The Sharing Economy and Housing Affordability: Evidence from Airbnb*

Kyle Barron† Edward Kung‡ Davide Proserpio§

April 1, 2018

Abstract

We assess the impact of home-sharing on residential house prices and rents. Using a dataset of Airbnb listings from the entire United States and an instrumental variables estimation strategy, we find that a 1% increase in Airbnb listings leads to a 0.018% increase in rents and a 0.026% increase in house prices at the median owner-occupancy rate zipcode. The effect is moderated by the share of owner-occupiers, a result consistent with absentee landlords reallocating their homes from the long-term rental market to the short-term rental market. A simple model rationalizes these findings.

Keywords: Sharing economy, peer-to-peer markets, housing markets, Airbnb
JEL Codes: R31, L86

---

*We thank Don Davis, Richard Green, Chun Kuang, Aske Egsgaard, Tom Chang, and seminar participants at the AREUA National Conference, the RSAI Annual Meetings, the Federal Reserve Bank of San Francisco, the UCI-UCLA-USC Urban Research Symposium, and the PSI Conference for helpful comments and suggestions. All errors are our own.

†NBER; barronk@nber.org.
‡Department of Economics, UCLA; ekung@econ.ucla.edu.
§Marshall School of Business, USC; proserpi@marshall.usc.edu.

1 Introduction

The sharing economy represents a set of peer-to-peer online marketplaces that facilitate matching between demanders and suppliers of various goods and services. The suppliers in these markets are often small (mostly individuals), and they often share excess capacity that might otherwise go unutilized—hence the term “sharing economy.” Economic theory would suggest that the sharing economy improves economic efficiency by reducing frictions that cause capacity to go underutilized, and the explosive growth of sharing platforms (such as Uber for ride-sharing and Airbnb for home-sharing) testifies to the underlying demand for such markets.[1] The growth of the sharing economy has also come at the cost of great disruption to traditional markets (Zervas et al., 2017), as well as new regulatory challenges, leading to contentious policy debates about how best to balance individual participants’ rights to freely transact, the efficiency gains from sharing economies, the disruption caused to traditional markets, and the role of the platforms themselves in the regulatory process.

Home-sharing, in particular, has been the subject of intense criticism. Namely, critics argue that home-sharing platforms like Airbnb raise the cost of living for local renters, while mainly benefitting local landlords and non-resident tourists.[2] It is easy to see the economic argument. By reducing frictions in the peer-to-peer market for short-term rentals, home-sharing platforms cause some landlords to switch from supplying the market for long-term rentals—in which residents are more likely to participate—to supplying the short-term market—in which non-residents are more likely to participate. Because the total supply of housing is fixed or inelastic in the short run, this drives up the rental rate in the long-term market. Concern over home-sharing’s

[1]These frictions could include search frictions in matching demanders with suppliers, and information frictions associated with the quality of the good being transacted, or with the trustworthiness of the buyer or seller. See Einav et al. (2016) for an overview of the economics of peer-to-peer markets, including the specific technological innovations that have facilitated their growth.

[2]Another criticism of Airbnb is that the company does not do enough to combat racial discrimination on its platform (see Edelman and Luca (2014); Edelman et al. (2017)), though we will not address this issue in this paper.

impact on housing affordability has garnered significant attention from policymakers, and has motivated many cities to impose stricter regulations on home-sharing.[3]

Whether or not home-sharing increases housing costs for local residents is an empirical question. There are a few reasons why it might not. The market for short-term rentals may be very small compared to the market for long-term rentals. In this case, even large changes to the short-term market might not have a measurable effect on the long-term market. The short-term market could be small—even if the short-term rental rate is high relative to the long-term rate—if landlords prefer more reliable long-term tenants and a more stable income stream. Alternatively, the market for short-term rentals could be dominated by housing units that would have remained vacant in the absence of home-sharing. Owner-occupiers, those who own the home in which they live, may supply the short-term rental market with spare rooms and cohabit with guests, or they may supply their entire home during temporary absences.[4] These otherwise vacant rentals could also be vacation homes that would not be rented to long-term tenants because of the restrictiveness of long-term leases. In either case, such owners would not make their homes available to long-term tenants, independently of the existence of a convenient home-sharing platform. Instead, home-sharing provides them with an income stream for times when their housing capacity would otherwise be underutilized.

In this paper, we study the effect of home-sharing on the long-term rental market using a comprehensive dataset of all US properties listed on Airbnb, the world's largest home-sharing platform. We first develop a simple model of house prices and rental rates when landlords can choose to allocate housing between long-term residents and short-term visitors. The effect of a home-

[3]For example, Santa Monica outlaws short-term, non-owner-occupied rentals of less than 30 days, as does New York State for apartments in buildings with three or more residences. San Francisco passed a 60-day annual hard cap on short-term rentals (which was subsequently vetoed by the mayor). It is unclear, however, the degree to which these regulations are enforced. We are aware of only one successful prosecution of an Airbnb host, occurring in Santa Monica in July 2016.

[4]A frequently cited example is that of the flight attendant who rents out his or her home on Airbnb while traveling for work.

sharing platform such as Airbnb is to reduce the frictions associated with renting on the short-term market. From the model we derive three testable predictions: 1) Airbnb increases both rental rates and house prices in the long-term market; 2) the increase in house prices is greater than the increase in rental rates, thus leading to an increase in the price-to-rent ratio; and 3) the effect on rental rates is smaller when a greater share of the landlords are owner-occupiers. Intuitively, the owner-occupancy rate matters because only non-owner-occupiers are on the margin of substituting their housing units between the long and short-term rental markets. Owner-occupiers interact with the short-term market only to rent out unused rooms or to rent while away on vacation, but they do not allocate their housing to long-term tenants.

To test the model, we collect primary data sources from Airbnb, Zillow, and the Census Bureau. We construct a panel dataset of Airbnb listings at the zipcode-year-month level from data collected from public-facing pages on the Airbnb website between the beginning of 2011 and the end of 2016, covering the entire United States. From Zillow, a website specializing in residential real estate transactions, we obtain a panel of house price and rental rate indices, also at the zipcode-year-month level. Zillow provides a platform for matching landlords with long-term tenants, and thus their price measures reflect sale prices and rental rates in the market for long-term housing. Finally, we supplement this data with a rich set of time-varying zipcode characteristics collected from the Census Bureau's American Community Survey (ACS), such as the median household income, population count, share of college graduates, and employment rate.

In the raw correlations, we find that the number of Airbnb listings in zipcode i in year-month t is positively associated with both house prices and rental rates. In a baseline OLS regression with no controls, we find that a 1% increase in Airbnb listings is associated with a 0.1% increase in rental rates and a 0.18% increase in house prices. Of course, these estimates should not be interpreted as causal, and may instead be picking up spurious correlations. For example, cities that are growing in population likely have rising rents, house prices, and numbers of Airbnb listings at the same time. We therefore exploit the panel nature of our dataset to control for unobserved zipcode level effects and arbitrary city level time trends. We include zipcode fixed effects to absorb any permanent differences between zipcodes, while fixed effects at the Core Based Statistical Area (CBSA)-year-month level control for any shocks to housing market conditions that are common across zipcodes within a CBSA.[5]

We further control for unobserved *zipcode-specific, time-varying* factors using an instrumental variable that is plausibly exogenous to local zipcode level shocks to the housing market. To construct the instrument, we exploit the fact that Airbnb is a young company that has experienced explosive growth over the past five years. Figure 1 shows worldwide Google search interest in Airbnb from 2008 to 2016. Demand fundamentals for short-term housing are unlikely to have changed so drastically from 2008 to 2016 as to fully explain the spike in interest, so most of the growth in Airbnb search interest is likely driven by information diffusion and technological improvements to Airbnb's platform as it matures as a company. Neither of these should be correlated with local zipcode level unobserved shocks to the housing market. By itself, global search interest is not enough for an instrument because we already control for arbitrary CBSA level time trends. We therefore interact the Google search index for Airbnb with a measure of how "touristy" a zipcode is in a base year, 2010. We define "touristy" to be a measure of a zipcode's attractiveness for tourists and proxy for it using the number of establishments in the food service and accommodations industry.[6] These include eating and drinking places, as well as hotels, bed and breakfasts, and other forms of short-term lodging. The identifying assumptions of our specification are that: 1) landlords in more touristy zipcodes are more likely to switch into the short-term rental market in response to learning about Airbnb than landlords in less touristy zipcodes; and 2) ex-ante levels of touristiness are not systematically correlated with ex-post unobserved shocks to the housing market at the zipcode level *that are*

[5]The CBSA is a geographic unit defined by the U.S. Office of Management and Budget that roughly corresponds to an urban center and the counties that commute to it.

[6]We focus on tourism because Airbnb has historically been frequented more by tourists than business travelers. Airbnb has said that 90% of its customers are vacationers, but is attempting to gain market share in the business travel sector.

also correlated in time with Google search interest for Airbnb. We discuss the instrument, its construction, and exercises supporting the exclusion restriction in more detail in Sections 4 and 4.1.

Using this instrumental variable, we estimate that for zipcodes with the median owner-occupancy rate (72%), a 1% increase in Airbnb listings leads to a 0.018% increase in the rental rate and a 0.026% increase in house prices. We also find that, as predicted by our theoretical model, the effect of Airbnb listings on rental rates and house prices is decreasing in the owner-occupancy rate. For zipcodes with a 56% owner-occupancy rate (the 25th percentile), the effect of a 1% increases in Airbnb listings is 0.024% for rents and 0.037% for house prices. For zipcodes with a 82% owner-occupancy rate (the 75th percentile), the effect of a 1% increase in Airbnb listings is only 0.014% for rents and 0.019% for house prices. These results are consistent with the model's predictions that the effect on both rental rate and house prices will be positive, that the effect on house prices will be larger than the effect on rents, and that the effect will be decreasing in owner-occupancy rate.

Next, we test the hypothesis that the effects we observe are partially due to absentee landlords substituting away from the rental and for-sale markets for long-term residents, and towards the short-term market. To do so, we consider the effect of Airbnb on housing vacancy rates. Because zipcode level data on vacancies are not available at a monthly—or even yearly—frequency, we focus on annual vacancy rates at the CBSA level. We find that annual CBSA vacancy rates have no association with the number of Airbnb listings. However, looking at the different types of vacancy we find that the number of Airbnb listings is positively associated with the share of homes that are vacant for seasonal or recreational use (likely to be part of the short-term rental market inventory) and negatively associated with the share of homes that are vacant-for-rent and vacant-for-sale (part of the long-term market inventory). These findings are consistent with absentee landlord switching from the long- to the short-term rental market.[7]

[7] Census Bureau methodology classifies a housing unit as vacant even if it is temporarily occupied by persons who usually live elsewhere.

Related literature

We are aware of only two other academic papers to directly study the effect of home-sharing on housing costs, and both of them focus on a specific US market. Lee (2016) provides a descriptive analysis of Airbnb in the Los Angeles housing market, while Horn and Merante (2017) use Airbnb listings data from Boston in 2015 and 2016 to study the effect of Airbnb on rental rates. They find that a one standard deviation increase in Airbnb listings at the census tract level leads to a 0.4% increase in asking rents. In our data, we find that a one standard deviation increase in listings at the within-CBSA zipcode level in 2015-2016 implies a 0.54% increase in rents.

We contribute to the literature concerning the effect of home-sharing on housing costs in three ways. First, we present a model that organizes our thinking about how home-sharing is expected to affect housing costs in the long-term market. Second, we provide direct evidence for the model's predictions, highlighting especially the role of the owner-occupancy rate and of the marginal landowner. Third, we present the first estimates of the effect of home-sharing on housing costs that uses comprehensive data from across the U.S.

Our paper also contributes to the growing literature on peer-to-peer markets. Such literature covers a wide array of topics, from the effect of the sharing economy on labor market outcomes (Chen et al., 2017; Hall and Krueger, 2017; Angrist et al., 2017) to entry and competition (Gong et al., 2017; Horton and Zeckhauser, 2016) to trust and reputation (Fradkin et al., 2017; Proserpio et al., 2017; Zervas et al., 2015). Because the literature on the topic is quite vast, we refer the reader to Einav et al. (2016) for an overview of the economics of peer-to-peer markets and to Proserpio and Tellis (2017) for a complete review of the literature on the sharing economy.

In terms of studies on Airbnb, both Zervas et al. (2017) and Farronato and Fradkin (2018) study the impact of Airbnb on the hotel industry. Zervas et al. (2017) focus on the effects on incumbents, while Farronato and Fradkin (2018) focus on the consumers gain in welfare. Our paper looks at a somewhat unique context in this literature because we focus on the effect of the sharing economy on the reallocation of goods from one purpose to another, which may cause local externalities. Local externalities are present here because the suppliers are local and the demanders are non-local; transactions in the home-sharing market, therefore, involve a reallocation of resources from locals to non-locals. Our contribution is therefore to study this unique type of sharing economy in which public policy may be especially salient.

The rest of the paper is organized as follows. In Section 2, we present a simple model of house prices and rental rates where landlords can substitute between supplying the long-term and the short-term market. In Section 3, we describe the data we collected from Airbnb and present some basic statistics. In Section 4, we describe our methodology and present exercises in support of the exclusion restriction of our instrument, and in Section 5 we discuss the results and present several robustness checks to reinforce the validity of our results. Section 6 discusses our findings, the limitations of our work, and provide concluding remarks.

