Impact of Short Term Rentals on the Rental Affordability in San Fransisco

The Case of Airbnb

Sukanya Sharma

Defended: June, 2018

Submitted in partial fulfillment of the requirements for the degree of Master of Urban Planning 2018 at the University of Illinois at Urbana-Champaign.

Masters Committee: Associate Professor Bumsoo Lee (advisor), Assistant Professor Andrew Greenlee

Abstract

Short term rentals listed on Airbnb and like platforms are common place all over the world. The advent of Airbnb operating on sharing economy models is a global phenomenon. Policy debates have naturally followed. Gauging impacts of such platforms is crucial and potentially can inform policy. Airbnb rentals all over the world, questions have been raised about their impacts (Meni, 2017). While one side discusses the democratization of tourism industry and how Airbnb makes it easier for tourists to travel and experience places, the other side brings forth the skirting of hotel taxes and negative neighborhood externalities that inevitably result from these rentals. Airbnb has attracted controversy in cities all over the world, with high-profile lawsuits centered around criticism for evasion of taxes and for avoiding regulatory oversight that is otherwise enforced on hotels and providers of similar services. The presented paper is an attempt to gauge the impact of Airbnb on rental affordability by using spatial econometric analyses. The study areas for the aforementioned research is the San Francisco Metropolitan Statistical Area (SF MSA).

The hypothesis is that an increase in the Airbnb listings (i.e. the short-term housing stock) in the study area is correlated negatively with rental affordability, causing it to decrease. Key research questions are does Airbnb impact the rental affordability in an area? If yes, then, to what extent? To answer these questions, both cross-sectional and longitudinal analyses are undertaken. Aiming to contribute to the body of literature, revolving around the debate through quantitative analyses and regulatory policy discussion, this study finds positive and statistically significant correlation between both Airbnb variables (percent Airbnb listings as a proportion of rental housing units and weighted Airbnb listings based on listing types) and variables representing rental affordability like percent rent-burdened and overburdened households, median rents and median house prices). Various models were considered for both cross-sectional and longitudinal analyses using different combinations of the aforementioned variables. The spatial econometric analysis answers one of our key questions in the affirmative – the presence of Airbnb rentals does impact the rental affordability in an area.

Having established a relationship, our second objective was to gauge this impact's extent. Simulations were run to understand the results of the spatial econometrics models to help visualize this impact. In the case of cross-sectional analysis for San Francisco MSA, these simulations showed that for a typical census tract (one with median percentage of Airbnb listings, as a fraction of the rental housing market) a 1% increase in Airbnb listings corresponded to a 0.06% rise in the rent overburdened household category. Hence, in the case of a census tract with 10,000 households, a 10% increase in percent Airbnb listings will correspond to 60 more households being added to the rent overburdened category. This effect is more pronounced for tracts with a lower number of Airbnb listings (10th or 25th percentiles). Additionally, tracts with no or a low percentage of Airbnb listings will have more households pushed to a rent-burdened category, with a similar rise in Airbnb listings.

In the case of longitudinal analysis of panel data for San Francisco City for a period of four years (2013– 2016), the simulations show that census tracts with a smaller presence of Airbnb listings (those below the 50th percentile) were more sensitive to an increase in Airbnb listings i.e. they saw a higher increase in the median house price per tract as compared to census tracts in the higher percentiles. This trend was consistent across all four years affirming the extent of the impact of Airbnb listings on the rental affordability in an area.

Chapter 1: Introduction

A major motivation for this research is the acute housing affordability crisis in US. California, like many other states currently faces a housing affordability crisis, the effects of which are particularly strong in the bay area. San Francisco city and MSA (including counties Alameda, Contra Costa, San Francisco, Marin and San Mateo) continue to show drops in rental as well as housing affordability levels. With the affordability situation this dire, a decrease in housing stock poses a major threat. In the case of rental affordability, the decrease in long-term rentals can lead to further increase in rent and consequently a decrease in rental affordability. This makes a case for this research and is one of the main motivations for this study.

Platforms like Airbnb are redefining tourist consumption patterns in ways that can impact housing affordability. From humble beginnings as a start-up in San Francisco in 2008, the platform is now valued at over 31 billion dollars[1], is present in 190 countries worldwide, while continuing to operate on the principles of sharing economy, peer-to-peer markets and the 'digital' economy. Airbnb and like platforms reaching such high valuations and becoming commonplace, it has become imperative to critically gauge their impacts.

