Data Open 2021

An Empirical Analysis On Airbnb influences in U.S. South Side Region

Team 7
Jenny Chen | Tess Tao | Charles Xu | Ray Xu

General Background

Airbnb, a peer-to-peer marketplace for short-term rentals, was founded in 2008 and experienced dramatic growth since then. Airbnb's rapid growth **prompted concerns and heated discussion about its impacts on cities and urban housing markets**.

Concern over the impact of home-sharing on housing affordability has garnered significant attention from policymakers and has motivated many cities to impose stricter regulations on home-sharing. However, past studies rarely focus on the U.S South, where Airbnb's potential impact usually goes understudied compared to more developed areas like New York City. Thus, we are motivated to **explore and validate the actual effect of Airbnb activities on the housing price of southern cities**. Furthermore, we are interested in comparing southern cities with other areas in order to get a better understanding of the **underlying mechanism as well as recommendations for future policy**.

Key Findings

Inspired by the ongoing debate around house-sharing platforms like Airbnb and their influence on the housing market, we conducted a **two-stage least squares regression** analysis as an effort to establish **causality between the number of Airbnb listings and the median housing price** in four major US southern cities (including Asheville, Austin, Nashville, and New Orleans), and Los Angeles.

Our results suggest that while Airbnb does expand the short-term rental market by giving hosts the power to flexibly rent out their homes across the year, it might do so through prompting some homeowners in the long-term rental market to switch to short-term home sharing, thus draining long-term supply in local housing market and pushing up the median house price in the coming years. Specifically, by constructing **LASSO-chosen instrumental variables** and running a two-stage least squares model, we arrive at the following insights concerning Airbnb rental activities in the South:

Observation 1: Overall, the results suggest that increase in Airbnb listings would lead to substantial increase in the median house price for a city with median occupancy rate. When we only examine southern cities, this effect decreases in its magnitude, with **a 1% increase in Airbnb listings leading to a 0.176% increase in house prices** for a southern city like New Orleans.

Observation 2: The effect of Airbnb listings on housing prices is decreasing in a city's home occupancy rate, **consistent with the hypothesis that Airbnb causes some homeowners to transfer from long term rental markets to short term home sharing**. Yet since such reallocation behaviors are more likely to take place in houses not physically occupied by their owners, this effect is more slender in cities with higher home occupancy rates.

Observation 3: When we are constructing instruments, we use Google Trends search index to interact with the number of establishments of food services and accommodation as a proxy for local touristiness of a city. As the first stage regression proves the relevance of our instrumenting approach, we conclude that **surge in Airbnb listings also drives local demand for space through increase in tourism.**

Observation 4: The effect also decreases when we only focus on southern cities in our dataset. This confirms our hypothesis that while increase in Airbnb activities contributes partially to the rise of house prices in southern regions, the effect is not as overwhelming as in other cities where Airbnb listings take up a higher proportion of the housing market.

Observation 5: Finally, while we initially hypothesize that Airbnb might reduce neighborhood qualities by creating safety concerns, noises, etc., which might have a negative effect on property values, our results suggest that the **negative externalities of Airbnb rentals are not large enough to overwhelm its positive influences on housing prices**.

Recommendation

Guided by our analysis, we propose a few policy recommendations for southern cities to i**mplement in response to the potential effect of Airbnb rental activities on local housing market**:

Recommendation 1: Local authorities should enforce an **annual hard cap on short-term, non-owner-occupied rentals** in areas where short term rentals are draining usual housing supply (for instance, San Francisco has previously passed a 60-day annual cap on such short-term rentals).

Recommendation 2: Airbnb should **encourage more owner-occupied rentals** on the platform, which could expand the home sharing market while not taking up too much space from long-term rental markets.