2 Model

2.1 Basic setup

We consider a housing market with a fixed stock of housing H, which can be allocated to short-term housing S, or long-term housing L. $S + L = H$. The rental rate of short-term housing is Q and the rental rate of long-term housing is R. The two housing markets are segmented—tenants who need long-term housing cannot rent in the short-term market and tenants who need short-term housing cannot rent in the long-term market.[8]

For now, we assume that all housing is owned by absentee landlords and will return to the possibility of owner-occupiers later. Each landlord owns

[8] In our view, the primary driver of this market segmentation is the length of lease and tenant rights. Local residents participating in the long-term rental market will typically sign leases of 6 months to a year, and are also granted certain rights and protections by the city. On the other hand, non-resident visitors participating in the short-term market will usually only rent for a few days and are not granted the same rights as resident tenants.

one unit of housing and decides to rent it on the short-term market or the long-term market, taking rental rates as given. A landlord will rent on the short-term market if $Q - c - \epsilon > R$, where $c + \epsilon$ is an additional cost of renting on the short-term market, with c being a common component and ϵ being an idiosyncratic component across landlords.[9] The share of landlords renting in the short-term market is therefore:

$$f(Q - R - c) = P(\epsilon < Q - R - c) \tag{1}$$

f is the cumulative distribution function of ϵ, and $f' > 0$. The total number of housing units in the short-term market are:

$$S = f(Q - R - c)H \tag{2}$$

Long-term rental rates are determined in equilibrium by the inverse demand function of long-term tenants:

$$R = r(L) \tag{3}$$

with $r' < 0$. Short-term rental rates are determined exogenously by outside markets.[10] The market is in steady state, so the house price P is equal to the present value of discounted cash flows to the landlord:

$$\begin{aligned} P &= \sum_{t=0}^{\infty} \delta^t E\left[R + \max\left\{0, Q - R - c - \epsilon\right\}\right] \\ &= \frac{1}{1-\delta}\left[R + g(Q - R - c)\right] \end{aligned} \tag{4}$$

where $g(x) = E[x - \epsilon | \epsilon < x] f(x)$ gives the expected net surplus of being able

[9]Renting in the short-term market could be costlier than in the long-term market because the technology for matching landlords with tenants may be historically more developed in the long-term market. Landlords may have idiosyncratic preferences over renting in the long-term market vs. the short-term market if they have different preferences for the stability provided by long-term tenants.

[10]For example, they could be determined by elastic tourism demand. Relaxing this assumption and allowing for price elasticity in the short-term market would not change the qualitative results.

to rent in the short-term market relative to the long-term market, and $g' > 0$.

2.2 The effect of home-sharing

The introduction of a home-sharing platform reduces the cost for landlords to advertise on the short-term market, implying a decline in c. This could happen for a variety of reasons. By improving the search and matching technology in the short-term market, the sharing platform may reduce the time it takes to find short-term tenants. By providing identity verification and a reputation system for user feedback, the platform may also help reduce information costs.

We consider how an exogenous change to the cost of listing in the short-term market, c, affects long-term rental rates and house prices. Equilibrium conditions (1)-(3) imply that:

$$\frac{dR}{dc} = \frac{r'f'H}{1 - r'f'H} < 0 \tag{5}$$

So, by decreasing the cost of listing in the short-term market, the home-sharing platform has the effect of raising rental rates. The intuition is fairly straightforward: the home-sharing platform induces some landlords to switch from the long-term market to the short-term market, reducing supply in the long-term market and raising rental rates.

For house prices, we can use Equation (4) to write:

$$\frac{dP}{dc} = \frac{1}{1-\delta}\left[\frac{dR}{dc} - \left(1 + \frac{dR}{dc}\right)g'\right] \tag{6}$$

We note from Equation (5) that $-1 < \frac{dR}{dc} < 0$, and so $\frac{dP}{dc} < \frac{1}{1-\delta}\frac{dR}{dc} < \frac{dR}{dc} < 0$. The latter inequality concludes that home-sharing increases house prices and that the house price response will be greater than the rental rate response. This is because home-sharing increases the value of homeownership through two channels. First, it raises the rental rate which is then capitalized into house prices. Yet, if this were home-sharing's only effect, then the price response and the rental rate response would be proportional by the discount factor.

Instead, the additional increase in the value of homeownership comes from the enhanced option value of renting in the short-term market. Because of this second channel, prices will respond even more than rental rates to the introduction of a home-sharing platform.

2.3 Owner-occupiers

We now relax the assumption that all homeowners are absentee landlords by also allowing for owner-occupiers. Let H_a be the number of housing units owned by absentee landlords and let H_o be the number of housing units owned by owner-occupiers. We still define L as the number of housing units allocated to long-term residents—including owner-occupiers—and therefore the number of renters is $L - H_o$. We assume that H_a is fixed, and that H_o will be determined by equilibrium house prices and rental rates.[11]

We allow owner-occupiers to interact with the short-term housing market by assuming that a fraction γ of their housing unit is excess capacity. This excess capacity can be thought of as the unit's spare rooms or the time that the owner spends away from his or her home. Owner-occupiers have the choice to either hold their excess capacity vacant, or to rent it out on the short-term market. They cannot rent excess capacity on the long-term market, due to the nature of leases and renter protections. The benefit to renting excess capacity on the short-term market is $Q - c - \epsilon$, where c and ϵ are again the cost and the idiosyncratic preference for listing on the short-term market, respectively. If excess capacity remains unused, the owner neither pays a cost nor derives any benefit from the excess capacity. Owner-occupiers will rent on the short-term market if $Q - c - \epsilon > 0$, and thus $f(Q - c)$ is the share of owner-occupiers who rent their excess capacity on the short-term market.

Note that the choice of the owner-occupier is to either rent on the short-term market, or to hold excess capacity vacant. Thus, participation in the

[11] If H_a is not fixed, then all of the housing stock will be owned by either absentee landlords or owner-occupiers, depending on which has the higher net present value of owning. In the Appendix, we numerically solve a model with heterogeneous agents which allows for an endogenous share of absentee landlords, and show that the qualitative results of this section still hold.

short-term market by owner-occupiers does not change the overall supply of housing allocated to the long-term market, L. It also does not change S, which is by definition equal to $H - L$ (we think of S as the number of units that are *permanently* allocated towards short-term housing, as determined by absentee landlords.) The equilibrium supply of short and long-term housing are therefore:

$$S = f(Q - R - c)H_a \tag{7}$$

$$L = H - f(Q - R - c)H_a \tag{8}$$

Rental rates in the long-term market continue to be determined by the inverse demand curve of residents, $r(L)$. The equilibrium response of rental rates to a change in c becomes:

$$\frac{dR}{dc} = \frac{r'f'H_a}{1 - r'f'H_a} \leq 0 \tag{9}$$

Equation (9) is similar to Equation (5) except that H is replaced with H_a. Equation (9) therefore makes clear that it is the absentee landlords who affect the rental rate response to Airbnb because it is they who are on the margin between substituting their units between the short and long-term markets. When the share of owner-occupiers is high, the rental rate response to Airbnb will be low. In fact, the response of rental rates to Airbnb could be zero if all landlords are owner-occupiers.

Since long-term residents are ex-ante homogeneous, an equilibrium with a positive share of both renters and owner-occupiers requires that house prices make residents indifferent between renting and owning:

$$P = \frac{1}{1 - \delta}\left[R + \gamma g(Q - c)\right] \tag{10}$$

Equation (10) says that the price that residents are willing to pay for a home is equal to the present value of long-term rents plus the present value of renting excess capacity to the short-term market. The response of prices to a change in c is:

$$\frac{dP}{dc} = \frac{1}{1-\delta}\left[\frac{dR}{dc} - \gamma g'\right] \tag{11}$$

So, again, we see that prices are more responsive to a decrease in c than rental rates.

To summarize the results of this section, we derived three testable implications. First, rental rates should increase in response to the introduction of a home-sharing platform. This is because home-sharing causes some landowners to substitute away from supplying the long-term rental market and into the short-term rental market. Second, house prices should increase as well, but by an even greater amount than rents. This is because home-sharing affects house prices through two channels: first by increasing the rental rate, which then gets capitalized into house prices, and second by directly increasing the ability for landlords to utilize the home fully. Finally, the rental rate response will be smaller when there is a greater share of owner-occupiers. This is because owner-occupiers are not on the margin of substituting between the long-term and short-term markets, whereas absentee landlords are.[12] We now turn to testing these predictions in the data.

3 Data and Background on Airbnb

3.1 Background on Airbnb

Recognized by most as the pioneer of the sharing economy, Airbnb is a peer-to-peer marketplace for short-term rentals, where the suppliers (hosts) offer different kinds of accommodations (i.e. shared rooms, entire homes, or even yurts and treehouses) to prospective renters (guests). Airbnb was founded in 2008 and has experienced dramatic growth, going from just a few hundred hosts in 2008 to over three million properties supplied by over one million

[12]Another class of homeowners we have yet to discuss is vacation-home owners. Owners of vacation homes can be treated either as owner-occupiers with high γ (here γ is the amount of time spent living in their primary residence), or as absentee landlords, depending on how elastic they are with respect to keeping the home as a vacation property vs. renting it to a long-term tenant. In either case, the key implications of the model will not change.

hosts in 150,000 cities and 52 countries in 2017. Over 130 million guests have used Airbnb, and with a market valuation of over $31B, Airbnb is one of the world's largest accommodation brands.

3.2 Airbnb listings data

Our main source of data comes directly from the Airbnb website. We collected consumer-facing information about the complete set of Airbnb properties located in the United States and about the hosts who offer them. The data collection process spanned a period of approximately five years, from mid-2012 to the end of 2016. Scrapes were performed at irregular intervals between 2012 to 2014, and at a weekly interval starting January 2015.

Our scraping algorithm collected all listing information available to users of the website, including the property location, the daily price, the average star rating, a list of photos, the guest capacity, the number of bedrooms and bathrooms, a list of amenities such as WiFi and air conditioning, etc., and the list of all reviews from guests who have stayed at the property.[13] Airbnb host information includes the host name and photograph, a brief profile description, and the year-month in which the user registered as a host on Airbnb.

Our final dataset contains detailed information about 1,097,697 listings and 682,803 hosts spanning a period of nine years, from 2008 to 2016. Because of Airbnb's dominance in the home-sharing market, we believe that this data represents the most comprehensive picture of home-sharing in the U.S. ever constructed for independent research.

3.3 Calculating the number of Airbnb listings, 2008-2016

Once we have collected the data, the next step is to define a measure of Airbnb supply. This task requires two choices: first, we need to choose the geographic

[13] Airbnb does not reveal the exact street address or coordinates of the property for privacy reasons; however, the listing's city, street, and zipcode correspond to the property's real location.

granularity of our measure; second, we need to define the entry and exit dates of each listing to the Airbnb platform. Regarding the geographic aggregation, we conduct our main analysis at the zipcode level for a few reasons. First, it is the lowest level of geography for which we can reliably assign listings without error (other than user input error).[14] Second, neighborhoods are a natural unit of analysis for housing markets because there is significant heterogeneity in housing markets across neighborhoods within cities, but comparatively less heterogeneity within neighborhoods. Zipcodes will be our proxy for neighborhoods. Third, conducting the analysis at the zipcode level as opposed to the city level helps with identification. This is due to our ability to compare zipcodes within cities, thus controlling for any unobserved city level factors that may be unrelated to Airbnb but all affect neighborhoods within a city, such as a city-wide shock to labor productivity.

The second choice, how to determine the entry and exit date of each listing, comes less naturally. First, our scraping algorithm did not constantly monitor a listing's status to determine whether it was active or not, but rather obtained snapshots of the property available for rent in the US at different points in time until the end of 2014, and at the weekly level starting in 2015. Second, even if it did so, measuring active supply would still be challenging.[15] Thus, to construct the number of listings going back in time, we employ a variety of methods following Zervas et al. (2017), which we summarize in Table 1.

[14]Airbnb does report the latitude and longitude of each property, but only up to a perturbation of a few hundred meters. So it would be possible, but complicated, to aggregate the listings to finer geographies with some error.

[15]Estimating the number of active listings is a challenge even for Airbnb. Despite the fact that Airbnb offers an easy way to unlist properties, many times hosts neglect to do so, creating "stale vacancies" that seem available for rent but in actuality are not. Fradkin (2015), using proprietary data from Airbnb, estimates that between 21% to 32% of guest requests are rejected due to this effect.

Table 1: Methods for Computing the Number of Listings

	Listing is considered active ...
Method 1	starting from host join date
Method 2	for 3 months after host join date, and after every guest review
Method 3	for 6 months after host join date, and after every guest review

Method 1 is our preferred choice to measure Airbnb supply and will be our main independent variable in all the analyses presented in this paper. This measure computes a listing's entry date as the date its host registered on Airbnb and assumes that listings never exit. The advantage of using the host join date as the entry date is that for a majority of listings, this is the most accurate measure of when the listing was first posted. The disadvantage of this measure is that it is likely to overestimate the listings that are available on Airbnb (and accepting reservations) at any point it time. However, as discussed in Zervas et al. (2017), such overestimation would cause biases only if, after controlling for several zipcode characteristics, it is correlated with the error term.[16]

Aware of the fact that method 1 is an imperfect measure of Airbnb supply, we also experiment with alternative definitions of Airbnb listings' entry and exit. Methods 2 and 3 exploit our knowledge of each listing's review dates to determine whether a listing is active. The heuristic we use is as follows: a listing enters the market when the host registers with Airbnb and stays active for m months. We refer to m as the listing's Time To Live (TTL). Each time a listing is reviewed the TTL is extended by m months from the review date. If a listing exceeds the TTL without any reviews, it is considered inactive. A listing becomes active again if it receives a new review. In our analysis, we test two different TTLs, 3 months and 6 months.

[16]The absence of bias in this measure is also confirmed by Farronato and Fradkin (2018) where using Airbnb propietary data resulted in the same estimates obtained by Zervas et al. (2017)(where the data collection and measures of Airbnb supply are similar to those used in this paper).

Despite the fact that our different measures of Airbnb supply rely on different heuristics and data, because of Airbnb's tremendous growth, all our measures of Airbnb supply are extremely correlated. The correlation between method 1 and each other measure is above 0.95 in all cases. In the Appendix, we present robustness checks of our main results to the different measures of Airbnb supply discussed above, and show that results are qualitatively and quantitatively unchanged.