There is an emerging body of literature investigating the relationship between Airbnb and housing related outcomes, such as average rent, tourism, rates of hotel usage, and more. These studies vary widely in their methods and represent a new field of study within housing policy and urban planning, some of which are discussed in the literature review of this research. The debate is divisive, and analysts is divided firstly on whether Airbnb has any effect at all and secondly whether these effects are positive or negative.

The presented research is an inquiry into the impacts of short-term rentals on the rent affordability of the study area. Due to the largest share among its competitors and its popularity, this study chose Airbnb as a representative platform for short-term rentals. One of the main motivations of this research is the affordability crisis afflicting most cities in the US and particularly San Francisco. This research finds its roots in understanding the complex dynamics of having Airbnb like platforms in an area which is undergoing a worsening affordability crisis. Additionally, this study aims to contribute to the discourse of one of the most pressing question – regulating sharing economy by adding data to the debate. The following section elaborates on the hypothesis, aims and methodology undertaken of this research.

1.1 Hypothesis

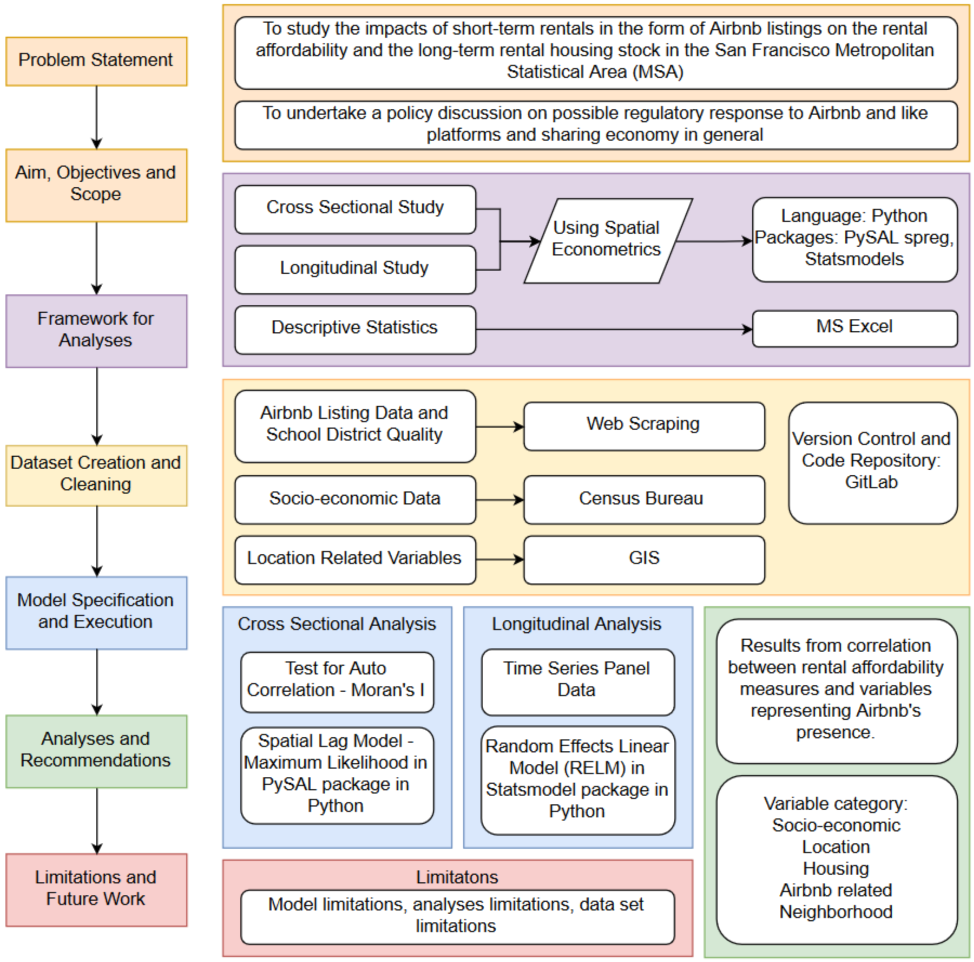
This research uses spatial econometrics techniques to examine the impact of Airbnb on the rental affordability and long-term rental housing stock in the San Francisco MSA. It hypothesizes is that with an increase in the Airbnb listings (i.e. short-term rental housing stock) there is a decrease in the rental affordability in the study area. This is studied using various locational, socio-economic and neighborhood level variables. More formally, the hypothesis states that the variables log percent Airbnb rentals (active and all rentals) and weighted Airbnb listings are significantly and positively correlated to rental affordability measures and rents (rent burdened, rent overburdened and gross median rent). Key research questions are *does Airbnb impact rental affordability in an area? If yes, then to what extent?*