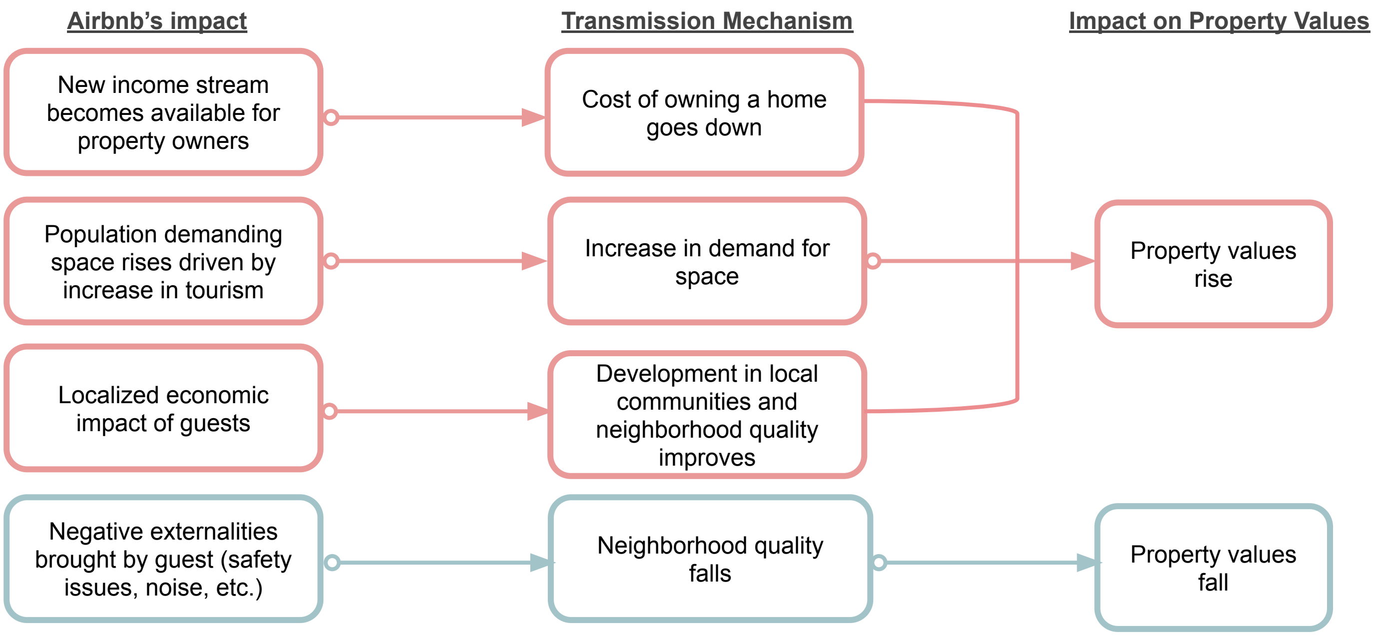

Figure 1: Transmission Mechanisms for the Impact of Airbnb Activity on Property Values

Background

Airbnb, a peer-to-peer marketplace that facilitates the matching between demand and supply of accommodations, has dramatically disrupted the housing market since its inception. As of 2020, the platform has over 7 million listings worldwide, with the United States being the most active listing country and taking up 900 thousand listings.

While the introduction of Airbnb should have **improved economic efficiency** since it helps utilize the spare spaces that would otherwise be underutilized, its influence on traditional housing markets has grown into a significant cause for concern, particularly when looking at its **impacts on local property value and communities**.

Our story

While past work has posed the underlying transmission mechanism for how Airbnb rental activities can affect property values, there currently lacks literature investigating the **causality and net influence of Airbnb rental activities on property values from a quantitative approach**.

Although there have been researches conducted to investigate the effects of Airbnb rental activities on property values in more developed regions, it is crucial for us to investigate the secondary impacts of Airbnb rental activities in the out-of-the-spotlight southern regions in the U.S. We picked Asheville, Austin, Nashville, and New Orleans as the representative southern cities per the given dataset. We also **incorporated Los Angeles into our model to serve as a comparison** against our cities of interest.

Analytical Approach Summary

In terms of our modelling approaches, we built on the **two-stage least squares approach** that had been implemented by current literatures and include **LASSO-chosen instrumental variables** in our final regression model.

Since the exogeneity of instrumental variables is hardly testable, using LASSO should improve the strength of our instruments and ensure that our tests have enough power to **validate the proposed causality between our variables of interest**.

We then ran the regression model on two datasets to distinguish the effect on southern cities from combined areas.

Finally, an additional **placebo test** was conducted in the end to prove that the observed effect is **not a consequence of development along the time period**. Framed in this way, our analysis of the effect of Airbnb on the housing market in southern cities should hold adequate statistical power.

New Orleans | Asheville | Los Angeles

Figure 2: Comparison in Airbnb Listings Density between 2013 and 2020

Exploratory Data Analysis

We utilized given datasets and sourced a handful of external data to perform preliminary graphical analysis.

Datasets Descriptions & Data Cleaning

To perform our city-level analysis, we need time-series data on multiple socio-economic indicators for the cities of interest, including Asheville, Austin, Nashville, and New Orleans. The given datasets provide listing data from 2016 to 2018 but as we are trying to investigate the **impact of Airbnb on local communities since its inception**, we sourced the internet and retrieved the more comprehensive data from Inside Airbnb.