3.4 Zillow: rental rates and house prices

Zillow.com is an online real estate company that provides estimates of house and rental prices for over 110 million homes across the U.S. In addition to giving value estimates of homes, Zillow provides a set of indexes that track and predict home values and rental prices at a monthly level and at different geographical granularities.

For house prices, we use the Zillow Home Value Index (ZHVI) which estimates the median transaction price for the actual stock of homes in a given geographic unit and point in time. The advantage of using the ZHVI is that it is available at the zipcode-month level for over 13,000 zipcodes.

For rental rates, we use the Zillow Rent Index (ZRI). Like the ZHVI, Zillow's rent index is meant to reflect the median monthly rental rate for the actual stock of homes in a geographic unit and point in time. Crucially, Zillow's rent index is based on rental *list prices* and is therefore a measure of prevailing rents for new tenants. This is the relevant comparison for a homeowner deciding whether to place her unit on the short-term or long-term market. Moreover, because Zillow is not considered a platform for finding short-term housing, the ZRI should be reflective of rental prices in the long-term market.

3.5 Other data sources

We supplement the above data with several additional sources. We use monthly Google Trends data for the search term "airbnb", which we downloaded directly from Google. This index measures how often people worldwide search for the term “airbnb” on Google, and is normalized to have a value of 100 at the peak month. We use County Business Patterns data to measure the number of establishments in the food services and accommodations industry (NAICS code 72) for each zipcode in 2010. We collect from the American Community Survey (ACS) zipcode level 5-year estimates of median household income, population, share of 25-60 years old with bachelors' degrees or higher, employment rate, and owner-occupancy rate. Finally, we obtain annual 1-year estimates of housing vacancy rates at the Core Based Statistical Area (CBSA) level from the same source.

3.6 Summary statistics

Figure 2 shows the geographic distribution of Airbnb listings in June 2011 and June 2016. The map shows significant geographic heterogeneity in Airbnb listings, with most Airbnb listings occurring in large cities and along the coasts. Moreover, there exists significant geographic heterogeneity in the growth of Airbnb over time. From 2011 to 2016, the number of Airbnb listings in some zipcodes grew by a factor of 30 or more; in others there was no growth at all. Figure 3 shows the total number of Airbnb listings over time in our dataset using methods 1-3. Using method 1 as our preferred method, we observe that from 2011 to 2016, the total number of Airbnb listings grew by a factor of 30, reaching over 1 million listings in 2016.

Table 2 gives a sense of the size of Airbnb relative to the housing stock at the zipcode level, for the 100 largest CBSAs by population in our data. Even in 2016, Airbnb remains a small percentage of the total housing stock for most zipcodes. The median ratio of Airbnb listings to housing stock is 0.21%, and the 90th percentile is 1.88%. When comparing to the stock of vacant homes, Airbnb begins to appear more significant. The median ratio of Airbnb listings to vacant homes is 2.63%, and the 90th percentile is 20%. Perhaps the most salient comparison—at least from the perspective of a potential renter—is the number of Airbnb listings relative to the stock of homes listed as vacant and for rent. This statistic reaches 13.7% in the median zipcode in 2016 and 129%

in the 90th percentile zipcode. This implies that in the median zipcode, a local resident looking for a long-term rental unit will find that about 1 in 8 of the potentially available homes are being placed on Airbnb instead of being made available to long-term residents. Framed in this way, concerns about the effect of Airbnb on the housing market do not appear unfounded.

4 Methodology

Let Y_{ict} be either the price index or the rent index for zipcode i in CBSA c in year-month t, let $Airbnb_{ict}$ be a measure of Airbnb supply, and let $oorate_{ic,2010}$ be the owner-occupancy rate in 2010.[17] We assume the following causal relationship between Y_{ict} and $Airbnb_{ict}$:

$$\ln Y_{ict} = \alpha + \beta Airbnb_{ict} + \gamma Airbnb_{ict} \times oorate_{ic,2010} + X_{ict}\eta + \epsilon_{ict} \quad (12)$$

where X_{ict} is a vector of observed time-varying zipcode characteristics, and ϵ_{ict} contains unobserved factors which may additionally influence Y_{ict}. If the unobserved factors are uncorrelated with $Airbnb_{ict}$, conditional on X_{ict}, then we can consistently estimate β and γ by OLS. However, ϵ_{ict} and $Airbnb_{ict}$ may be correlated through unobserved factors at the zipcode, city, and time levels. We allow ϵ_{ict} to contain unobserved zipcode level factors δ_i, and unobserved time-varying factors that affect all zipcodes within a CBSA equally, θ_{ct}. Writing: $\epsilon_{ict} = \delta_i + \theta_{ct} + \xi_{ict}$, Equation (12) becomes:

$$\ln Y_{ict} = \alpha + \beta Airbnb_{ict} + \gamma Airbnb_{ict} \times oorate_{ic,2010} + X_{ict}\eta + \delta_i + \theta_{ct} + \xi_{ict} \quad (13)$$

Even after controlling for unobserved factors at the zipcode and CBSA-year-month level, there may still be some unobserved *zipcode-specific*, *time-varying* factors contained in ξ_{ict} that are correlated with $Airbnb_{ict}$. To address

[17] We use the owner-occupancy rate in 2010 to minimize concerns about endogeneity of the owner-occupancy rate. In the Appendix, we show that the results are robust to using the contemporaneous owner-occupancy rate calculated from ACS 5-year estimates from 2011 to 2016.

this issue, we construct an instrumental variable which is plausibly uncorrelated with local shocks to the housing market at the zipcode level, ξ_{ict}, but likely to affect the number of Airbnb listings.

Our instrument begins with the worldwide Google Trends search index for the term "airbnb", g_t, which measures the quantity of Google searches for "airbnb" in year-month t. Such trends represent a measure of the extent to which awareness of Airbnb has diffused to the public, including both demanders and suppliers of short-term rental housing. Figure 1 plots g_t from 2008 to 2016, and it is representative of the explosive growth of Airbnb over the past ten years. Crucially, the search index is *not* likely to be reflective of growth in overall tourism demand, because it is unlikely to have changed so much over this relatively short time period. Moreover, it should not be reflective of overall growth in the supply of short-term housing, *except* to the extent that it is driven by Airbnb.

The CBSA-year-month fixed effects θ_{ct} already absorb any unobserved variation at the year-month level. Therefore, to complete our instrument we interact g_t with a measure of how attractive a zipcode is for tourists in base year 2010, $h_{i,2010}$. We measure "touristiness" using the number of establishments in the food services and accommodations industry (NAICS code 72) in a specific zipcode. Zipcodes with more restaurants and hotels may be more attractive to tourists because these are services that tourists need to consume locally—thus, it matters how many of these services are near the tourist's place of stay. Alternatively, the larger number of restaurants and hotels may reflect an underlying local amenity that tourists value.

Our operating assumption is that landlords in more touristy zipcodes are more likely to switch from the long-term market to the short-term market in response to learning about Airbnb. Landlords in more touristy zipcodes may be more likely to switch because they can book their rooms more frequently, and at higher prices, than in non-touristy zipcodes. We can verify this assumption by examining the relationship between Google trends and the difference in Airbnb listings for more touristy and less touristy zipcodes. Figure 4 shows that such difference increases as Airbnb awareness increases confirming our hypothesis.

In order for the instrument to be valid, $z_{ict} = g_t \times h_{i,2010}$ must be uncorrelated with the zipcode-specific, time-varying shocks to the housing market, ξ_{ict}. This would be true if either ex-ante touristiness in 2010 ($h_{i,2010}$) is independent of zipcode level shocks (ξ_{ict}), or growth in worldwide Airbnb searches (g_t) is independent of zipcode level shocks. To see how our instrument addresses potential confounding factors, consider changes in zipcode level crime rate as an omitted variable. It is unlikely that changes to crime rates across all zipcodes are systematically correlated in time with worldwide Airbnb searches. Even if they were, they would have to correlate in such a way that the correlation is systematically stronger or weaker in more touristy zipcodes. Moreover, these biases would have to be systematically present within all cities in our sample. Of course, we cannot rule this possibility out completely. We therefore now turn to a detailed discussion of the instrument and its validity, and present some exercises that suggest that the exogeneity assumption is likely satisfied.

4.1 Discussion: Validity of the instrumental variable

The construction of an instrumental variable using the interaction of a plausibly exogenous time-series (Google trends) with a potentially endogenous cross-sectional exposure variable (the touristiness measure) is an approach that was popularized by Bartik (1991) and that has been used in many prominent recent papers (Peri (2012); Dube and Vargas (2013); Nunn and Qian (2014); Hanna and Oliva (2015); Diamond (2016)).

The approach is popular because one can often argue that some aggregate time trend, which is exogenous to local conditions, will affect different spatial units systematically along some cross-sectional exposure variable. In the classic Bartik (1991) example, national trends in industry-specific productivity are interacted with the historical local industry composition to create an instrument for local labor demand. Such instrument will be valid if the interaction of the aggregate time trend with the exposure variable is independent of the error term. This could happen if either the time trend is independent of the error term ($E[g_t, \xi_{ict}] = 0$) or if the exposure variable is independent of the error term ($E[h_{i,2010}, \xi_{ict}] = 0$). While this may seem plausible at first glance, Christian and Barrett (2017) point out that if there are long-run time trends in the error term, and if these long-run trends are systematically different along the exposure variable, then the exogeneity assumption may fail. In our context, a story that may be told is the following. Suppose there is a long-run trend towards gentrification, which leads to higher house prices over time. Suppose also that the trend of gentrification is higher in more touristy zipcodes. Since there is also a systematic long-run trend in the time-series variable, g_t, the instrument $g_t h_{i,2010}$ is no longer independent of the error term, and 2SLS estimates may reflect the effects of gentrification rather than home-sharing.

We now proceed to make four arguments for why the exogeneity condition is likely to hold in our setting.

Parallel pre-trends

As Christian and Barrett (2017) noted, the first stage of this instrumental variable approach is analogous to a difference-in-differences (DD) coefficient estimates. In our case, since the specification includes year-month and zipcode fixed effects, the variation in the instrument comes from comparing Airbnb listings between high- and low-Airbnb awareness year-months, and between high- and low-touristiness zipcodes. Because of this, Christian and Barrett (2017) suggest testing whether spatial units with different levels of the exposure variable have parallel trends in periods before g_t takes effect. This is similar to testing whether control and treatment groups have parallel pre-trends in DD analysis. To do this, we plot the Zillow house price index for zipcodes in different quartiles of 2010 touristiness ($h_{i,2010}$), from 2009 to the end of 2016.[18] The results are shown in Figure 5. The figure shows that there are no differential pre-trends in the Zillow Home-Value Index (ZHVI) for zipcodes in different quartiles of touristiness until after 2012, which also happens to be when interest in Airbnb began to grow according to Figure 1. This is true

[18]We cannot repeat this exercise with rental rates because Zillow rental price data did not begin until 2011 or 2012 for most zipcodes.

both when computing the raw averages of ZHVI within quartile (top panel) and when computing the average of the residuals after controlling for zipcode and CBSA-year-month fixed effects (bottom panel). The lack of differential pre-trends suggests that zipcodes with different levels of touristiness do *not* generally have different long-run house price trends, but they only began to diverge after 2012 when Airbnb started to become well known.

Placebo test

The above test is not perfect, especially because 2012 happens to be the year in which house prices began to recover from the Great Recession. Because of this, it is possible that touristy zipcodes have a different recovery pattern than non-touristy zipcodes. We therefore consider a second test to support the validity of the instrument. Recall that our instrumental variable relies on the assumption that increases in Airbnb awareness (measured using Google trends) will differentially affect the number of Airbnb listings in high-touristiness zipcodes and in low-touristiness zipcodes. Following Christian and Barrett (2017) we implement a form of randomization inference to test whether this type of instrument is really exogenous. The idea behind this test is that by randomizing the endogenous variable of interest (the number of Airbnb listings is a specific zipcode) while holding constant everything else should eliminate (or at least attenuate) the causal effect of Airbnb.

To do so we keep constant touristiness, Google trends, the zipcodes experiencing any Airbnb entry, observable time-varying zipcode characteristics, housing market variables, and the aggregate number of Airbnb listings in any year-month period. However, among the zipcodes with positive Airbnb entry, we randomly assign the specific number of Airbnb listings among these zipcodes; for example, we randomly assign to zipcode i the variable $Airbnb_{jct}$ (i.e., the Airbnb counts of zipcode j of CBSA c for every t spanning the period from 2011 to 2016).

Note that this new dataset still preserves possible sources of endogeneity such as zipcode touristiness and spurious time trends; however, the randomization eliminates a major source of variation needed for our instrument to work because now it is not necessarily the case that, for the same level of Airbnb awareness, high-touristiness zipcodes experience stronger Airbnb growth than low-touristiness zipcodes. This means that a 2SLS estimate of the effect of Airbnb using this dataset should produce results that are indistinguishable from zero (or much smaller than the estimates on the real dataset), unless there is some spurious correlation between the instrument and our dependent variable (i.e., the exclusion restriction does not hold).

We estimate the 2SLS specification on this dataset for 100 draws of randomized allocations of Airbnb listings among zipcodes, and find that the measured effect of Airbnb completely disappears for all of our dependent variables, i.e., rent index, price index, and price-to-rent ratio.[19] Thus, this test strongly supports the validity of our instrument.