1.2 Objectives and Methodology

The objectives of this research are as follows:

- To study the impacts of short-term rentals in the form of Airbnb listings on the rental affordability in the San Francisco Metropolitan Statistical Area (MSA) and San Francisco City.
- To undertake a policy discussion on possible regulatory response to Airbnb and like platforms and sharing economy in general.

[1] Bensinger, Greg. Airbnb Valued at $31 Billion After New Funding Round; Home-sharing site adds $1 billion cash cushion to stave off IPO, Wall Street Journal (Online); New York, N.Y. 09 Mar 2017. Retrieved from: `https://www.wsj.com/articles/airbnb-valued-at-31-billion-after-new-funding-round-1489086240`



Source: author.

Figure 1: Research methodology.

For the purposes of this research, both cross-sectional and longitudinal research was conducted. The research methodology presented in the figure above shows the steps that were undertaken and details of those steps. As shown in the figure, analyses were divided into two main sections—the cross-sectional analysis and longitudinal. A detailed treatment of these analyses is carried out in Chapters 3 and 4.

Chapter 3: Data and Methods

As mentioned in the chapter 1, this study conducts both cross-sectional and longitudinal analysis. This chapter details out the methods used, the theory behind those methods, dataset generation, our unit of analysis and the variables considered. It is important to note here that the cross-sectional analysis was carried out for the San Francisco MSA (a five-county region) whereas the longitudinal analysis was carried out for San Francisco city due to limited availability of data on Airbnb listings.

Note: The section on Longitudinal Analysis has been omitted from this extract for the sake of brevity.

3.1 Cross-sectional Analysis

The cross-sectional analysis was carried out to provide information on the characteristics of and statistical relationships between selected dependent and independent variables, at a specific moment in time – 2017 for this research.

Study Area

The chosen area for the cross-sectional study was the San Francisco-Oakland-Hayward MSA (Metropolitan Statistical Area) which is a five-county region in California with a population of 4,679,166 (2016 ACS estimate) and an area of 2,474 square miles (6,410 km^2). It consists of Alameda County, Contra Costa County, San Francisco County, San Mateo County, and Marin County. To better gauge the extent of spatial correlation and expand the analysis, the MSA area was selected so as to not confine the analysis to San Francisco City. This five-county area includes major urban centers as well as suburban development and few rural areas. This diversity also helped in understanding the impact of Airbnb listings beyond just a city/urban area. Figure 2 shows the five-county study region and Airbnb listings in it.