Once we have collected the data, the next step is to **define a measure of Airbnb supply at a given time**. The Inside Airbnb dataset provides information such as a listing's review information, geographical information, and various host information. While a common method computes a listing's entry date as the date its host registered on Airbnb and assumes that listings never exit, we believe that the **calculation is inherently flawed and overestimate the supply**. Instead, we used a **listing's first and last review date as its active time** because it represents a more accurate estimation of active listing.

We also included city-level data on GDP, unemployment rate, personal income, population retrieved from FRED and Census Bureau. Other than traditional data cleaning methods such as removing missing values, we paid special attention to the **consistency of time granularity across datasets**. The raw data for GDP, personal income, and population are on a yearly basis and after examining the growth trend for each variable, we found that they all followed a linear growth. We then interpolated the data to monthly level in order to align with the listing data we have.

For variables including Housing Price Index, Airbnb listings, GDP, personal income, and population, we chose to **take log of these values to focus on the percentage change** rather than the absolute value in order to improve the interpretability and consistency of our results.

EDA (continued) & Modeling

Fixed Effects
OLS

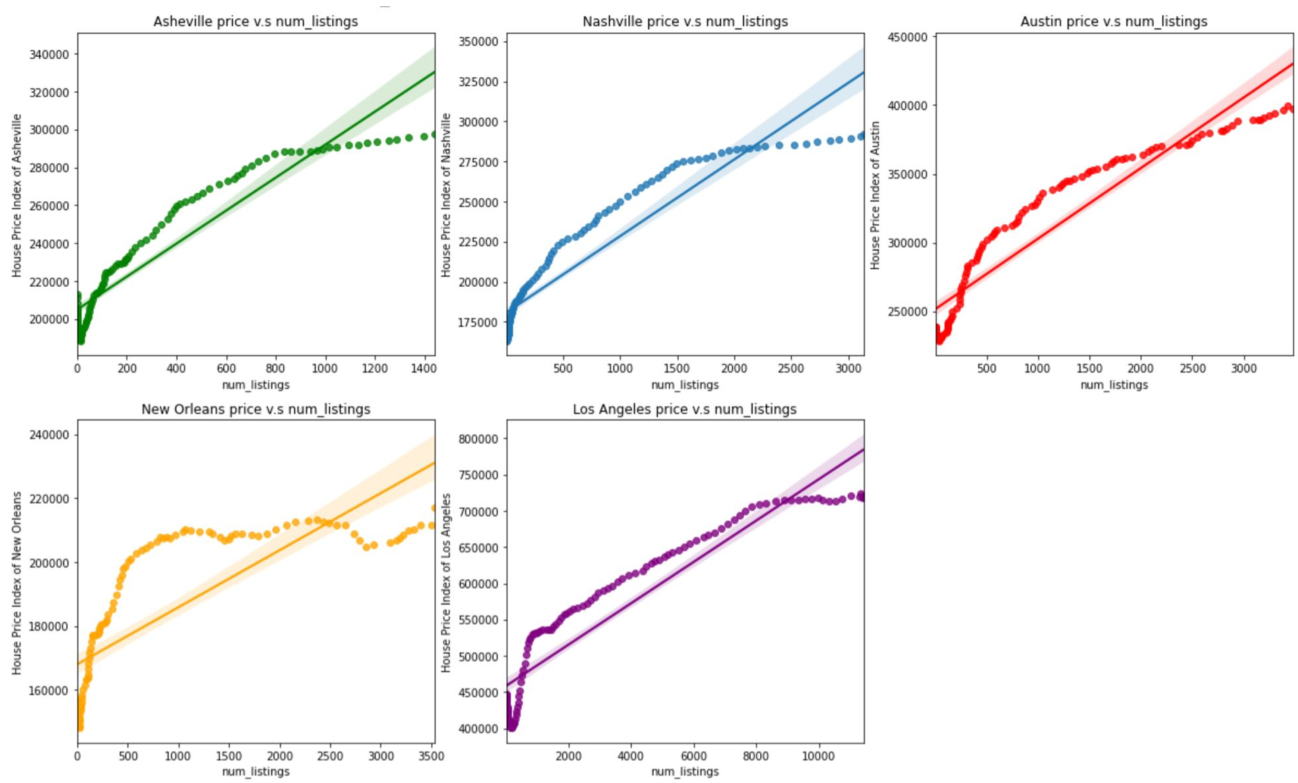

Figure 3: Simple Regression of Housing Price Index on Num_Listings for 5 Cities

Looking at Figure 2, despite the difference in local development and demographics, we still saw **similarity in the growth trend of Airbnb** for the cities of interest between 2013 and 2020. Furthermore, all five of them displayed a **sparse geographical distribution for listings** with LA having the widest reach across the board all over the city. This leads to our hypothesis that **Airbnb listing activities may have similar city-level impacts on property values for both southern regions and non-southern regions**.