IV has no effect in non-Airbnb zipcodes

To further provide support to the validity of our instrument we perform another test which consists of checking whether the instrumental variable predicts house prices and rental rates in zipcodes that were never observed to have any Airbnb listings. If the instrument is valid, then it should only be correlated to house prices and rental rates through its effect on Airbnb listings, so in areas with no Airbnb we should not see a positive relationship between the instrument and house prices and rental rates.[20] To test this, we regress the Zillow rent index, house price index, and price-to-rent ratio (our three outcomes of interest) on the instrumental variable directly, using only data from zipcodes in which we never observed any Airbnb listings. Table 3 reports the results of

[19]The median estimate (standard error) of β and γ are 0.17 (1.25) and 8.77e-07 (8.44e-07) for the rent index, -.23 (1.08) and 1.53e-06 (1.04e-06) for the price index, and -.27 (1.45) and 1.56e-06 (1.27e-06) for the price-to-rent ratio.

[20]This exercise is similar in spirit to an exercise performed in Martin and Yurukoglu (2017) to support the validity of an instrument. In Martin and Yurukoglu (2017), the channel position of Fox News in the cable line up is used as an instrument for the effect of Fox viewership on Republican voting. They show that the future channel position of Fox News is not correlated with Republican voting in the time periods before Fox News. This is analogous to us showing that our instrument is not correlated with house prices and rents in zipcodes without Airbnb.

these regressions and shows that, conditional on the fixed effects and zipcode demographics, we do not find any statistically significant relationship between the instrument and house prices/rental rates in zipcodes without Airbnb. If anything, we find that there is a *negative* relationship between the instrument and house prices/rental rates in zipcodes without Airbnb, though the estimates are imprecise and the sample size is considerably reduced when considering only such zipcodes.[21] Thus, there does not seem to be any evidence that the instrument would be positively correlated with house prices/rental rates, except through its effect on short-term rentals.

Robustness to the inclusion of demographic controls

Of course, the above test to support the validity of the instrument is not perfect either. The sample of zipcodes that never had any Airbnb listings could be fundamentally different from the sample of zipcodes that did.[22] We therefore make one final argument to support the validity of our instrument, which is that the regression results we will present in Section 5 are robust to the inclusion of zipcode demographic characteristics. Because the included demographic controls (population, household income, share of college-educated, and employment rate) are fairly basic measurements of zipcode level economic outcomes, they are likely to be highly correlated with other unobserved factors that affect zipcode level housing markets. Therefore, the fact that our results are not affected by these controls suggests that it is unlikely that the instrument is correlated with other unobserved zipcode level factors that affect housing markets. To see this, consider the story about gentrification posited above. If the relationship between Airbnb listings and house prices/rental rates is spuriously driven by gentrification, then one would expect the estimated effect to be reduced once controlling for neighborhood level income and education; however, since this does not happen, gentrification seems unlikely to be an omitted driver of the results.

[21]If we regress house prices and rental rates on the instrument for zipcodes *with* Airbnb, we find a positive and statistically significant relationship.

[22]Indeed, Table 4 shows that there is a significant difference when comparing zipcodes that observed and never observed any Airbnb listings.

5 Results and Extensions

5.1 The effect of Airbnb on house prices and rents

We begin by reporting results in which $Airbnb_{ict}$ is measured as the log of one plus the number of listings as measured by method 1 in Table 1.[23] Doing so, we estimate a specification similar to that used in Zervas et al. (2017) and Farronato and Fradkin (2018), where the authors estimate the impact of Airbnb on the hotel industry.

We consider three dependent variables: the log of the Zillow Rent Index, the log of the Zillow Home-Value Index, and the log of the price-to-rent ratio. In order to maintain our measure of touristiness, $h_{i,2010}$, as a pre-period variable, only data from 2011 to 2016 are used. This time frame covers all of the period of significant growth in Airbnb (see Figure 3). We also include only data from the 100 largest CBSAs, as measured by 2010 population.[24] Since the regression in Equation 13 has two endogenous regressors ($Airbnb_{ict}$ and $Airbnb_{ict} \times oorate_{ic,2010}$), two instruments are used for the two-stage least squares estimation ($g_t \times h_{i,2010}$ and $g_t \times h_{i,2010} \times oorate_{ic,2010}$).

Table 5 reports the regression results when the dependent variable is the log Zillow rent index. Column 1 reports the results from a simple OLS regression of log ZRI on log listings and no controls. Without controls, a 1% increase in Airbnb listings is associated with a 0.098% increase in rental rates. Column 2 includes zipcode and CBSA-year-month fixed effects. With the fixed effects, the estimated coefficient on Airbnb declines by an order of magnitude. Column 3 includes the interaction of Airbnb listings with the zipcode's owner-occupancy rate. Column 3 shows the importance of controlling for owner-occupancy rate, as it significantly mediates the effect of Airbnb listings. Column 4 includes time-varying zipcode level characteristics, including the log total population, the log median household income, the share of 25-60

[23]We add one to the number of listings to avoid taking logs of zero. In the Online Appendix, we show that our results are robust to dropping observations with 0 listings and using ln(listings) instead.

[24]The 100 largest CBSAs constitute the majority of Airbnb listings (over 80%). In the Online Appendix we show that our results are robust to the inclusion of more CBSAs.

years old with Bachelors' degrees or higher, and the employment rate. Because these measures are not available at a monthly frequency, we linearly interpolate them to the monthly level using ACS 5-year estimates from 2011 to 2016.[25] Column 4 shows that the results are robust to the inclusion of these zipcode demographics. Finally, columns 5 and 6 report the 2SLS results using the instrumental variable without and with time-varying zipcode characteristics as controls. Using the results from column 6 – our preferred specification – we estimate that a 1% increase in Airbnb listings in zipcodes with the median owner-occupancy rate (72%) leads to a 0.018% increase in rents. As predicted by our model, the effect of Airbnb is significantly declining in the owner-occupancy rate. At 56% owner-occupancy rate (the 25th percentile), the effect of a 1% increase in Airbnb listings is to increase rents by 0.024%, and at 82% owner-occupancy rate (the 75th percentile), the effect of a 1% increase in Airbnb listings is to increase rents by 0.014%.

Table 6 repeats the regressions with the log Zillow house price index as the dependent variable. As with the rental rates, we find that controlling for owner-occupancy rate is very important, as the estimated direct effect of Airbnb listings increases by an order of magnitude when controlling for the interaction vs. not. Further, including demographic controls still does not affect the results. Using the coefficients reported in column 6 of Table 6, we estimate that a 1% increase in Airbnb listings leads to a 0.026% increase in house prices for a zipcode with a median owner-occupancy rate. The effect increases to 0.037% in zipcodes with an owner-occupancy rate equal to the 25th percentile, and decreases to 0.019% in zipcodes with an owner-occupancy rate equal to the 75th percentile.

It is worth noting that in both the rental rate and house price regressions, the 2SLS estimates (columns 5 and 6 of Tables 5 and 6) are about twice as large as the OLS estimates (columns 3 and 4 of Tables 5 and 6) . This goes against our initial intuition that omitted factors (such as gentrification) are most likely to be positively correlated with both Airbnb listings and house prices/rents, thus creating a positive bias. However, we note that the OLS estimate may

[25]Results are not sensitive to different types of interpolations.

also be negatively biased or biased towards zero for two reasons. First, there may be measurement error in the true amount of home-sharing, leading to attenuation bias. Measurement error may arise from the fact that we only estimate the number of Airbnb listings, and we do not know their exact entry and exit. Measurement error may also arise from the fact that there are other home-sharing platforms besides Airbnb, that we do not measure. Our estimate for the number of listings is therefore a noisy measure of the true number of short-term rentals. Second, simultaneity bias may be negative if higher rental rates in the long-term rental market would cause a decrease in the number of Airbnb listings, *ceteris paribus*. This is true in our model because an increase in the long-term rental rate (holding Q fixed), would decrease the number of landlords choosing to supply the short-term market, and it is likely to be true in reality as well.

Finally, Table 7 reports the regression results when log price-to-rent ratio is used as the dependent variable. Column 6 shows that the effect of Airbnb listings on the price-to-rent ratio is positive, and that, similarly to rents and prices, the effect is declining in owner-occupancy rate. At the median owner-occupancy rate, a 1% increase in Airbnb listings leads to a statistically significant 0.01% increase in the price-to-rent ratio.

To summarize the results in Tables 5-7, we showed that 1) an increase in Airbnb listings leads to both higher house prices and rental rates; 2) the effect is higher for house prices than it is for rental rates; and 3) the effect is decreasing in the zipcode's owner-occupancy rate. These results are all consistent with the model presented in Section 2, thus providing evidence that home-sharing indeed increases housing costs by reallocating long-term rentals to the short-term market, but also that home-sharing increases homeowners' option value for utilizing excess capacity.

5.2 Robustness checks

We now report a number of robustness checks to reinforce the validity of our estimates. First, we re-estimate our specification using different subsamples of the data. The main purpose of these checks is to confirm that the results are not being driven by only a select number of cities, zipcodes, or time periods. In doing so, our goal is to further reduce concerns about possible omitted variables correlated with location and time that may drive the results presented in Section 5.1. For example, consider the zipcode location and specifically whether it is located close to the city center. One may argue that zipcodes close to the city center would have experienced a positive increase in rents and house prices independently of the presence of Airbnb (and of course such zipcodes are also more likely to have a higher number of Airbnb listings). Second, we perform a specification test that uses an alternative functional form of Airbnb supply. This test guards against concerns related to our choice of using a log-log specification to estimate the impact of Airbnb on the housing market.

Zipcodes near and far from the city-center

First, we repeat the 2SLS regressions with full controls separately for zipcodes that are "near" to their CBSA's city center and for zipcodes that are "far" from the city center. The city center is obtained using Microsoft's Bing Maps API, and zipcode centroids are obtained from the Census Bureau. A zipcode is counted as "near" to the CBD if it is closer than the CBSA median, and "far" otherwise. The first two rows of Table 8 report the results. The qualitative results hold in both the near and far samples, though it seems that the effects are larger in the far group. This confirms that the results are not being solely driven by a few zipcodes close to downtown areas, and that home-sharing is having an impact even on zipcodes that are further from the city center.

Early and late time periods

Second, we repeat the regressions separately for two time periods: 2011-2013 and 2014-2016. Rows 3-4 of Table 8 report these results. Again, the main qualitative results can be seen in both time periods, though the effect of owner-occupancy rate seems to be a lot weaker in the earlier period than in the later period. We speculate that this could be due to the possibility that Airbnb first attracted those users with spare rooms or houses not on the long-term market (e.g., vacation rentals), and that only recently Airbnb became an attractive option for landlords that previously rented in the long-term market.

Large and small CBSAs

Finally, we repeat the regressions separately for the 30 largest CBSAs, and for the CBSAs ranked 31-100 in 2010 population. Rows 5-6 of Table 8 report the results. The qualitative results hold for both samples, though the results are not statistically significant in the rank 31-100 sample when the outcome is price-to-rent ratio. The effects of Airbnb appear to be stronger in the larger cities, which could be driven by a number of factors, including differences in housing demand and housing supply elasticities.

Log-density specification

In our main results, we have used a log-log specification to measure the effect of Airbnb listings on house prices and rental rates. This is because such specification provides us with easily, interpretable coefficients in the form of elasticity that is often used in competitive settings, and it has been used in the past in the context of Airbnb (Farronato and Fradkin, 2018; Zervas et al., 2017). However, as Zervas et al. (2017) observed, the log-log specification implies constant elasticity, an assumption that might not hold in our settings.

To make sure that our results are not driven by the log-log choice we use an alternative specification in which $Airbnb_{ict}$ in Equation (13) is measured as the number of Airbnb listings divided by the total occupied housing stock.[26] We call this measure "Airbnb density."

We report the results using the log-density specification in Table 9. We report OLS results in column 1 and 2SLS results in column 2. The main results continue to hold qualitatively: 1) higher Airbnb density leads to higher house prices and rental rates; 2) the effect is higher for house prices than rental rates;

[26] Data on total occupied housing stock is from ACS 5-year estimates from 2011 to 2016.

and 3) the effect is decreasing in owner-occupancy rate.

One of the downsides of the log-density specification is that Airbnb density is extremely skewed[27] and using $g_t \times h_{i,2010}$ as the instrument, the first stage becomes very weak and we fail to reject underidentification.[28] We therefore report results using an augmented set of instruments formed by interacting second order polynomials of g_t, $h_{i,2010}$, and $oorate_{i,2010}$. In the Appendix, we show that the qualitative results are robust to a number of different sets of instruments, but that the coefficients are somewhat sensitive to the choice of instruments. This is why the log-log specification, which has proven to be very robust, remains our preferred specification.

Additional checks

In the Online Appendix, we report a number of additional robustness checks, such as using alternative measures of Airbnb listings, the effect of including even smaller CBSAs, and the effect of dropping zipcodes with zero or a small number of listings. The main results are robust to all these alternative specifications.

5.3 Effect magnitudes

In this section we consider the economic significance of our estimated effects. Our baseline result is that a 1% increase in Airbnb listings leads to a 0.018% increase in rents and a 0.026% increase in house prices, at a median owner-occupancy rate zipcode. The median year-on-year growth rate in Airbnb listings was 28% across zipcodes in the top 100 CBSAs. Taken at the sample median, then, Airbnb growth explains 0.5% in annual rent growth and 0.7% of annual price growth.

Another way to calculate effect size is to calculate the Airbnb contribution to year-over-year rent and house price growth for each zipcode by multiplying median year-over-year changes in log listings by the estimated coefficients $\hat{\beta} +$

[27]The skewness is 129.58 compared to a mean of 0.007 and variance of 0.06.

[28]In the rent regression, an underidentification test using the Kleibergen and Paap (2006) rk LM statistic fails to reject underidentification with a p-value of 0.6650.