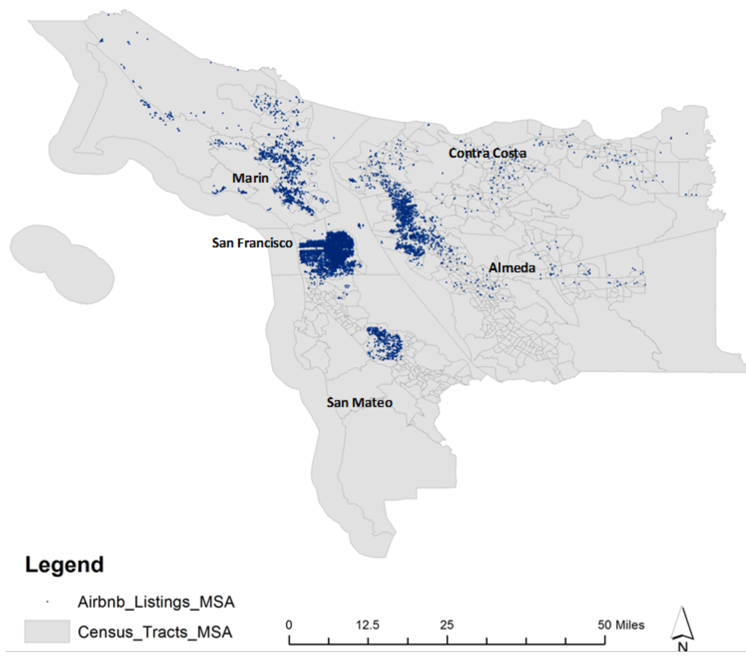


Figure 2: Airbnb listings in San Francisco MSA as of 2017.

3.2 Dataset Generation

For the purposes of dataset creation, various methods were used to collect relevant data. For the Airbnb listing data, a script[2] was developed in Python using selenium and PostgreSQL packages to scrape data off of airbnb.com. The scraping was undertaken for all five counties in the study area and listings data was extracted with their location coordinates using the bounding box method. The scraping was undertaken in December 2017. In addition to using Airbnb listings as a percentage of rental housing units[3], a composite score index was created which incorporated the difference of listing type. A weight of 1 was given to entire house listings, 0.5 for private rooms and 0.2 for shared rooms/couches etc. The distribution of percent Airbnb listing and weighted listings indicate positively skewed distributions as shown in the figure below. Hence to account for heteroskedasticity, all variables used in the model are natural logarithms (except the dummy variable) were log-transformed.

[2] Available at GitLab (`https://gitlab-beta.engr.illinois.edu/sukanya3/Airbnb_Spatial_Econometrics`) under open license

[3] Rental housing data obtained from 2016 ACS estimates

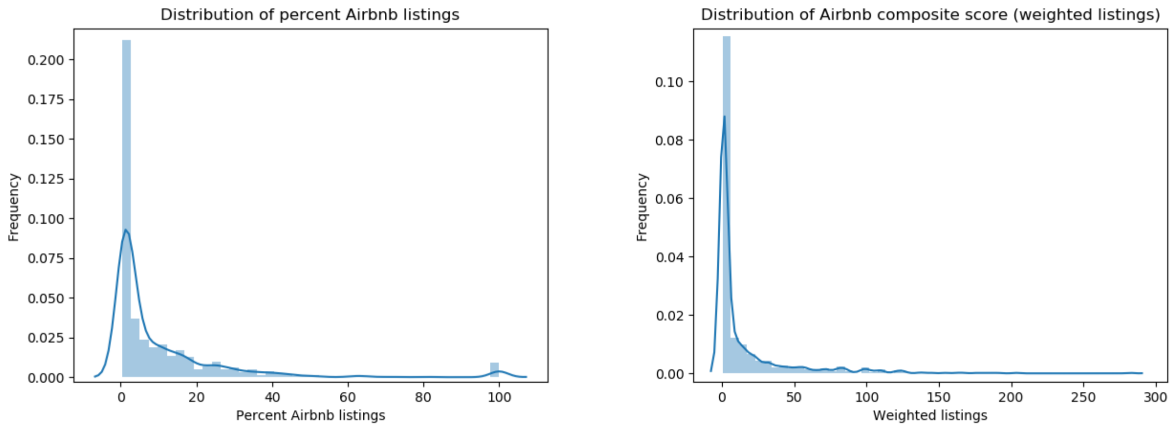


Figure 3: Distribution of Airbnb listings in the study area

Data for independent (control) variables like `log percent unemployed population`, `log percent foreign-born population`, `log percent non-white population`, `log rent burdened households` and `log rent overburdened households` was obtained from the American Community Survey, using 2016 estimates for the study area. This data was collected at the census-tract level.

Another independent variable in the dataset is `log school district quality` which is in the form of a score assigned by an independent non-profit — greatschools.org. The school quality is reflected from a score called summary rating that the website gives based on various criteria. According to the nonprofit, the summary rating is a multi-measure school quality metric intended to reflect a school's characteristics and quality across multiple dimensions, ultimately representing the school's overall quality in preparing students for postsecondary success. It is an aggregation of the school's "sub-ratings", which include test score, student progress, academic progress, equity, college readiness, and advanced courses ratings, as well as a flag for discipline and attendance issues[4]. The data for this variable was web-scraped from the greatschools.org website using a Python script[5].