The Figure 3 shows simple regression of housing price index on airbnb listings for all cities. At first sight, we plotted the raw data of housing prices against the number of Airbnb listings for each city. We can clearly observe a **similar pattern of positive correlation between the housing price and the numbers of Airbnb listings across the selected cities**. This leads to our hypothesis that airbnb listing activities brings more positive effects to the property value than negative sides, and further validates our earlier-posed hypothesis about the similar impacts that Airbnb listings have on property values across cities.

Modelling - Fixed Effects OLS

After identifying the control groups and the subject of interests, we now employ comparative case study methods to examine our hypothesis and infer the **causality between the patterns of airbnb rental activity and the property values in the U.S. southern area.**

Among the existing methodologies employed by past literatures, the most intuitive model to study the effect of Airbnb activities is the **fixed effects OLS** model. Multiple regression analysis estimated by OLS can provide us a way to mitigate the problem of omitted variable bias by including additional regressors other than the main variable that we would like to explore. Using OLS, we can calculate the amount of variance in the property price that is accounted for by the variation in each of the independent variables that we choose. Then, the size and statistical significance for each estimated coefficient will be determined.

Let Y_{it} be either the price index or the rent index for city i in time t (weekly). We chose the to be a measure of Airbnb supply whi $Num\_listings_{it}$ nain independent variable. Then, we chose other factor that could possibly be omitted variable which are occupancy rates, GDP factor, Personal Income, Population, Unemployment Rate. In order for the time and entity fixed effects, the model will equivalently be represented by 5-1=4 entity binary indicators and 11-1 = 10 time binary indicators to avoid the issue of multicollinearity. Then, since all variables are not on the same scale, we will use the log-log model instead. Therefore, we assume the following causal relationship between house price and Airbnb listings:

$$
\begin{aligned}
ln(Y_{it}) = \beta_0 + \beta_1 ln(Num\_listings_{it}) + \beta_2 ln(Num\_listings_{it} \times Occupancy_{it}) \\
+ \beta_3 ln(PI_{it}) + \beta_4 ln(Pop_{it}) + \beta_5 Unemploy_{it} + \Upsilon_i + \delta_t + \varepsilon_{it}
\end{aligned}
$$

where unemployment, ϒi, is the entity fixed effects, and δt is the time fixed effects for each year. Since we have cleaned data for four southern cities and LA, we ran regression twice, with LA data for the first time and without LA data for the second to see the difference. Referring to the FE column of Table 1, we can find out that a **1% increase in the number of Airbnb listings would lead to a consequent 0.12% increase in the housing price index** when we examine the overall dataset. In general, it validates our hypothesis that there is a **positive relationship between the increasing numbers of rental activities and the property price**. From the statistical inference standpoint, the p-value of approximately zero with a F-statistics of 132 should provide strong support of our conclusion that effect is significant. From the plot which contains the dependent variable and fitted values with confidence intervals vs. the independent variable chosen, we can say that the regression is fairly accurate and the model assumption is valid.

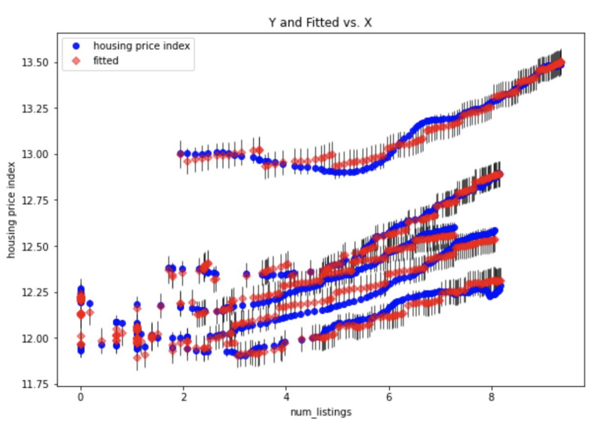

Figure 4: Fitted Variables with confidence intervals

Justification for Using IV

Initial results from our previous fixed effects OLS already suggest a strong positive correlation between the number of Airbnb listings and the housing price index in any city. Specifically, when all other demographic control variables are held constant, the results show that a 1% increase in the number of Airbnb listings would lead to a consequent 0.12% increase in the housing price index. **However, this methodology still has potential limitations concerning the issue of endogeneity, which justify our use of proper instrumental variables for our variables of interest.**