$\hat{\gamma} \times oorate_{i,2010}$. We report these effects in Table 10 for the median zipcodes in the 10 largest CBSAs, as well as for the median zipcode in our sample of 100 largest CBSAs. We also include the average year-on-year rent and price growth for comparison. While the size of the Airbnb contribution may seem large, we caution that estimating the effect at the sample median masks substantial heterogeneity in the actual experiences of different zipcodes, and ignores the very likely possibility of heterogeneous treatment effects. We also note that our estimated effects are consistent with those found in Horn and Merante (2017), who study the effect of Airbnb on rents in Boston from 2015-2016. They found that a one standard deviation increase in Airbnb listings led to a 0.4% increase in rents. In our data, the within-CBSA standard deviation in log listings is 0.27 for 2015-2016, which at the median owner-occupancy rate implies a 0.54% increase in rents using our estimates.

5.4 The effect of home-sharing on housing reallocation

We close the paper by presenting some suggestive evidence that home-sharing affects rental rates and house prices through the reallocation of housing stock. To do this, we investigate the effect of Airbnb on housing vacancies. Because vacancy data is not available at the zipcode level at a monthly or annual frequency, we focus on annual CBSA level vacancies. We regress vacancy rates at the CBSA-year level on the number of Airbnb listings, year fixed effects, and CBSA fixed effects. Data on vacancies come from annual ACS 1-year estimates at the CBSA level.[29] Table 11 reports the results.

The first thing to note in Table 11 is that the number of Airbnb listings at the CBSA level appears uncorrelated with the total number of vacancies, once controlling for CBSA and year fixed effects (column 1). However, when we break the vacancy rate down by the type of vacancy, we find a positive and statistically significant relation with the share of homes classified as vacant for seasonal or recreational use and a negative and statistically significant

[29] We compute the total number of vacancies as sum of the number of vacant seasonal units, vacant-for-rent units, and vacant-for-sale units. We ignore vacant units that are for migrant workers, and we ignore vacant units for which the reason for vacancy is unknown.

association with the share of homes that are vacant-for-rent and vacant-for-sale.

It is important to note that the Census Bureau classifies homes as vacant even if they are temporarily occupied by persons who usually live elsewhere. Thus, homes allocated permanently to the short-term market are supposed to be classified as vacant, and will likely also be classified as seasonal or recreational homes by their owners and/or neighbors.[30] The positive association of Airbnb with vacant-seasonal homes, and the negative association with vacant-for-rent and vacant-for-sale homes is therefore consistent with absentee landlords substituting away from the rental and for-sale markets for long-term residents and allocating instead to the short-term market.

6 Discussion & Conclusion

The results presented in this paper suggest that the increased ability to home-share has led to increases in both rental rates and house prices. The increases in rental rates and house prices occur through two channels. In the first channel, home-sharing increases rental rates by inducing some landlords to switch from supplying the market for long-term rentals to supplying the market for short-term rentals. The increase in rental rates through this channel is then capitalized into house prices. In the second channel, home-sharing increases house prices directly by enabling homeowners to generate income from excess housing capacity. This raises the value of owning relative to renting, and therefore increases the price-to-rent ratio directly.

The results in this paper contribute to the debate surrounding home-sharing and its impact on the housing market. While Airbnb and proponents of the sharing economy argue that the platform is not responsible for higher house prices and rental rates,[31] critics of home-sharing argue that Airbnb does

[30] When a home is vacant, Census workers will interview neighbors about the occupancy characteristics of the home.

[31] For example, Airbnb disputed the findings of a recent report on the effects of the platform on the housing market in New York City. See: `https://www.citylab.com/equity/2018/03/what-airbnb-did-to-new-york-city/552749/`.

raise housing costs for local residents. This paper provides evidence confirming this latter hypothesis, and it does so using the most comprehensive dataset about home-sharing in the US available to date. Moreover, this paper also provides evidence that home-sharing increases the value of homes by allowing owners to better utilize excess capacity, for example by allowing owners to rent spare bedrooms, or the entire home when on vacation.

Turning to how cities and municipalities should deal with the steady increase in home-sharing, our view is that regulations on home-sharing should (at most) seek to limit the reallocation of housing stock from long-term rentals to short-term rentals, without discouraging the use of home-sharing by owner-occupiers. One regulatory approach could be to only levy occupancy tax on home sharers who rent the entire home for an extended period of time, or to require a proof of owner-occupancy in order to avoid paying occupancy tax.

Of course, this research does not come without limitations. First, we must recognize that our Airbnb data is imperfect: while we observe properties listed on Airbnb, we do not observe exact entry and exit of these properties. However, using Airbnb proprietary data Farronato and Fradkin (2018) obtain very similar elasticity estimates to Zervas et al. (2017) who use a similar approach to ours to obtain Airbnb data and measure Airbnb supply. This, along with our extensive set of robustness checks, reassures us about the validity of our results.

Second, we need to keep in mind that in settings where the effects are likely to be heterogeneous, a 2SLS estimate does not represent the Average Treatment Effect (ATE) but instead a Local Average Treatment Effect (LATE), or the effect of Airbnb on the subset of "complier" zipcodes – those zipcodes that are induced by the instrument to change the value of the endogenous regressor. Thus, our estimate do not necessarily reflect the average effect of Airbnb on any zipcodes. Despite this limitation, however, we estimate magnitudes that are similar to those obtained by Horn and Merante (2017) for the city of Boston. Finally, our model does not take into account possible spillover effects the neighboring zipcodes can have on each other.

To summarize the state of the literature on home-sharing, research (including this paper) has found that home-sharing 1) raises local rental rates by causing a reallocation of the housing stock; 2) raises house prices through both the capitalization of rents and the increased ability to use excess capacity; and 3) induces market entry by small suppliers of short-term housing who compete with traditional suppliers (Zervas et al. (2017); Farronato and Fradkin (2018)). More research is needed, however, in order to achieve a complete welfare analysis of home-sharing. For example, home-sharing may have positive spillover effects on local businesses if it drives a net increase in tourism demand. On the other hand, home-sharing may have negative spillover effects if tourists create negative externalities, such as noise or congestion, for local residents. Moreover, home-sharing introduces an interesting new mechanism for scaling down the local housing supply in response to negative demand shocks—a mechanism that was not possible when all of the residential housing stock was allocated to the long-term market. Understanding the impact of such mechanism on the housing market is an open question to date. We leave these research questions for future work.

References

Angrist, Josh, Sydnee Caldwell, and Jonathan Hall, "Uber versus taxi: A driver's eye view," 2017.

Bartik, Timothy J., *Who Benefits from State and Local Economic Development Policies?*, Kalamazoo, MI: W.E. Upjohn Institute for Employment Research, 1991.

Chen, M. Keith, Judith A. Chevalier, Peter E. Rossi, and Emily Oehlsen, "The Value of Flexible Work: Evidence from Uber Drivers," Working Paper 23296, NBER 2017.

Christian, Paul and Christopher P. Barrett, "Revisiting the Effect of Food Aid on Conflict: A Methodological Caution," Working Paper 8171, World Bank 2017.

Diamond, Rebecca, "The Determinants and Welfare Implications of US WOrkers' Diverging Location Choices by Skill: 1980-2000," *American Economic Review*, 2016, *106* (3), 479–524.

Dube, Oeindrila and Juan F Vargas, "Commodity price shocks and civil conflict: Evidence from Colombia," *The Review of Economic Studies*, 2013, *80* (4), 1384–1421.

Edelman, Benjamin G and Michael Luca, "Digital discrimination: The case of airbnb.com," Working Paper, Harvard Business School 2014.

Edelman, Benjamin, Michael Luca, and Dan Svirsky, "Racial discrimination in the sharing economy: Evidence from a field experiment," *American Economic Journal: Applied Economics*, 2017, *9* (2), 1–22.

Einav, Liran, Chiara Farronato, and Jonathan Levin, "Peer-to-Peer Markets," *Annual Review of Economics*, October 2016, *8*, 615–635.

Farronato, Chiara and Andrey Fradkin, "The welfare effects of peer entry in the accommodation market: The case of airbnb," Technical Report, National Bureau of Economic Research Working Paper 2018.

Fradkin, Andrey, "Search frictions and the design of online marketplaces," Working Paper, Massachusetts Institute of Technology 2015.

_ , Elena Grewal, and David Holtz, "The determinants of online review informativeness: Evidence from field experiments on Airbnb," Technical Report, Working Paper 2017.

Gong, Jing, Brad N. Greenwood, and Yiping Song, "Uber Might Buy Me a Mercedez Benz: An Empirical Investigation of the Sharing Economy and Durable Goods Purchase," Working Paper, SSRN 2017.

Hall, Jonathan V. and Alan B. Krueger, "An Analysis of the Labor Market for Uber's Driver-Partners in the United States," *ILR Review*, 2017.

Hanna, Rema and Paulina Oliva, “The effect of pollution on labor supply: Evidence from a natural experiment in Mexico City,” *Journal of Public Economics*, 2015, *122*, 68–79.

Horn, Keren and Mark Merante, “Is Home Sharing Driving Up Rents? Evidence from Airbnb in Boston,” *Journal of Housing Economics*, 2017.

Horton, John J. and Richard J. Zeckhauser, “Owning, Using and Renting: Some Simple Economics of the “Sharing Economy”,” Working Paper 22029, NBER 2016.

Kleibergen, Frank and Richard Paap, “Generalized Reduced Rank Tests Using the Singular Value Decomposition,” *Journal of Econometrics*, 2006, *133* (1), 97–126.

Lee, Dayne, “How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendations,” *Harvard Law & Policy Review*, 2016, pp. 229–254.

Martin, Gregory J. and Ali Yurukoglu, “Bias in Cable News: Persuasion and Polarization,” *American Economic Review*, 2017, *107* (9), 2565–2599.

Nunn, Nathan and Nancy Qian, “US food aid and civil conflict,” *American Economic Review*, 2014, *104* (6), 1630–66.

Peri, Giovanni, “The effect of immigration on productivity: Evidence from US states,” *Review of Economics and Statistics*, 2012, *94* (1), 348–358.

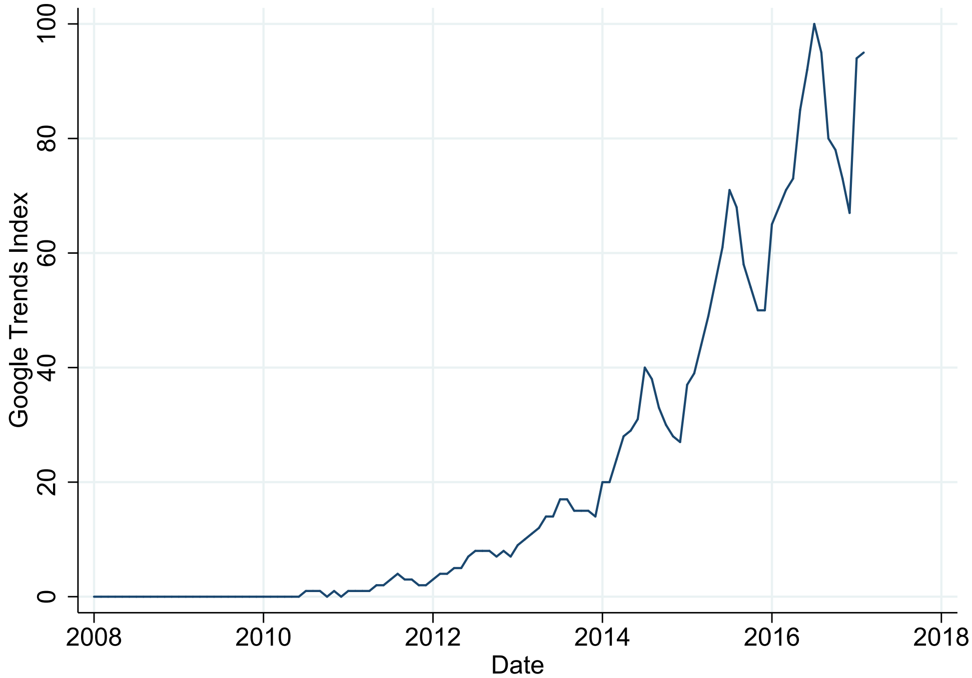
Proserpio, Davide and Gerard Tellis, “Baring the Sharing Economy: Concepts, Classification, Findings, and Future Directions,” *Working Paper*, 2017.

_ , Wendy Xu, and Georgios Zervas, “You Get What You Give: Theory and Evidence of Reciprocity in the Sharing Economy,” 2017.

Zervas, Georgios, Davide Proserpio, and John Byers, “A first look at online reputation on Airbnb, where every stay is above average,” 2015.

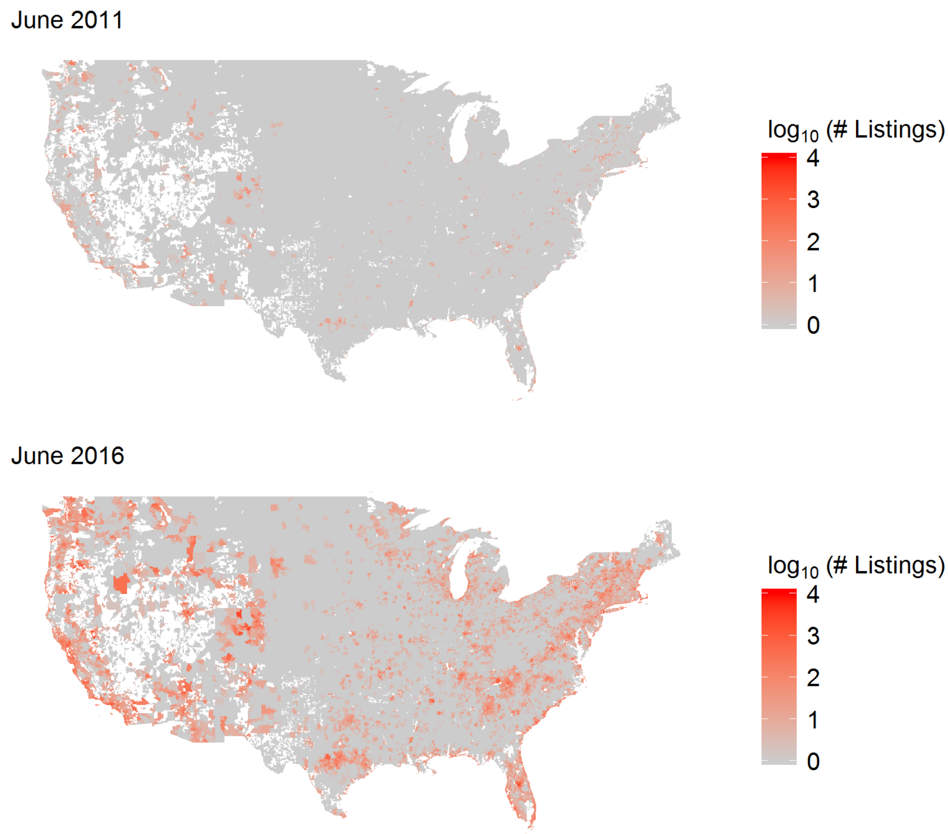
_ , _ , and John W. Byers, "The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry," *Journal of Marketing Research*, 2017, *54* (5), 687–705.

