respond always	0.285	
	respond fast	-0.379	
	# of listings by host	-0.092	
	c^{menu}	2.230	
	$c^{\text{on/off}}$	84.10	
	$\bar{\Delta}_0$	0.025	
	$\sigma^2_{\Delta_0}$	10.00	— (fixed)
	σ^2_{ν}	1.597	

Note: The algorithm-related parameters $\{c^{\text{on/off}}, \bar{\Delta}_0, \sigma^2_{\nu}\}$ are estimated using a subset of observations due to time constraints. The final estimates will follow in the next version of the paper.

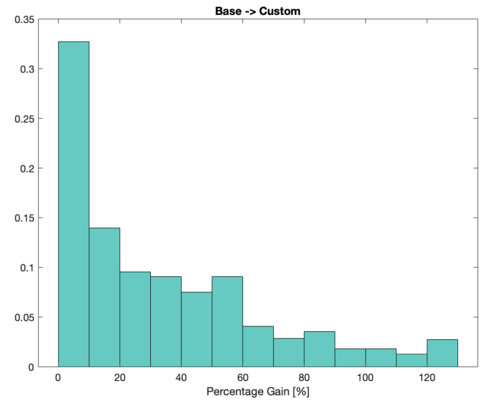
Figure 5: Histogram of Estimated Marginal Costs and Price-Cost Margins



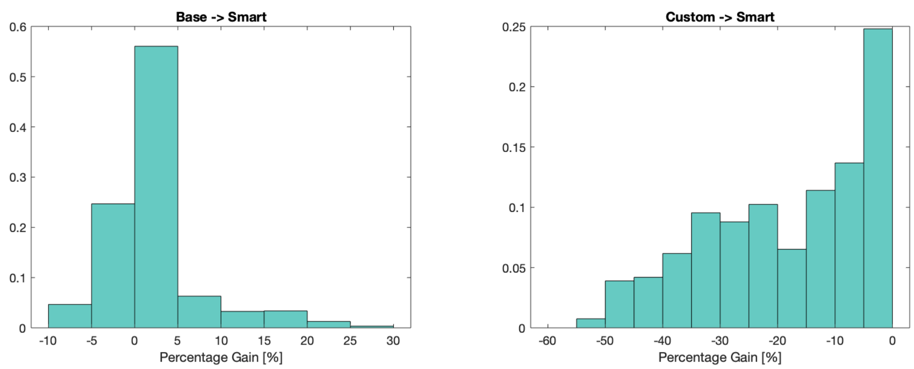
Note: Red lines represent the median values.

Figure 6: Histogram of Percentage Gains from Change of Pricing Methods

(a) Base –> Custom



(b) Base –> Smart & Custom –> Smart

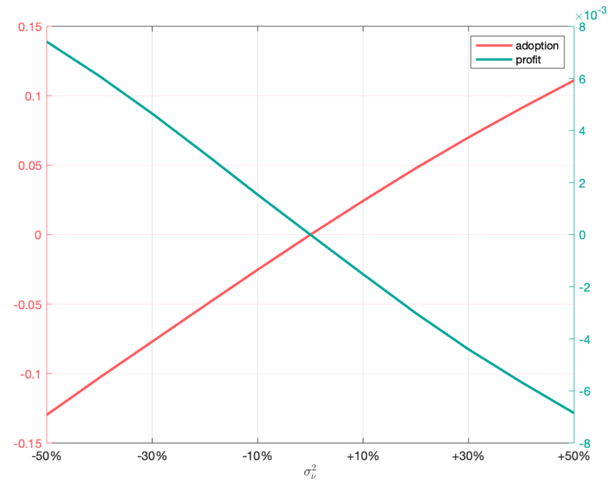


hosts' profit by changing the pricing method from base to smart (left panel) and from custom to smart (right panel). The results indicate the presence of algorithmic bias in the Smart Pricing algorithm. Changing the pricing method from base to custom does not always increases profit; 42.5% of host-month observations are actually worse off from using algorithmic price instead of base price. Compared to using custom price (i.e., optimal price), using the algorithmic price decreases the profit by 18.3% on average.