Even after controlling for unobserved factors at the city and year level, there could still be some unobserved time-changing, city-specific factors in the error term that could be correlated with num_listingsit. If this was the case, our estimate of the coefficient on num_listingsit will be seriously biased by the unseen correlation between them. As can be shown in the Figure 5, **simultaneous causality seems to exist between our variable of interest,** number of Airbnb listings and the Zillow HPI when we conduct simple linear regression both ways on Los Angeles data. To address this issue, a valid instrument needs to be constructed that is uncorrelated with specific shocks to the local housing market at the city level, but **only affects housing price through its impact on the number of airbnb listings**. In this way, the direct effect of airbnb listings on housing prices could be effectively isolated.

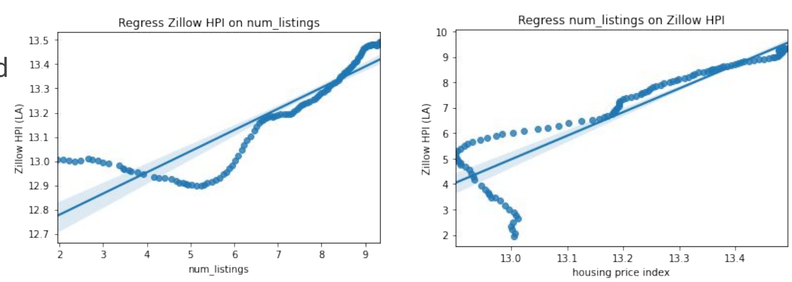

Figure 5: Simultaneous causality

Our choice of candidate instruments would include the worldwide Google Trends search index for the word “airbnb”, which is a standardized measure of the quantity of Google searches for “airbnb” in any given year-month. Moreover, we would interact the Google Trends term with an additional measurement of the density of food services and accommodation establishments in every city (standardized by local population).

From a theoretical perspective, **the aggregate Google Trends, as it is uncorrelated with local demographic and economic conditions, will only affect spatial units along some systematic cross-sectional variable,** in our case the density of food services and accommodation, which supposedly represents how attractive a city is to tourists in 2009 before Airbnb trend started to take effect. This approach of using the interaction of an exogenous time serious with a potentially cross-sectional exposure variable could be validated by two main justifications:

Justification 1:
Our first assumption in the instrument's power is that in response to the climbing popularity of Airbnb worldwide, home owners are more likely to rent their property in the short-term market. This assumption is supported by Figure 6, which shows the relationship between Google Trends search index for "Airbnb" and the difference in the number of Airbnb listings between high- and low-touristiness areas. We use the median number of establishments density as the standard to split the cities into two groups of touristiness, and the result shows that **such difference increases as the Google Trends index increases over time**, thus confirming our assumption.

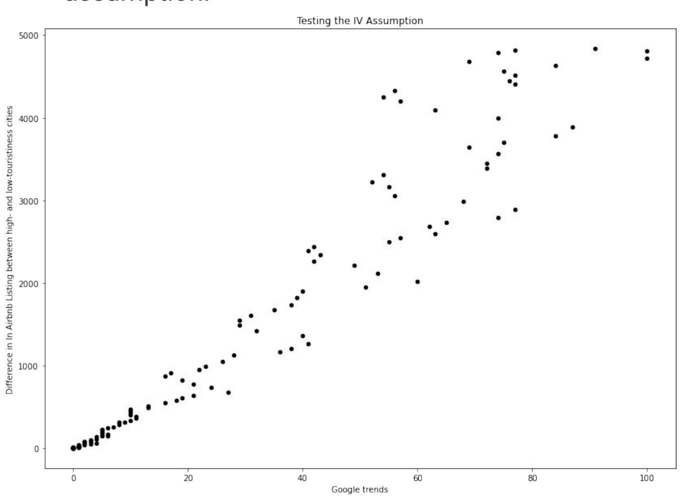

Figure 6: Testing the IV Assumption

Justification 2:
Since we already control for time and entity fixed effects in our regression model, we can use a difference-in-difference analysis in the trends between more attractive cities and less attractive cities, and between time periods with more and less Airbnb awareness, to gauge variations in the instrument variable. To do this, we plot the Zillow house price index for cities with different levels of 2009 touristiness across the timeframe after standardizing house price index using 2012 year as the baseline. The result is shown in Figure 7, which shows that **there exists no obvious differences in pre-trends in the Zillow House price index for cities until after 2012, which is also when the awareness of Airbnb starts to grow worldwide.** This suggests that cities with different levels of touristiness do not originally have divergent house price trends, and the difference in house price trends only starts to emerge after 2012 when Airbnb has gained its worldwide popularity. It is worth noting that in the second figure, the house price trend in Los Angeles experienced a much more significant surge after 2012, compared to other southern cities. This also justifies our exercise of comparing LA with other southern cities in the formal regression model in order to explore the differences in the effect size of Airbnb supply.