Figure 1: Google Trends Search Index for Airbnb (Worldwide, 2008-2017)



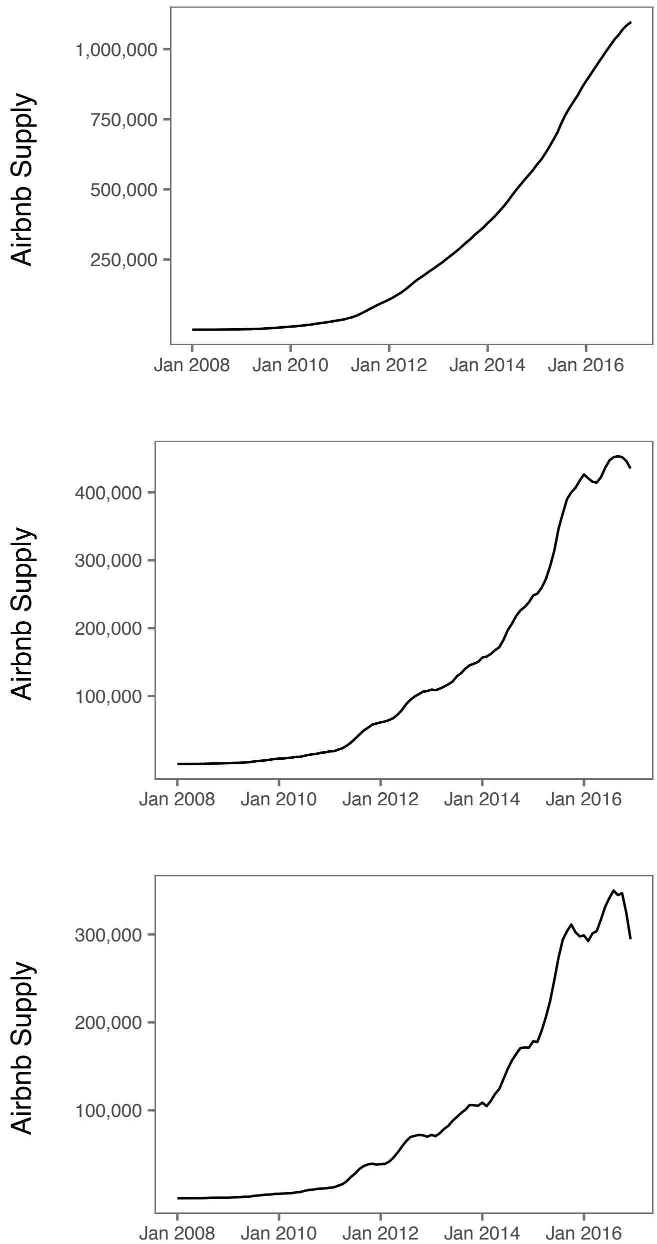
Note: Weekly Google Trends index for the single English search term "Airbnb", from any searches worldwide. Google Trends data are normalized so that the date with the highest search volume is given the value of 100.

Figure 2: Map of Airbnb Listings by Zipcode, 2011-2016



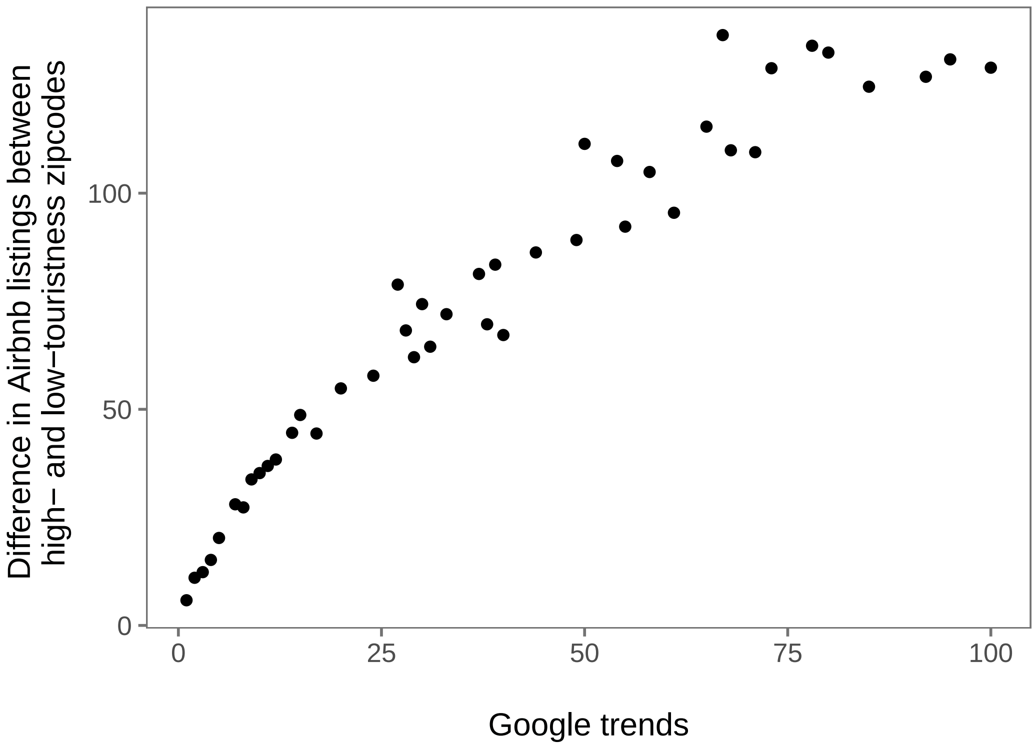
Note: The figure shows the spatial distribution of Airbnb listings in June 2011 and June 2016, where the number of listings is calculated using method 1 in Table 1. Listings are reported in logs, and log listings is set to zero if there are zero listings. Geographic areas without zipcode boundary information are colored white.

Figure 3: Total Number of Airbnb Listings (US, 2008-2016)



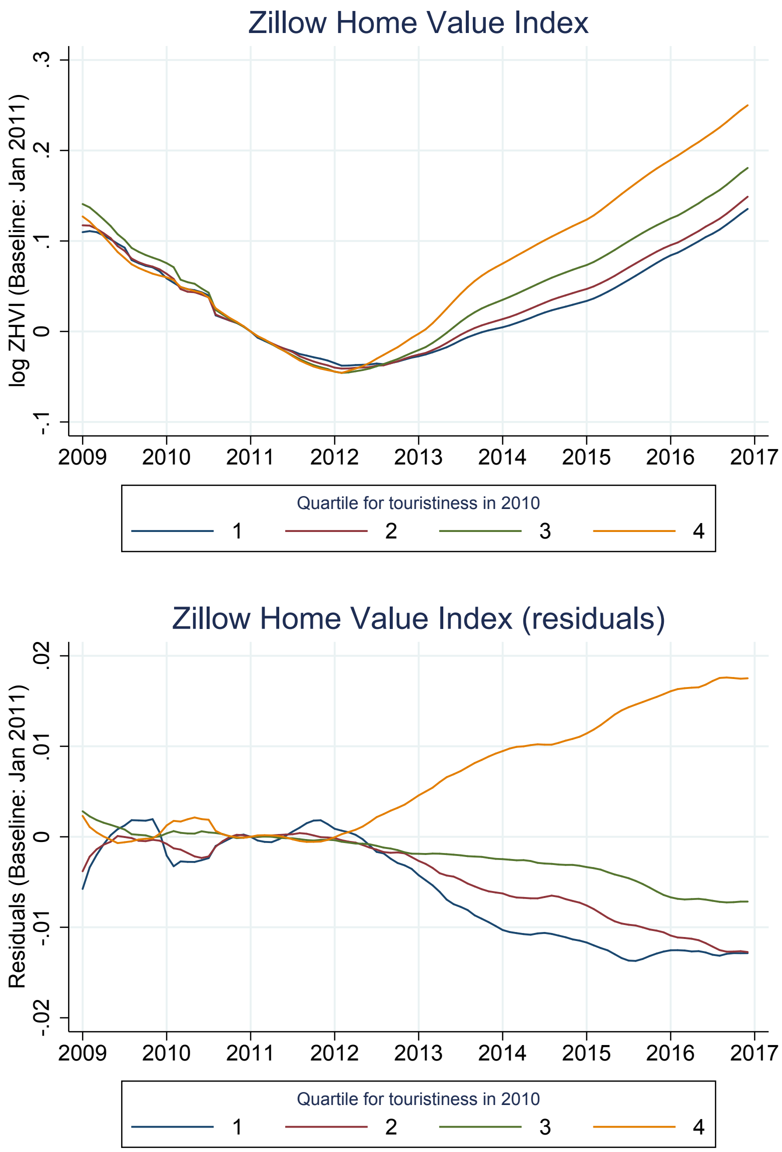
Note: This figure plots the number of Airbnb listings over time, using each of the 3 methods described in Table 1.

Figure 4: Testing the IV operating assumption



Note: This figure plots the difference in the number of Airbnb listings for high- and low-touristiness zipcode over the Google trend values. We use the sample median value of touristiness to create two equally sized groups of high- and low-touristiness zipcodes.

Figure 5: Trends in Zillow Home Value Index by "Tourstiness" of Zipcode



Note: The top panel plots the ZHVI index, normalized to January 2011=0, averaged within different groups of zipcodes based on their level of "touristiness" in 2010. Touristiness is measured as the number of establishments in the food services and accommodations sector (NAICS code 72) in 2010, and the zipcodes are separated into four equally sized groups. The bottom panel plots the residuals from a regression of the ZHVI on zipcode fixed effects and CBSA-month fixed effects.

Table 2: Size of Airbnb Relative to the Housing Stock (zipcodes, 100 largest CBSAs)

	p10	p25	p50	p75	p90
June 2011					
Airbnb Listings	0	0	0	2	7
Housing Units	1,058	2,813	7,437	12,829	18,037
Airbnb Listings as a Percentage of					
Total Housing Units	.00	.00	.00	.02	.09
Renter-occupied Units	.00	.00	.00	.06	.33
Vacant Units	.00	.00	.00	.20	.92
Vacant-for-rent Units	.00	.00	.00	1.01	5.06
June 2016					
Airbnb Listings	1	4	13	44	144
Housing Units	1,097	2,926	7,610	13,219	18,443
Airbnb Listings as a Percentage of					
Total Housing Units	.03	.08	.21	.60	1.88
Renter-occupied Units	.13	.33	.87	2.50	7.31
Vacant Units	.37	.99	2.63	7.19	20.00
Vacant-for-rent Units	1.72	4.65	13.70	42.80	129.00

Note: This table reports the size of Airbnb relative to the housing stock, by zipcodes for the 100 largest CBSAs as measured by 2010 population. The number of Airbnb listings is calculated using method 1 in Table 1. Data on housing stocks, occupancy characteristics, and vacancies come from ACS zipcode level 5-year estimates.

Table 3: IV Validity Check: Correlation Between Instrument and Rents/Prices in Zipcodes Without Airbnb

	(1) Dep var: ln ZRI	(2) Dep var: ln ZHVI	(3) Dep var: ln ZHVI/ZRI
$g_t \times h_{i,2010}$	−0.000	−0.000	−0.000
	(0.000)	(0.000)	(0.000)
ln Population	0.011	0.045***	0.032
	(0.013)	(0.016)	(0.020)
ln Median HH Income	−0.002	−0.001	0.004
	(0.011)	(0.016)	(0.020)
College Share	0.054*	0.120***	0.076
	(0.032)	(0.038)	(0.052)
Employment Rate	0.045	−0.017	−0.063
	(0.031)	(0.033)	(0.047)
Zipcode FE	Yes	Yes	Yes
CBSA-year-month FE	Yes	Yes	Yes
Observations	61854	50875	43164
R^2	0.979	0.994	0.964

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

Note: This table reports regression results when outcomes of interest are regressed on the instrumental variable directly, for zipcodes that were never observed to have any Airbnb listings. Because zipcode demographic characteristics are not available at a monthly frequency, zipcode-month measures for household income, population, college share, and employment rate are interpolated from the 2011 thru 2016 ACS 5-year estimates. Clustered standard errors at the zipcode level are reported in parenthesis.

Table 4: Comparing Airbnb and non-Airbnb zipcodes

	Airbnb Zipcodes	Non-Airbnb Zipcodes	Difference
Touristiness	43.73	7.40	36.33***
ln Median Income	11.02	10.87	0.14***
ln Population	9.47	8.25	1.21***
Share with bachelors' degree	0.35	0.20	0.15***
Employment rate	0.73	0.71	0.02***

Note: This table reports differences in demographic variables between zipcodes that were never observed to have any Airbnb listings and zipcodes that were.