The centroids of the census-tracts were used as address/locations for all schools in that census-tract. The highest school quality score was selected in case of multiple schools existing in the same tract. There was an element of reverse geocoding the location coordinates of the census tract centroids by using Google API to create addresses that the greatschools.org website accepts for locating relevant schools at the high school level.

For the location variables like `Log BART dist` (log of Euclidian distance of BART stations from census-tract centroid), `Coastline tracts` (dummy variable where 1 denotes a coastline tract) and `Log CBD dist` (log of Euclidian distance between nearest central business district and centroid of the census-tract) were computed using ArcMap in ArcGIS and its functions in ArcToolbox and network analyst.

For gauging job accessibility, two variables from the Smart Location Database were used. The primary variables – `D5ar` and `D5br` from the destination accessibility dataset were used because they measure jobs or working- age population within a 45-minute commute via automobile (`D5ar`) or transit (`D5br`). The "r" reflects the accessibility of job from residences to jobs. This data was collected at the census-tract level for all five counties in the study area. Table 1 shows the variable categories and description in brief.

3.3 Methods

For the purpose of this analysis, spatial econometrics was used. Spatial econometrics is a subfield of econometrics that deals with spatial interaction (spatial autocorrelation) and spatial structure (spatial heterogeneity) in regression models for cross-sectional and panel data (Paelinck and Klaassen, 1979; Anselin, 1988a). It is used in theoretical models which involve interactions between different entities and for models with data observations which are not truly independent due to spatial auto-correlations and neighborhood effects. Figure 4 shows the step by step diagnostic flowchart of arriving at an appropriate model for a cross-sectional analysis.

[4]Methodology used by Great Schools non-profit. Retrieved from `https://www.greatschools.org/gk/ratings-methodology/`

[5]Inspired by the script developed and shared by Yongsung Lee, a Ph.D. candidate at School of City and Regional Planning, Georgia Institute of Technology

Variable Category	Variable Name	Description	Data Source	Remarks
Independent Variables				
Airbnb	Log Percent Airbnb	Log of Airbnb listings as a percentage of rental housing units	Airbnb.com web scrape and US Census TIGER/Line Files	
	Log Weighted Airbnb listings	Log of weighted Airbnb listings (1 for entire home, 0.5 for private room, 0.2 for shared room)	Airbnb.com web scrape	
Location	Log BART dist	Log Euclidean distance in meter between census tract centroid and nearest BART station	ArcMAP analysis and TIGER/Line shapefiles from Census Bureau	
	Log CBD dist	Log Euclidean distance in meter between census tract centroid and nearest Central Business District	ArcMAP analysis and TIGER/Line shapefiles from Census Bureau	
	Coastal tracts (Dummy)	if a census tract is at the coast line; otherwise 0	ArcMAP analysis and TIGER/Line shapefiles from Census Bureau	
Demographic	Log unemployment rate	Log of percentage unemployed people as a percentage of the civilian labor force	US Census TIGER/Line Files	
	Log percent non-white	Log of percentage population which is not white	US Census TIGER/Line Files	
	Log percent foreign-born	Log of percentage population who is not US citizen at birth, includes those who become US citizens through naturalization	US Census TIGER/Line Files	
Neighborhood Level	Log school district quality	Log of school district score as given by greatschools.org	Greatschool.org web scrape	
Job Accessibility	Log accessibility by car	Log of jobs within 45 minutes auto/car travel time; time-decay (network travel time) weighted	Smart Location Database, US EPA Smart Growth Program	
	Log accessibility by transit	Jobs within 45-minute transit commute, distance decay (walk network travel time) weighted	Smart Location Database, US EPA Smart Growth Program	
Dependent Variable				
Rental Affordability Measures	Log rent burdened	Log of percentage households spending 30% or more of gross monthly income towards total housing costs	US Census TIGER/Line Files	Inversely related to rental affordability
	Log rent overburdened	Log of percentage households spending 50% or more of gross monthly income towards total housing costs	US Census TIGER/Line Files	Inversely related to rental affordability
Housing variables	Log median rent	Log of median gross rents or monthly housing cost expenses for renters	US Census TIGER/Line Files	Inversely related to rental affordability
	Log median house price	Log of median house prices	US Census TIGER/Line Files	Inversely related to rental affordability

Table 1: Variable categories and descriptions.