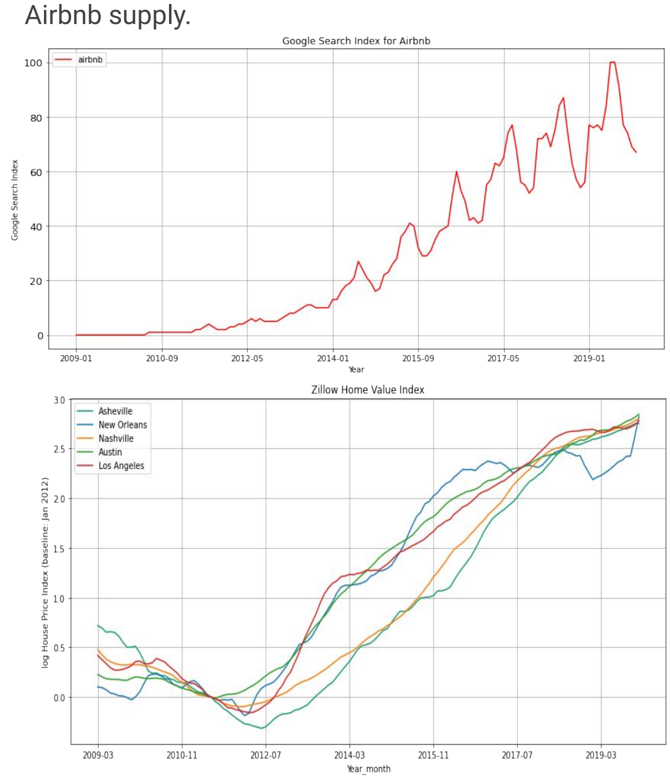

Figure 7: Testing the IV Assumptions

Construction of IV (Lasso)

We introduced five instruments for the first-stage. Even though we could include all the tests as instruments, that would risk overfitting the first stage. We want to optimally instrument an **endogenous regressor using the fewest possible instruments** and that's why we incorporated Lasso into the selection process. Building on top of a regular regression, LASSO regression works by adding bias (penalizing coefficients) to create a sparser model. To reduce variance/bias in the estimation, we incorporate control factors with instrument variables. We first **residualize y and x by regressing each variable in y and x on the set of control variables**, and then run lasso regression using the residualized variables. We performed the lasso regression with **different sets of hyperparameters to test the robustness of our process of selecting instruments**, and finally picked $airbnb * Establishment$ and $airbnb * estab\_density\_occup$ as our final instruments for $num\_listings_{it}$ and $num\_listings_{it}\ x\ occupancy_{it}$.

Two Stage Regression

Using the LASSO-constructed instrumental variables, we are able to come up with our final two stage fixed effects model as follows:

First stage is composed of two equations:

$$ln(Num\_listings_{it}) = \beta_0 + \beta_1 ln(Num\_listings_{it}) \times Estab_{it} + \beta_2 ln(Num\_listings_{it}) \times Estab\_Density_{it} \times Occupancy_{it} + X_{it}\eta + \Upsilon_i + \delta_t + \varepsilon_{it}$$

$$ln(Num\_listings_{it}) \times Occupancy_{it} = \beta_0 + \beta_1 ln(Num\_listings_{it}) \times Estab_{it} + \beta_2 ln(Num\_listings_{it}) \times Estab\_Density_{it} \times Occupancy_{it} + X_{it}\eta + \Upsilon_i + \delta_t + \varepsilon_{it}$$

where Xit is the vector of observed time-changing demographic characteristics including GDP, personal income, population, and unemployment, ϒi is the entity fixed effects and δt is our time fixed effects.