Table 5: The Effect of Airbnb on Rental Rates

	(1)	(2)	(3)	(4)	(5)	(6)
ln Airbnb Listings	0.098***	0.008***	0.022***	0.021***	0.046***	0.043***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.003)
... × Owner-occupancy Rate (2010)			−0.023***	−0.022***	−0.038***	−0.035***
			(0.002)	(0.002)	(0.003)	(0.003)
ln Population				0.050***		0.042***
				(0.007)		(0.007)
ln Median HH Income				0.021***		0.017***
				(0.005)		(0.006)
College Share				0.063***		0.057***
				(0.013)		(0.013)
Employment Rate				0.048***		0.036***
				(0.014)		(0.014)
Zipcode FE	No	Yes	Yes	Yes	Yes	Yes
CBSA-year-month FE	No	Yes	Yes	Yes	Yes	Yes
Instrumental Variable	No	No	No	No	Yes	Yes
Observations	649841	649841	649841	649697	649841	649697
R^2	0.170	0.991	0.991	0.991	0.991	0.991
Kleibergen-Paap F Statistic					817.3	804.2

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

Note: The number of Airbnb listings is calculated using method 1 in Table 1. To avoid taking the log of a zero, one is added to the number of Airbnb listings before taking logs. The instrumental variables are $g_t \times h_{i,2010}$ and $g_t \times h_{i,2010} \times oorate_{ict}$. Because zipcode demographic characteristics are not available at a monthly frequency, zipcode-month measures for household income, population, college share, and employment rate are interpolated from the 2011 thru 2016 ACS 5-year estimates. Clustered standard errors at the zipcode level are reported in parenthesis.

Table 6: The Effect of Airbnb on House Prices

	(1)	(2)	(3)	(4)	(5)	(6)
ln Airbnb Listings	0.175***	0.009***	0.040***	0.038***	0.079***	0.076***
	(0.004)	(0.001)	(0.002)	(0.002)	(0.005)	(0.005)
... × Owner-occupancy Rate (2010)			−0.048***	−0.046***	−0.073***	−0.070***
			(0.003)	(0.003)	(0.006)	(0.006)
ln Population				0.078***		0.064***
				(0.010)		(0.010)
ln Median HH Income				0.012		0.005
				(0.008)		(0.008)
College Share				0.073***		0.061***
				(0.018)		(0.018)
Employment Rate				0.098***		0.070***
				(0.020)		(0.020)
Zipcode FE	No	Yes	Yes	Yes	Yes	Yes
CBSA-year-month FE	No	Yes	Yes	Yes	Yes	Yes
Instrumental Variable	No	No	No	No	Yes	Yes
Observations	572858	572858	572858	572805	572858	572805
R^2	0.188	0.996	0.996	0.996	0.996	0.996
Kleibergen-Paap F Statistic					660.7	645.4

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

Note: The number of Airbnb listings is calculated using method 1 in Table 1. To avoid taking the log of a zero, one is added to the number of Airbnb listings before taking logs. The instrumental variables are $g_t \times h_{i,2010}$ and $g_t \times h_{i,2010} \times oorate_{ict}$. Because zipcode demographic characteristics are not available at a monthly frequency, zipcode-month measures for household income, population, college share, and employment rate are interpolated from the 2011 thru 2016 ACS 5-year estimates. Clustered standard errors at the zipcode level are reported in parenthesis.

Table 7: The Effect of Airbnb on Price-to-Rent Ratio

	(1)	(2)	(3)	(4)	(5)	(6)
ln Airbnb Listings	0.077***	0.002**	0.016***	0.015***	0.032***	0.031***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.004)	(0.004)
... × Owner-occupancy Rate (2010)			−0.022***	−0.022***	−0.031***	−0.031***
			(0.003)	(0.003)	(0.005)	(0.005)
ln Population				0.030***		0.025**
				(0.010)		(0.010)
ln Median HH Income				−0.013		−0.016*
				(0.009)		(0.009)
College Share				0.011		0.006
				(0.019)		(0.019)
Employment Rate				0.046**		0.034
				(0.022)		(0.022)
Zipcode FE	No	Yes	Yes	Yes	Yes	Yes
CBSA-year-month FE	No	Yes	Yes	Yes	Yes	Yes
Instrumental Variable	No	No	No	No	Yes	Yes
Observations	537157	537142	537142	537089	537142	537089
R^2	0.154	0.979	0.979	0.979	0.979	0.979
Kleibergen-Paap F Statistic					627.7	614.7

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

Note: The number of Airbnb listings is calculated using method 1 in Table 1. To avoid taking the log of a zero, one is added to the number of Airbnb listings before taking logs. The instrumental variables are $g_t \times h_{i,2010}$ and $g_t \times h_{i,2010} \times oorate_{ict}$. Because zipcode demographic characteristics are not available at a monthly frequency, zipcode-month measures for household income, population, college share, and employment rate are interpolated from the 2011 thru 2016 ACS 5-year estimates. Clustered standard errors at the zipcode level are reported in parenthesis.

Table 8: Robustness Checks (Alternative Samples)

	Panel A Dep var: ln ZRI		Panel B Dep var: ln ZHVI		Panel C Dep var: ln ZHVI/ZRI	
	Coefficient:		Coefficient:		Coefficient:	
	airbnb	... $\times$ *oorate*	*airbnb*	... $\times$ *oorate*	*airbnb*	... $\times$ *oorate*
Sample:	(1)	(2)	(1)	(2)	(1)	(2)
Zipcodes: Near city center	0.030***	-0.022***	0.058***	-0.047***	0.028***	-0.024***
	(0.003)	(0.004)	(0.006)	(0.007)	(0.006)	(0.007)
Zipcodes: Far from city center	0.058***	-0.051***	0.097***	-0.095***	0.035***	-0.039***
	(0.005)	(0.005)	(0.008)	(0.009)	(0.006)	(0.007)
Years: 2011-2013	0.034***	-0.003	0.046***	-0.003	0.005	0.011*
	(0.003)	(0.005)	(0.004)	(0.006)	(0.004)	(0.006)
Years: 2014-2016	0.032***	-0.033***	0.088***	-0.126***	0.061***	-0.094***
	(0.006)	(0.006)	(0.009)	(0.010)	(0.009)	(0.010)
CBSAs: pop. rank 1-30	0.054***	-0.041***	0.096***	-0.083***	0.040***	-0.039***
	(0.004)	(0.004)	(0.007)	(0.007)	(0.005)	(0.005)
CBSAs: pop. rank 31-100	0.022***	-0.016***	0.031***	-0.025***	0.009	-0.004
	(0.003)	(0.004)	(0.006)	(0.008)	(0.006)	(0.008)

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

Notes: This table repeats the regressions reported in column 6 of Tables 5-7, performed separately on different subsamples. “Near to city center” is the sample of zipcodes that that are below the median distance to CBD, where the median is taken within CBSAs. “Far from city center” is the sample zipcodes that are above the median distance to CBD. City center coordinates are recovered using the Microsoft Bing API, and zipcode centroid coordinates are from the U.S. Census Bureau.

Table 9: Robustness Check (log-density specification)

	Panel A Dep var: ln ZRI		Panel B Dep var: ln ZHVI		Panel C Dep var: ln ZHVI/ZRI	
	(1)	(2)	(1)	(2)	(1)	(2)
Airbnb Density	0.913***	1.571***	1.843***	2.679***	0.976***	1.075***
	(0.135)	(0.182)	(0.225)	(0.318)	(0.189)	(0.267)
... × Owner-occupancy Rate (2010)	−1.223***	−2.609***	−3.063***	−3.608***	−1.942***	−1.754***
	(0.209)	(0.555)	(0.340)	(0.893)	(0.308)	(0.675)
ln Population	0.052***	0.044***	0.066***	0.069***	0.018*	0.022*
	(0.007)	(0.009)	(0.010)	(0.013)	(0.010)	(0.012)
ln Median HH Income	0.015***	0.010*	0.004	−0.005	−0.013	−0.016*
	(0.006)	(0.006)	(0.008)	(0.009)	(0.009)	(0.009)
College Share	0.058***	0.058***	0.053***	0.042**	0.004	−0.000
	(0.013)	(0.015)	(0.018)	(0.019)	(0.018)	(0.019)
Employment Rate	0.046***	0.047***	0.103***	0.089***	0.051**	0.045**
	(0.014)	(0.015)	(0.019)	(0.021)	(0.021)	(0.022)
Zipcode FE	Yes	Yes	Yes	Yes	Yes	Yes
CBSA-year-month FE	Yes	Yes	Yes	Yes	Yes	Yes
Instrumental Variable	No	Yes	No	Yes	No	Yes
Observations	613245	613245	538990	538990	504260	504260
R^2	0.991	0.991	0.996	0.996	0.979	0.979
Kleibergen-Paap F Statistic		9.954		10.92		10.54

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

Note: The number of Airbnb listings is calculated using method 1 in Table 1. Instruments in column 2 are interacted second order polynomials of g_t, $h_{i,2010}$, and $oorate_{i,2010}$. Because zipcode demographic characteristics are not available at a monthly frequency, zipcode-month measures for household income, population, college share, and employment rate are interpolated from the 2011 thru 2016 ACS 5-year estimates. Clustered standard errors at the zipcode level are reported in parenthesis.

Table 10: Effect Magnitudes for 10 Largest CBSAs

	Year-over-Year Airbnb Contribution		Year-over-Year Growth	
CBSA	Rent	Price	Rent	Price
Top 100 CBSAs	0.59%	0.82%	3.18%	5.70%
New York-Newark-Jersey City, NY-NJ-PA	0.60%	0.83%	3.64%	3.55%
Los Angeles-Long Beach-Anaheim, CA	1.14%	1.79%	4.92%	9.66%
Chicago-Naperville-Elgin, IL-IN-WI	0.34%	0.44%	2.25%	3.98%
Dallas-Fort Worth-Arlington, TX	0.70%	1.01%	4.18%	8.21%
Miami-Fort Lauderdale-West Palm Beach, FL	1.02%	1.51%	4.51%	11.72%
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.54%	0.73%	1.94%	2.05%
Houston-The Woodlands-Sugar Land, TX	0.95%	1.37%	4.67%	8.34%
Washington-Arlington-Alexandria, DC-VA-MD-WV	0.70%	0.96%	1.28%	4.41%
Atlanta-Sandy Springs-Roswell, GA	0.75%	1.07%	3.11%	8.42%
Detroit-Warren-Dearborn, MI	0.16%	0.21%	2.41%	8.54%

Note: Airbnb contribution is calculated as $\hat{\beta} + \hat{\gamma} oorate_{ic,2010}$ multiplied by the median year-over-year growth in log Airbnb listings for each zipcode, and then taken at the median zipcode. Estimates from columns 6 of Tables 5 and 6 are used.

Table 11: The Effect of Airbnb on Vacancy Rates

	(1) All Vacant Units	(2) Seasonal Homes	(3) Vacant-for-Rent	(4) Vacant-for-Sale
ln Airbnb Listings	0.001	0.008**	−0.005***	−0.002***
	(0.004)	(0.003)	(0.001)	(0.001)
Zipcode FE	Yes	Yes	Yes	Yes
CBSA-year-month FE	Yes	Yes	Yes	Yes
Observations	600	600	600	600
R^2	0.929	0.923	0.841	0.722

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

Note: Vacancy rate is regressed on the log number of Airbnb listings at the CBSA-year level. The number of Airbnb listings is calculated using method 1 in Table 1. To avoid taking the log of a zero, one is added to the number of Airbnb listings before taking logs. The dependent variable is the number of vacant units divided by the total number of housing units. Data on vacancies comes from annual ACS 1-year estimates. Seasonal homes are housing units described as being for seasonal, recreational, or occasional use. Note that according to Census methodology, housing units occupied temporarily by persons who usually live elsewhere are classified as vacant units.

For Online Publication: Appendix

A Model with Endogenous Owner-Occupiers

The model in Section 2 can be extended to allow the share of owner-occupiers to be endogenous. However, ex-ante heterogeneity in potential buyers needs to be introduced or else an equilibrium with all three of renters, owner-occupiers, and absentee landlords would require that Equations (4) and (10) both be equal. If they were not, then either long-term residents will outbid absentee landlords to own all the housing, or the opposite will happen.

We introduce heterogeneity in the most parsimonious way possible. Consider a set of N individuals who potentially interact with a local housing market. Each individual can choose to be a renter, an owner-occupier, an absentee landlord, or none of the above. Let us normalize the utility for "none of the above" to zero. The present value of utility that person i gets from being a renter is:

$$\begin{aligned} u_{i,r} &= U - \frac{1}{1-\delta} R + \epsilon_{i,r} \\ &= u_r + \epsilon_{i,r} \end{aligned}$$

Here, U is the present value of amenities that the individual gets from being a resident in this market. $\frac{1}{1-\delta} R$ is the present value of rents. $\epsilon_{i,r}$ is an idiosyncratic utility shock which is known ex-ante. The present value that person i gets from being an owner is:

$$\begin{aligned} u_{i,o} &= U - P + \frac{1}{1-\delta} \gamma g(Q-c) + \epsilon_{i,o} \\ &= u_o + \epsilon_{i,o} \end{aligned}$$

Here, U is again the present value of amenities, P is the purchase price of housing, and $\frac{1}{1-\delta} \gamma g(Q-c)$ is the present value of rents received from selling excess capacity on the peer-to-peer market. Finally, the present value that person i gets from being an absentee landlord is:

$$
\begin{aligned}
u_{i,a} &= -P + \frac{1}{1-\delta}\left[R + g(Q - R - c)\right] + \epsilon_{i,a} \\
&= u_a + \epsilon_{i,a}
\end{aligned}
$$

For analytical tractability, let the utility shocks ϵ_i be distributed i.i.d. type 1 extreme value. The share of individuals that choose option j out of $j = \{r, o, a\}$ is:

$$
s_j = \frac{\exp u_j}{1 + \sum_{k \in \{r,o,a\}} \exp u_k}
$$

The equilibrium conditions determining R and P are:

$$
(s_a + s_o)N = H
$$

and:

$$
\left[1 - f(Q - R - c)\right] s_a N = s_r N
$$

The first condition is the market clearing condition for the housing market as a whole; i.e. the number of absentee landlords plus owner-occupiers is equal to the housing stock. The second condition is the market clearing condition for the long-term rental market; i.e. the number of renters is equal to the number of absentee landlords allocating housing to the long-term market.