7 Counterfactuals

In the counterfactual analysis, we vary two parameters related to hosts' belief about algorithmic bias and examine how they change the algorithm adoption, and thus, hosts' profit. First, we analyze hosts benefit from having more correct beliefs. Hosts currently form belief about algorithmic bias by learning about bias from the difference between the algorithmic suggested prices and the optimal prices under realized demand. In the counterfactual, we simulate the case in which hosts have relatively more correct belief; their belief

Figure 7: Percentage Change in Hosts' Adoption and Profit by Varying σ_ν^2



is the actual ratio of the suggested prices to the optimal prices. In other words, hosts have a correct belief for this time period, though it may not be correct for the future periods. Under the relatively more correct belief, the adoption decreases by 1.66% and hosts' profit increases by 0.49% on average. Second, we vary the variance of signal noise, σ_ν^2, from 50% to 150% of the current value. Figure 7 presents the percentage change in hosts' adoption and profit by varying the variance of signal noise. We find that more precise signals lead to less adoption and more profit. In other words, hosts are better off with more precise signal of the price suggestions.

8 Conclusion

The goal of this paper is to investigate societal impact of Airbnb's Smart Pricing algorithm. To achieve the goal, we estimate a structural model of hosts' pricing and algorithm adoption using the Airbnb listings data. The results show that using the algorithm often reduces hosts' profit due to algorithmic bias. Providing more correct beliefs about algorithmic bias and more precise signals from the price suggestions reduces the adoption and increases hosts' profit.

References

[1] Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pp. 841–890.

[2] Bellman, R. (1957). Dynamic Programming. Princeton University Press.

[3] Datta, A., Tschantz, M. C., and Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1), pp. 92-112.

[4] Doraszelski, U., Lewis, G., and Pakes, A. (2018). Just Starting Out: Learning and Equilibrium in a New Market. *American Economic Review*, Vol. 108, No. 3, pp. 565-615.

[5] Einav, L., Farronato, C., and Levin, J. (2016). Peer-to-Peer Markets. *Annual Review of Economics*, 8(1), 615-635.

[6] Ericson, R. and Pakes, A. (1995). Markov-Perfect Industry Dynamics: A Framework for Empirical Work. *The Review of Economic Studies*, 62(1), pp. 53-82.

[7] Gandhi, A., Lu, Z., and Shi, X. (2019). Estimating Demand for Differentiated Products with Zeroes in Market Share Data. Working paper.

[8] Huang, Y., Ellickson, P. B., and Lovett, M. J. (2019). Learning to Set Prices. Working paper.

[9] Huang, G., Luo, H., and Xia, J. (2019). Invest in Information or Wing It? A Model of Dynamic Pricing with Seller Learning. *Management Science*, 65(12), pp. 5556-5583.

[10] Jeon, J. (2019). Learning and Investment under Demand Uncertainty in Container Shipping. Working paper.

[11] Lambrecht, A. and Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent GenderBased Discrimination in the Display of STEM Career Ads. *Management Science*, 65(7), pp. 2966-2981.

[12] Li, J., Moreno, A., and Zhang, D. (2016). Pros vs Joes: Agent Pricing Behavior in the Sharing Economy (August 28, 2016). Ross School of Business Paper No. 1298. Available at SSRN: https://ssrn.com/abstract=2708279.

[13] Olimov, F. (2019). Dynamic Pricing in a Decentralized Platform: The Case of Airbnb. Working paper.

[14] Pan, Q. (2019). Costly Price Adjustment and Automated Pricing: The Case of Airbnb. Working paper.

[15] Pavlov, V., and Berman, R. (2019). Price Manipulation in Peer-to-Peer Markets. Working paper.

[16] Sweeney, L. (2013). Discrimination in online ad delivery. *ACM Queue*, 11(3), 10:10-10:29.

[17] Weintraub, G., Benkard, C., and Van Roy, B. (2008). Markov Perfect Industry Dynamics with Many Firms. *Econometrica*, 76(6):1375–1411.

[18] Weintraub, G., Bendard, C., and Van Roy, B. (2010). Computational Methods for Oblivious Equilibrium. *Operations Research*, 58(4), pp. 1247–1265.

[19] Ye, P., Qian, J., Chen, J., Wu, C., Zhou, Y., De Mars, S., Yang, F., and Zhang, L. (2018). Customized Regression Model for Airbnb Dynamic Pricing. KDD 2018, August 19-23, 2018, London, United Kingdom.