As the above model shows, in the first stage, we **instrument both endogenous variables with the LASSO-constructed variables, and use the fitted values of endogenous variables to conduct the second stage regression**:

	With Los Angeles			U.S. South Only		
	(OLS)	(FE)	(2SLS)	(OLS)	(FE)	(2SLS)
ln Airbnb Listings	0.070*** (0.015)	0.1207*** (0.017)	0.803*** (0.174)	0.203*** (0.037)	0.176*** (0.016)	0.172*** (0.039)
ln Airbnb Listing x Occupancy Rate(2009)	-0.022 (0.022)	-0.108*** (0.018)	-0.697*** (0.159)	-0.116*** (0.041)	-0.149*** (0.017)	-0.122*** (0.028)
ln GDP	-2.300*** (0.147)	3.973*** (0.327)	14.328*** (2.593)	-4.055*** (0.448)	-2.485*** (0.771)	-2.716*** (0.934)
ln Personal Income	1.137*** (0.256)	-4.202*** (0.522)	-23.667*** (4.700)	1.848*** (0.418)	1.276 (0.832)	0.98 (1.087)
ln Population	2.859*** (0.162)	-2.262*** (0.249)	-8.165*** (1.47)	5.037*** (0.533)	4.396*** (0.750)	4.603*** (0.951)
Unemployment Rate	-0.052*** (0.005)	-0.006** (0.003)	-0.0087*** (0.003)	-0.109 (0.008)	-0.007*** (0.003)	-0.0027 (0.003)
City FE	No	Yes	Yes	No	Yes	Yes›
Year FE	No	Yes	Yes	No	Yes	Yes
Instrumental Variable	No	No	Yes	No	No	Yes
Observations	600	600	600	480	480	480
R^2	0.951	0.993	0.995	0.871	0.982	0.986
F Statistic		123	117		247	210

Table 1: The Effect of Airbnb on House Price Index

Observation 1:

Table 1 reports the two stage regression results on all cities and on southern cities only in our dataset. Just as the initial fixed effects model suggests, we found that the interaction between Airbnb listings and occupancy rates is necessary to be controlled for, since it exhibits significant mediating effects for both models: the effect of Airbnb activities on a city's median housing price would decrease in the city's house occupancy rate. This observation conforms to our intuition that t**he effect of Airbnb listings mainly acts on the rental housing market where homeowners are more flexible in shifting their property from long-term rentals to short-term rentals on Airbnb.** Thus, for areas in any time period where home occupancy rate is high, the effect of airbnb listings on housing prices would be significantly reduced.

Observation 2:

By comparison of all cities and southern cities, it is worth noting that the use of instruments produces different results for the two groups. When we only include southern cities in our dataset, the coefficient on Airbnb listings decreased compared to the fixed effects result. Considering the relative homogeneity of southern cities, this conforms to **our initial hypothesis that certain unobserved trends (like gentrification) are most likely to be positively correlated with Airbnb listings as time varies.** However, fixed effects estimates could also be biased negatively due to several reasons.
Firstly, attenuation bias could arise as a result of measurement error in our pre-processing stage of data. Since there exists other similar peer-to-peer platforms for short-term rentals besides Airbnb, it is likely that we have obtained a noisy estimate of the true supply. Secondly, the bias could also be negative if the number of Airbnb listings would actually decrease with the increase of rental prices in the city. If this was the case, **our previous estimate of the effect of Airbnb listings using fixed effect OLS would be seriously underestimated**, and we would expect a larger coefficient for Airbnb listings by using the two stage model.

.

As our two-stage regression results suggest, simultaneous causality indeed functions differently in southern cities and in LA. Overall, the coefficient of airbnb listings suggest that a 1% increase in Airbnb listing would cause a 0.80% increase in the median house price for a city with median occupancy rate. When we only look at cities in the south, this effect decreases in its magnitude, suggesting that a 1% increase in Airbnb listings would lead to a 0.176% increase in house prices for a southern city like New Orleans, while holding everything else constant.

Observation 3:

Finally, in an effort to answer our initial hypothesis that the impact of Airbnb activities on housing prices exhibits different patterns for the southern cities compared to the more developed cities like Los Angeles, we conducted regression analysis on two datasets: one that includes LA housing price trends and one that excludes. As our results show, it seems that when Los Angeles data are taken into consideration, the **estimated effect of Airbnb activities becomes greater in magnitude even after controlling for local demographics and economic characteristics**.