We leave the derivation of analytical results for this model to future work or enterprising students. For this Appendix, we will simply present some numerical results which are consistent with all the key predictions in Section 2. Choosing $N = 10$, $H = 2$, $U = \$500,000$, $\delta = 0.95$, $\gamma = 0.1$, $Q = \$25,000$, and letting the distribution of idiosyncratic costs to listing in the short-term market be uniform from \$0 to \$100,000, we consider a change of c from ∞ (no home-sharing) to $c = 0$ (costless home-sharing). Table 12 below shows the results. Consistent with the model, the introduction of home-sharing under these model parameters results in a modest increase in both rental rates and house prices, and the increase in house prices is larger than the increase in rental rate. The qualitative results are robust to different parameter choices.

Table 12: Simulation Results

	$c = \infty$	$c = \$50k$	Δ
Rent	\$25,069	\$25,193	0.49%
Price	\$502,773	\$507,702	0.98%

B Additional Robustness Checks

Alternative measures of Airbnb supply

In this section, we perform a number of additional robustness checks. First, we show that our main results are robust to the alternative methods of calculating Airbnb supply, as discussed in Section 3. Rows 1 and 2 of Table 13 report the regression results when methods 2 and 3 are used to measure Airbnb supply instead of method 1. The results are barely changed, which is not surprising given the high correlation between the three measures, despite level differences.

Alternative CBSA sample

Second, we show that our main results are robust to the inclusion of smaller cities, beyond the 100 largest CBSAs. In rows 3 and 4 of Table 13, we report regression results when the sample includes the top 150 CBSAs and the top 200 CBSAs. Again, the results are not much changed, suggesting that the inclusion of smaller cities will not drive the results downwards significantly.

Excluding observations with zero or a small number of listings

Finally, one issue with the log-log specification is that we take the log of one plus the number of listings to avoid taking logs of zero. We now show that the results are robust to this choice. Row 5 of Table 13 reports regression results when instead of adding 1 to the number of listings, we instead simply drop all zipcode-month observations in which the number of listings is zero. The RHS variable is therefore log(#listings) instead of log(1+#listings). Row

6 additionally drops all zipcode-month observations in which the number of listings is less than 5. The results remain qualitatively and quantitatively similar under this alternative choice.

Using contemporaneous owner-occupancy rate

As described in Section 4, we interact $Airbnb_{ict}$ with $oorate_{ic,2010}$, the owner-occupancy rate in 2010, to reduce endogeneity concerns. However, the results are robust to using the contemporaneous owner-occupancy rate, $oorate_{ict}$.[32] Row 7 of Table 13 reports the results when we use contemporaneous owner-occupancy rate.

C 2SLS Results using Airbnb Density

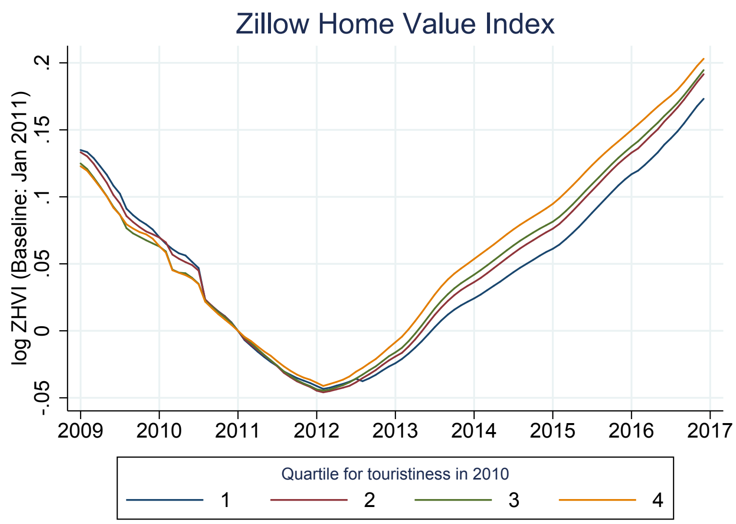
In this section, we report some 2SLS results using various choices of instruments for the log-density specification to show that the qualitative results are robust this choice. However, as we shall show, the magnitudes are somewhat sensitive. As noted in the main text, using $g_t \times h_{i,2010}$ as the instrument results in underidentification. In practice, we find that using $g_t \times h_{i,2010}/stock_{i,2010}$ as the instrument, where $stock_{i,2010}$ is the total housing stock in 2010, gives reasonable results. Figure 6 and Table 14 repeat the IV validity support exercises discussed in Section 4.1 for this instrument. Alternatively, higher order polynomials of the instrument (without dividing by $stock_{i,2010}$) appear to work as well, though the estimates are quite sensitive to the specific choice of instruments.

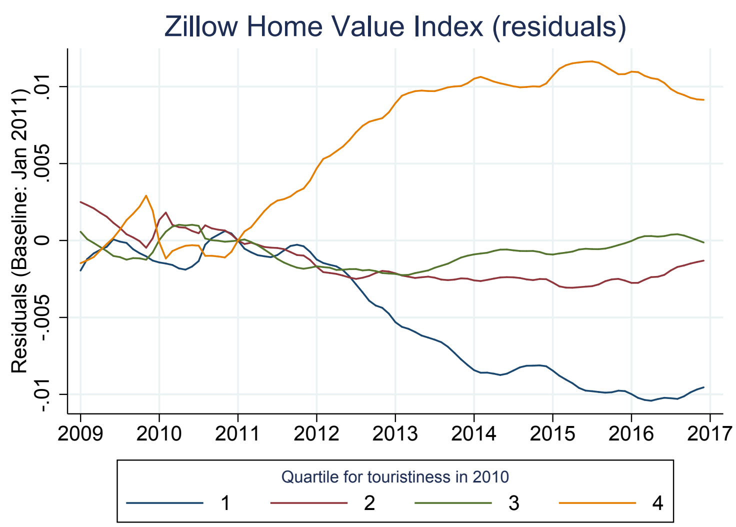
We report results for three 2SLS regression using different sets of instruments in Table 15. In columns (1) of each panel, the instruments are $g_t \times h_{i,2010}/stock_{i,2010}$ interacted with $oorate_{i,2010}$. In columns (2), the instruments are a third order polynomial of $g_t \times h_{i,2010}$ interacted with $oorate_{i,2010}$. In columns (3), the instruments are the full interactions between second order polynomials of g_t, $h_{i,2010}$, and $oorate_{i,2010}$. The general qualitative result is

[32] Contemporaneous owner-occupancy rate is interpolated to the monthly level using ACS 5-year estimates from 2011 to 2016.

that the direct effect of Airbnb density is positive, while the interaction with owner-occupancy rate is negative, consistent with the results using the log-log specification.

Figure 6: Trends in Zillow Home Value Index by $h_{i,2010}/stock_{i,2010}$





Note: The top panel plots the ZHVI index, normalized to January 2011=0, averaged within different groups of zipcodes based on $h_{i,2010}/stock_{i,2010}$, i.e. the number of establishments in food services and accommodations sector in 2010 divided by the housing stock in 2010. The zipcodes are then separated into four equally sized groups. The bottom panel plots the residuals from a regression of the ZHVI on zipcode fixed effects and CBSA-month fixed effects.

Table 13: Additional Robustness Checks

	Panel A Dep var: ln ZRI		Panel B Dep var: ln ZHVI		Panel C Dep var: ln ZHVI/ZRI	
	Coefficient:		Coefficient:		Coefficient:	
	airbnb	... $\times$ *oorate*	*airbnb*	... $\times$ *oorate*	*airbnb*	... $\times$ *oorate*
Robustness Check:	(1)	(2)	(1)	(2)	(1)	(2)
Method 2 for calculating # listings	0.048***	-0.040***	0.087***	-0.082***	0.036***	-0.037***
	(0.003)	(0.004)	(0.006)	(0.008)	(0.005)	(0.006)
Method 3 for calculating # listings	0.048***	-0.041***	0.087***	-0.083***	0.036***	-0.037***
	(0.003)	(0.004)	(0.006)	(0.008)	(0.005)	(0.006)
CBSAs pop. rank 1-150	0.040***	-0.033***	0.071***	-0.067***	0.030***	-0.031***
	(0.003)	(0.003)	(0.005)	(0.006)	(0.004)	(0.005)
CBSAs pop. rank 1-200	0.038***	-0.031***	0.067***	-0.065***	0.027***	-0.030***
	(0.002)	(0.003)	(0.004)	(0.006)	(0.003)	(0.004)
Drop obs. with zero listings	0.048***	-0.041***	0.092***	-0.084***	0.042***	-0.039***
	(0.006)	(0.004)	(0.010)	(0.007)	(0.009)	(0.005)
Drop obs. with <5 listings	0.034**	-0.043***	0.081***	-0.096***	0.046**	-0.049***
	(0.014)	(0.005)	(0.022)	(0.009)	(0.020)	(0.008)
Contemporaneous owner-occ rate	0.042***	-0.035***	0.074***	-0.070***	0.030***	-0.031***
	(0.003)	(0.003)	(0.005)	(0.006)	(0.004)	(0.005)

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

Notes: This table reports results from robustness checks described in Appendix section B. In each case, 2SLS results are reported where the instrument is $g_t \times h_{i,2010}$. 1 is added before taking the log of the number of listings, except in rows 5 and 6 where the log(#listings) is taken directly.

Table 14: IV Validity Check for $g_t \times h_{i,2010}/stock_{i,2010}$

	(1) Dep var: ln ZRI	(2) Dep var: ln ZHVI	(3) Dep var: ln ZHVI/ZRI
$g_t \times h_{i,2010}/stock_{i,2010}$	0.007	0.013	−0.004
	(0.012)	(0.011)	(0.014)
ln Population	0.011	0.045***	0.032
	(0.013)	(0.016)	(0.020)
ln Median HH Income	−0.002	−0.001	0.004
	(0.011)	(0.016)	(0.020)
College Share	0.054*	0.120***	0.077
	(0.032)	(0.038)	(0.051)
Employment Rate	0.046	−0.016	−0.063
	(0.031)	(0.033)	(0.047)
Zipcode FE	Yes	Yes	Yes
CBSA-year-month FE	Yes	Yes	Yes
Observations	61854	50875	43164
R^2	0.979	0.994	0.964

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

Note: This table reports regression results when outcomes of interest are regressed on the instrumental variable directly, for zipcodes that were never observed to have any Airbnb listings. Because zipcode demographic characteristics are not available at a monthly frequency, zipcode-month measures for household income, population, college share, and employment rate are interpolated from the 2011 thru 2016 ACS 5-year estimates. Clustered standard errors at the zipcode level are reported in parenthesis.

Table 15: 2SLS Results for Log-Density Specification

	Panel A Dep var: ln ZRI			Panel B Dep var: ln ZHVI			Panel C Dep var: ln ZHVI/ZRI		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Airbnb Density	1.002***	1.888***	1.571***	2.484***	3.601***	2.679***	1.476***	1.677***	1.075***
	(0.215)	(0.215)	(0.182)	(0.304)	(0.367)	(0.318)	(0.323)	(0.279)	(0.267)
... × Owner-occupancy Rate (2010)	−1.102*	−3.976***	−2.609***	−3.781***	−6.829***	−3.608***	−2.969***	−3.173***	−1.754***
	(0.594)	(0.709)	(0.555)	(0.874)	(1.138)	(0.893)	(0.966)	(0.731)	(0.675)
ln Population	0.054***	0.033***	0.044***	0.064***	0.040**	0.069***	0.011	0.011	0.022*
	(0.008)	(0.012)	(0.009)	(0.013)	(0.017)	(0.013)	(0.015)	(0.013)	(0.012)
ln Median HH Income	0.014**	0.008	0.010*	−0.002	−0.005	−0.005	−0.016*	−0.018**	−0.016*
	(0.006)	(0.008)	(0.006)	(0.008)	(0.010)	(0.009)	(0.010)	(0.009)	(0.009)
College Share	0.053***	0.070***	0.058***	0.046**	0.040*	0.042**	0.002	−0.001	−0.000
	(0.015)	(0.019)	(0.015)	(0.018)	(0.021)	(0.019)	(0.019)	(0.019)	(0.019)
Employment Rate	0.044***	0.053***	0.047***	0.097***	0.113***	0.089***	0.052**	0.050**	0.045**
	(0.015)	(0.016)	(0.015)	(0.021)	(0.023)	(0.021)	(0.023)	(0.022)	(0.022)
Zipcode FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
CBSA-year-month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Instrumental Variable	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	613245	613245	613245	538990	538990	538990	504260	504260	504260
R^2	0.991	0.990	0.991	0.996	0.996	0.996	0.979	0.979	0.979
Kleibergen-Paap F Statistic	15.32	5.880	9.954	10.83	5.877	10.92	9.418	5.661	10.54

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

Note: This table reports 2SLS results using the log-density specification, for various choices of instrumental variables. In columns (1), the instruments are $g_t \times h_{i,2010}/stock_{i,2010}$ and the interaction with $oorate_{i,2010}$. In columns (2), the instruments are a third order polynomial of $g_t \times h_{i,2010}$ interacted with $oorate_{i,2010}$. In columns (3), the instruments are fully interacted second order polynomials of g_t, $h_{i,2010}$, and $oorate_{i,2010}$.