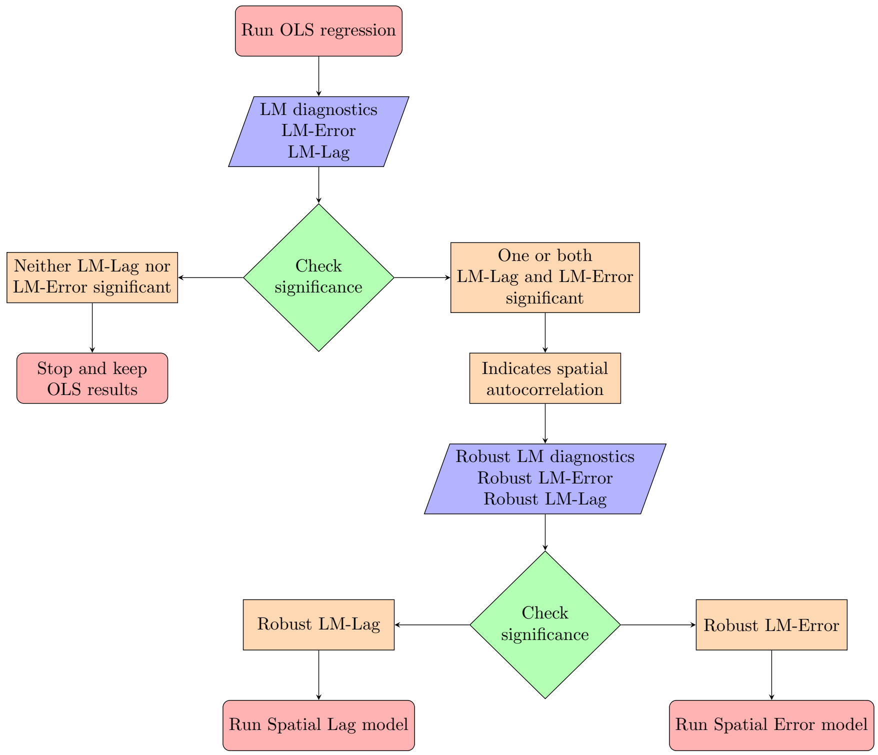


Figure 4: Steps undertaken for diagnostics of spatial specifications

The data was tested for spatial correlation using the Moran's I test, before proceeding with regression analysis to understand the correlation between rental affordability and Airbnb's presence. As expected, the data tested positive for spatial autocorrelation. Chapter 4 contained the details of the test for spatial autocorrelation and also contains a report for Global Moran's I. Hence the focus of the quantitative analysis was to undertake spatial analysis on the principles of spatial econometrics. The spatial lag model can be written as:

$$\mathbf{y} = \rho\mathbf{W}\mathbf{y} + \mathbf{X}\beta + \mathbf{u}$$

where $\mathbf{y}$ is an $(n \times 1)$-vector of observations on the dependent variable, $\mathbf{W}$ is an $(n \times n)$-spatial lag operator and $\mathbf{W}\mathbf{y}$ is a spatial lag term with spatial auto-regressive parameter ρ, $\mathbf{X}$ is an $(n \times k)$-matrix of observations on exogenous (independent) explanatory variables with $(k \times 1)$-coefficient vector β, and an $(n \times 1)$-vector of errors $\mathbf{u}$.

For this analysis, Maximum Likelihood (ML) approach was used in conjunction with the aforementioned spatial lag model. ML does not allow for the presence of multiple endogenous (dependent) variables for the model specification. For our analysis, a single independent variable suffices. ML approach assumes homoskedasticity of the error term $\mathbf{u}$.

For easier understanding, the model can be thought of in the following way:

- A shapefile is used to construct a spatial weights matrix (which assigns weights based on k nearest neighbors for each unit of analysis i.e. census-tract). In this model k was chosen to be 4, meaning each census-tract is assumed to be impacted by 4 nearest census-tracts around it.
- A log-likelihood variable can be defined as a function of parameter: β, ρ, and σ^2, where σ^2 is the variance of the error distribution.
- The ML estimates for these three parameters are found by equating their first derivatives to zero and solving the resulting equations.
- Finally, the maximum log-likelihood can be computed by numerically estimating the single parameter ρ. These steps were carried out with the help of `spdep` and `car` packages in R. Chapter 4 details out the results obtained from these computations.