Appendix

A. Derivations

A.1. Optimal Price

(i) $ms_{jt} \neq \bar{ms}_{jt}$

$$\Pi_{jt} = (p_{jt} - c_j) \cdot \frac{\exp(b_{jt} - \alpha p_{jt})}{\exp(-\psi_{it}) + \exp(b_{jt} - \alpha p_{jt}) + \sum_{j' \neq j} \exp(b_{j't} - \alpha p_{j't})} \cdot M$$

Taking the first order condition $\frac{\partial \Pi_{jt}}{\partial p_{jt}} = 0$,

$$1 - \alpha(p^*_{jt} - c_j) \frac{\exp(-\psi_{it}) + \sum_{j' \neq j} \exp(b_{j't} - \alpha p_{j't})}{\exp(-\psi_{it}) + \exp\left(b_{jt} - \alpha p^*_{jt}\right) + \sum_{j' \neq j} \exp(b_{j't} - \alpha p_{j't})} = 0$$

$$\exp\left(b_{jt} - \alpha p^*_{jt}\right) = (\alpha p^*_{jt} - \alpha c_j - 1)\left(\exp(-\psi_{it}) + \sum_{j' \neq j} \exp(b_{j't} - \alpha p_{j't})\right) \tag{13}$$

$$(\alpha p^*_{jt} - b_{jt} + b_{jt} - \alpha c_j - 1)\left(\exp(-\psi_{it}) + \sum_{j' \neq j} \exp(b_{j't} - \alpha p_{j't})\right) \exp\left(\alpha p^*_{jt} - b_{jt}\right) = 1$$

Denote $m \equiv b_{jt} - \alpha c_j - 1$, $n \equiv \exp(-\psi_{it}) + \sum_{j' \neq j} \exp(b_{j't} - \alpha p_{j't})$, and $x \equiv \alpha p^*_{jt} - b_{jt}$,

$$(x + m) \cdot n \cdot e^x = 1$$

Denote $u = x + m$,

$$ue^u = \frac{e^m}{n}$$

By the definition of the Lambert W function,

$$u = W\left(\frac{e^m}{n}\right)$$

$$x + m = W\left(\frac{e^m}{n}\right)$$

$$\alpha p^*_{jt} - b_{jt} + b_{jt} - \alpha c_j - 1 = W\left(\frac{\exp(b_{jt} - \alpha c_j - 1)}{\exp(-\psi_{it}) + \sum_{j' \neq j} \exp(b_{j't} - \alpha p_{j't})}\right)$$

$$p_{jt}^* = c_j + \frac{1}{\alpha} + \frac{1}{\alpha} W\left(\frac{\exp\left(b_{jt} - \alpha c_j - 1\right)}{\exp\left(-\psi_{it}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)}\right)$$

(ii) $ms_{jt} = \bar{m}s_{jt}$

$$\left(p_{jt}^* - c_j\right) \cdot \frac{\exp\left(b_{jt} - \alpha p_{jt}^*\right)}{\exp\left(-\psi_{it}\right) + \exp\left(b_{jt} - \alpha p_{jt}^*\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)} = \left(p_{jt}^* - c_j\right) \cdot \frac{T_t}{M}$$

$$1 - \frac{\exp\left(-\psi_{it}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)}{\exp\left(-\psi_{it}\right) + \exp\left(b_{jt} - \alpha p_{jt}^*\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)} = \frac{T_t}{M}$$

$$\exp\left(b_{jt} - \alpha p_{jt}^*\right) = \frac{T_t}{M - T_t} \cdot \left[\exp\left(-\psi_{it}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)\right]$$

$$b_{jt} - \alpha p_{jt}^* = \log\left(\frac{T_t}{M - T_t}\right) + \log\left(\exp\left(-\psi_{it}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)\right)$$

$$p_{jt}^* = \frac{1}{\alpha}\left[b_{jt} - \log\left(\frac{T_t}{M - T_t}\right) - \log\left(\exp\left(-\psi_{it}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)\right)\right]$$

From (i) and (ii),

$$p_{jt}^* = \begin{cases} c_j + \frac{1}{\alpha} + \frac{1}{\alpha} W\left(\frac{\exp(b_{jt} - \alpha c_j - 1)}{\exp(-\psi_{it}) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)}\right) & \text{if } ms_{jt} \neq \bar{m}s_{jt} \\ \frac{1}{\alpha}\left[b_{jt} - \log\left(\frac{T_t}{M - T_t}\right) - \log\left(\exp\left(-\psi_{it}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)\right)\right] & \text{if } ms_{jt} = \bar{m}s_{jt} \end{cases}$$

A.2. Learning about Seasonality Shock

The optimal price satisfies Equation (13).