While we must note that estimating the effect on median housing prices might overlook significant dissimilarity within a specific city (such as difference between suburbs and urban neighborhoods), and that we should be cautious with extending our conclusion to any smaller geographic units, the **disparity in the two sets of coefficients could still highlight strong evidences that are consistent with our hypothesis that while the increase of Airbnb activities contributes partially to the rising of house prices in the southern cities of US, the effect is not as overwhelming as in other regions where Airbnb listings take up a higher proportion of the housing market.**

		With Los Angeles		U.S. South Only	
		(Original)	(Placebo)	(Original)	(Placebo)
num_listings	airbnb * estab_density_ occup	0	0.532	0	0.723
	airbnb * establishment	0	0.959	0	0.681
num_listings * occupancy	airbnb * estab_density_ occup	0	0.461	0	0.647
	airbnb* establishment	0	0.913	0	0.596

Table 2: Placebo Test P-Value Results

Placebo Test

We implemented a form of randomization inference as a final exercise to **test whether our instrument is truly exogenous or primarily driven by spurious time trends**. To do so, we randomly swapped the industrument variables airbnb*establishment and airbnb*estab_density_occup among the five cities within the same month and keep other variables constant. Therefore, the randomized regressor preserves the overall time trends in the number of Airbnb listings but randomizes the identity of which city had how much Airbnb growth and thus eliminates the impact of touristiness on the margin of Airbnb listings. We ran the same 2SLS and found that the first-stage becomes very weak when regressing the randomized regressor.

on the instrument, leading to statistically insignificant estimates (Table x). If the results are primarily driven by a spurious time trend that interacts with the Airbnb listings, this placebo test will produce 2SLS estimates that continue to be positively and statistically significant.

However, the result we get is insignificant which proves that the effect of increment on Airbnb listings is really what matters. Taken together, the preceding results paint a strong picture in support of the validity of our instrument.

Caveats, challenges, and future research areas

During the course of our analysis, we used city-level data due to the limited data availability of socio-economic indicators at a lower level of geography.

By intuition, conducting the analysis on a neighborhood level would be more ideal because there might be **significant heterogeneity in housing markets across neighborhoods within cities but comparatively less heterogeneity within neighborhoods**. However, we believe our methodology is a principally grounded way to understand the impact of Airbnb listing activity on local property values and we would recommend extending a similar methodology to a **lower level of geography as well as other cities in the United States.**

Appendix

For our submission, we included this report, and a zip containing code and a data folder for our external datasets. The files are aptly named towards their purpose in the analysis.

References

1. Sheppard, S., & Udell, A. (2016). Do Airbnb properties affect house prices. *Williams College Department of Economics Working Papers*, *3*(1), 43. https://web.williams.edu/Economics/wp/SheppardUdellAirbnbAffectHousePrices.pdf
2. Barron, K., Kung, E., & Proserpio, D. (2021). The effect of home-sharing on house prices and rents: Evidence from Airbnb. *Marketing Science*, *40*(1), 23-47.https://pubsonline.informs.org/doi/abs/10.1287/mksc.2020.1227

Data Sources

Provided Data Sources:

1. **listings**: descriptive information on tens of thousands of Airbnb listings in the U.S. South.
2. **calendar**: basic information on the Airbnb listing calendar from 2016 – 2018.

Additional Data Sources:

1. **passenger_data.csv**: Time-series data for targeted cities on the number of airport arrivals (https://machinelearningmastery.com/autoregression-models-time-series-forecasting-python/)
2. **gdp.csv**: Time-series GDP data for targeted cities (https://fred.stlouisfed.org/series/NGMP12420, https://fred.stlouisfed.org/series/NGMP11700, https://fred.stlouisfed.org/series/NGMP31080, https://fred.stlouisfed.org/series/NGMP34980, https://fred.stlouisfed.org/series/NGMP35380)
3. **personal_income.csv**: Time series Personal Income per capita data for targeted cities (https://fred.stlouisfed.org/series/AUST448PCPI, https://fred.stlouisfed.org/series/ASHE737PCPI, https://fred.stlouisfed.org/series/PCPI06037, https://fred.stlouisfed.org/series/NASH947PCPI, https://fred.stlouisfed.org/series/NEWO322PCPI)
4. **unemployment.csv**: Time-series data for targeted cities on unemployment rate (https://data.bls.gov/PDQWeb/la)
5. **population.csv**: Time-series population data for targeted cities (https://stats.oecd.org/Index.aspx?DataSetCode=CITIES)
6. **occupancy.csv**: Home occupancy rate for targeted cities in 2010 (https://data.census.gov/cedsci/table?g=0400000US22_1600000US2255000&d=ACS%205-Year%20Estimates%20Data%20Profiles&tid=ACSDP5Y2010.DP04)

Code

For our analysis, we used the following programming languages, tools, and packages:

- **Python** and **Jupyter Notebook** were primarily used for data wrangling, EDA, preliminary visualization, Lasso, and Two Stage Regression modelling. We used standard libraries such as **pandas, numpy, matplotlib, seaborn, statsmodels, sklearn,** etc.
- **Kepler.gl** by Uber was an external mapping tool we used for geographical visualizations: https://docs.kepler.gl/