Chapter 4: Results

This chapter presents the results of the cross-sectional and longitudinal analysis carried out by using the models described in the last chapter. One of the first steps for both cross-sectional and longitudinal analyses was the test for autocorrelation. Spatial autocorrelation is an integral concept in spatial statistics as it enables the investigation for spatial interpolation. Simply put, Spatial autocorrelation is a measure of similarity or correlation between nearby observations. To test for spatial autocorrelation, the Moran's I test was conducted. Moran's I test suggests that:

- -1 is perfect clustering of dissimilar values or perfect dispersion
- 0 is no autocorrelation or perfect randomness and
- $+1$ indicates perfect clustering

Moran's I is an inferential statistic and hence there is a need to assess whether the index generated is significant or not. This is done with a simple hypothesis test of calculating Z-score and its associated p-value.

- The null hypothesis for the test states that data is randomly disbursed.
- The alternate hypothesis states that it is more spatially clustered.

Two possible scenarios then become that positive Z-value will mean that the data is spatially clustered whereas a negative Z-value will mean that the data is clustered in a competitive way. For example, high values are repelling high values or negative values are repelling negative values. Table 2 shows the Global Moran's I test report generated on the percent Airbnb listing dataset and the results show that there is clustering, and hence spatial autocorrelation is significant and present. Hence a simple OLS regression for the dataset will not yield credible results and there is a need to use models that account for the spatial autocorrelation for both the cross-sectional and longitudinal analysis.

Moran's Index	0.029840
Expected Index	-0.000095
Variance	0.000001
Z-score	37.464236
p-value	0.000000

Table 2: Global Moran's I-test for spatial auto-correlation.

Following a positive test for spatial autocorrelation, the following results were obtained from the cross-sectional analysis. The study area is the San Francisco MSA and as mentioned earlier and all variables are natural logarithms used to tackle the issue of heteroskedasticity. Table 3 summarizes the results from spatial lag model used in the cross-sectional analysis. The table shows eight models which are varying combinations of independent and dependent variables. Each model is a combination of one of the rental affordability measures as Y variable (`log rent burdened`, `log rent overburdened`, `log median rent` and `log house price`), and two sets of X variables (with `log percent Airbnb listings` or `log composite score` as one of the key variables along with all demographic, neighborhood level and location variables).

Following observations can be made from the results of the spatial lag model. To interpret the models, we pay attention to Probability values (should be $p < 0.1$ for a significant correlation), Coefficient values to ascertain the dependence of independent variable on Y.

- Both Airbnb variables (percent and weighted listings) have positive coefficient for all eight models. Additionally, they are all statistically significant.
- Location variables, `log BART dist.` and `log CBD dist.` show a positive coefficient. This is as expected since proximity to BART stations and downtown/central business districts is usually accompanied by higher rents, house prices and hence more number of rent burdened and overburdened households. The dummy location variable accounting for whether or not a census tract is on the coast shows negative coefficients for models 1, 2, 3 and 4 as expected. This can be attributed to the fact that housing in coastal tracts (with views) usually is premium housing and therefore attract only higher income groups leading to lower rent burdened households. However, models 5, 6 7 and 8 either show statistically insignificant results or counter-intuitive signs (negative). This indicates an opportunity to use more nuanced variables representing coastal locations in the study.