$$\exp\left(b_{jt} - \alpha p_{jt}^*\right) = \left(\alpha p_{jt}^* - \alpha c_j - 1\right)\left(\exp\left(-\psi_{it}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)\right)$$

$$b_{jt} - \alpha p_{jt}^* = \log\left(\alpha p_{jt}^* - \alpha c_j - 1\right) + \log\left(\exp\left(-\psi_{it}\right) + \sum_{j' \neq j} \exp\left(b_{j't} - \alpha p_{j't}\right)\right) \tag{14}$$

Taylor expansion of $\log\left(\alpha p_{jt}^* - \alpha c_j - 1\right)$ at p^{o*} is:

$$\log\left(\alpha p_{jt}^* - \alpha c_j - 1\right) = \log\left(\alpha p_{jt}^{o*} - \alpha c_j - 1\right) + \frac{\alpha(p^* - p^{o*})}{\alpha p_{jt}^{o*} - \alpha c_j - 1} \tag{15}$$

Taylor expansion of $\log\left(\exp\left(-\psi_{it}\right)+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)\right)$ at $\psi_{it}=0$ is:

$$\log\left(\exp\left(-\psi_{it}\right)+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)\right)=\log\left(1+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)\right)-\frac{\psi_{it}}{1+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)} \tag{16}$$

Substituting Equations (15) and (16) into Equation (14),

$$\begin{aligned}b_{jt}-\alpha p^*_{jt}&=\log\left(\alpha p^{o*}_{jt}-\alpha c_j-1\right)+\frac{\alpha(p^*-p^{o*})}{\alpha p^{o*}_{jt}-\alpha c_j-1}\\&\quad+\log\left(1+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)\right)-\frac{\psi_{it}}{1+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)}\end{aligned}$$

Denote $\lambda_1=1+\sum_{j'\neq j}\exp\left(b_{j't}-\alpha p_{j't}\right)$ and $\lambda_2=\alpha p^{o*}_{jt}-\alpha c_j-1$,

$$b_{jt}-\alpha p^*_{jt}=\log\left(\lambda_1\lambda_2\right)+\frac{\alpha(p^*-p^{o*})}{\lambda_2}-\frac{\psi_{it}}{\lambda_1}$$

$$\psi_{it}=\lambda_1\left[\log\left(\lambda_1\lambda_2\right)-\frac{\alpha p^{o*}}{\lambda_2}-b_{jt}\right]+\alpha\lambda_1\left(\frac{1}{\lambda_2}+1\right)p^*_{jt} \tag{17}$$

Substituting $p^{Atip}_{jt}=p^*_{jt}+p^{o*}_{jt}(\Delta_{jt}+\nu_{jt})$ from Equation (3) and $\Delta_{jt}=\bar{\Delta}_{jt}+\eta_{jt}$ where $\eta_{jt}\sim N(0,\sigma^2_{\Delta_{jt}})$ into Equation (17),

$$\psi_{it}=\lambda_1\left[\log\left(\lambda_1\lambda_2\right)-\frac{\alpha p^{o*}}{\lambda_2}-b_{jt}\right]+\alpha\lambda_1\left(\frac{1}{\lambda_2}+1\right)\left(p^{Atip}_{jt}-p^{o*}_{jt}\bar{\Delta}_{jt}-p^{o*}_{jt}(\eta_{jt}+\nu_{jt})\right)$$

$$\psi_{it}+\alpha p^{o*}_{jt}\lambda_1\left(\frac{1}{\lambda_2}+1\right)(\eta_{jt}+\nu_{jt})=\lambda_1\left[\log\left(\lambda_1\lambda_2\right)-\frac{\alpha p^{o*}}{\lambda_2}-b_{jt}+\alpha\left(\frac{1}{\lambda_2}+1\right)\left(p^{Atip}_{jt}-p^{o*}_{jt}\bar{\Delta}_{jt}\right)\right]$$

A.3. Learning about Bias of the Algorithm

Substituting $p^{Atip}_{jt}=p^*_{jt}+p^{o*}_{jt}(\Delta_{jt}+\nu_{jt})$ into Equation (17),

$$\Delta_{jt}+\nu_{jt}=\frac{1}{\alpha}\frac{1}{p^{o*}_{jt}}\frac{\lambda_2}{1+\lambda_2}\left[\log\left(\lambda_1\lambda_2\right)-\frac{\psi_{it}}{\lambda_1}-\frac{\alpha p^{o*}}{\lambda_2}-b+\alpha\left(1+\frac{1}{\lambda_2}\right)p^{Atip}_{jt}\right]$$