Independent variables (X)		Dependent variable (Y)							
Category	Variable name	Log Rent burdened		Log Rent overburdened		Log median rent		Log median house price	
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Airbnb	Log percent Airbnb	0.001**		0.026***		0.071***		0.181***	
		(0.013)		(0.019)		(0.025)		(0.055)	
	Log Weighted Airbnb listings		0.021**		0.013***		0.142***		0.223***
			(0.013)		(0.019)		(0.025)		(0.054)
Location	Log BART distance	0.020	0.016	0.051**	0.053**	0.083**	0.064*	0.288***	0.271***
		(0.018)	(0.018)	(0.026)	(0.026)	(0.034)	(0.033)	(0.073)	(0.073)
	Log CBD distance	0.015	0.040*	0.034	0.041	0.085**	0.194***	0.190**	0.301***
		(0.020)	(0.022)	(0.029)	(0.033)	(0.038)	(0.042)	(0.082)	(0.093)
	Coastal tracts (dummy)	−0.105	−0.112	−0.343**	−0.348**	−0.047	−0.100	−0.915**	−1.000**
		(0.101)	(0.101)	(0.147)	(0.147)	(0.192)	(0.190)	(0.418)	(0.418)
Demographic	Log unemployment rate	0.293***	0.300***	0.392***	0.393***	0.037***	−0.001***	−0.325***	−0.273**
		(0.029)	(0.029)	(0.042)	(0.043)	(0.055)	(0.055)	(0.120)	(0.121)
	Log percent non-white	0.144***	0.146***	0.210***	0.210***	0.093**	−0.104*	0.077	−0.094**
		(0.029)	(0.029)	(0.042)	(0.042)	(0.054)	(0.054)	(0.118)	(0.118)
	Log percent foreign-born	0.014***	0.015	0.025*	0.025**	0.029**	0.030**	0.241***	0.241***
		(0.011)	(0.011)	(0.015)	(0.015)	(0.020)	(0.020)	(0.044)	(0.044)
Neighborhood level	Log school district quality	−0.002**	−0.004*	−0.043*	−0.041*	0.027	0.022***	0.171***	0.170***
		(0.015)	(0.015)	(0.022)	(0.022)	(0.029)	(0.029)	(0.064)	(0.064)
Job accessibility	Log accessibility by car	0.031	−0.038**	−0.007	−0.008	0.139***	0.174***	0.333***	0.371***
		(0.019)	(0.019)	(0.028)	(0.028)	(0.036)	(0.036)	(0.078)	(0.080)
	Log accessibility by transit	−0.009**	−0.010**	−0.003**	−0.003**	0.011	−0.017**	−0.022	−0.029
		(0.004)	(0.004)	(0.006)	(0.006)	(0.008)	(0.008)	(0.018)	(0.018)
Constant	Intercept	2.067***	1.833***	1.488***	1.548***	3.306***	2.378***	5.496***	4.475***
		(0.327)	(0.344)	(0.420)	(0.449)	(0.624)	(0.644)	(1.291)	(1.349)
Tests and statistics	Number of observations	975	975	975	975	975	975	975	975
	Log likelihood	−598.806	−597.526	−966.312	−966.977	−1230.074	−1217.766	−1987.387	−1984.384
	Rho	0.033989***	0.035772***	0.012907***	0.012026*	0.16286***	0.14753***	0.071619***	0.066014**
	AIC	1223.612	1221.051	1958.625	1959.954	2486.149	2461.533	4000.774	3994.767

Notes: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 3: Coefficients and standard errors for Spatial Lag models

- All the demographic and neighborhood level variables show significant correlation and similarly show expected signs.
- The job accessibility variables show mixed results. While most of the coefficients are significant, it is hard to intuitively grasp the reason behind their signs without resorting to more detailed accessibility models.

The tests and statistics give the log likelihood and Akaike Information Criteria results. Both these statistics indicate the quality of models. A lower value of log likelihood and a larger value of AIC indicate a better-quality model. Additionally, Rho statistic is significant for all models and indicates high spatial dependence within the dataset. All the models with percent Airbnb as the independent variables show better test values than the weighted listings variables. Hence the results from these models are further simulated below for more intuitive understanding.

Since both the independent and dependent variables were log-transformed and fairly low in magnitude, we simulate the results by plotting values estimated by the model at the 0^{th} (min), 25^{th}, 50^{th} (median), 75^{th}, 90^{th} and 100^{th} (max) percentile of the X variable. These simulations show that for a typical census tract (one with median percentage of Airbnb listings, as a fraction of the rental housing market) a 1% increase in percent Airbnb listings corresponded to a 0.06% rise in the rent overburdened household category. Hence, in the case of a census tract with 10,000 households, a 10% increase in percent Airbnb listings will correspond to 60 more households being added to the rent overburdened category. This effect is more pronounced for tracts with a lower number of Airbnb listings (10^{th} or 25^{th} percentile).

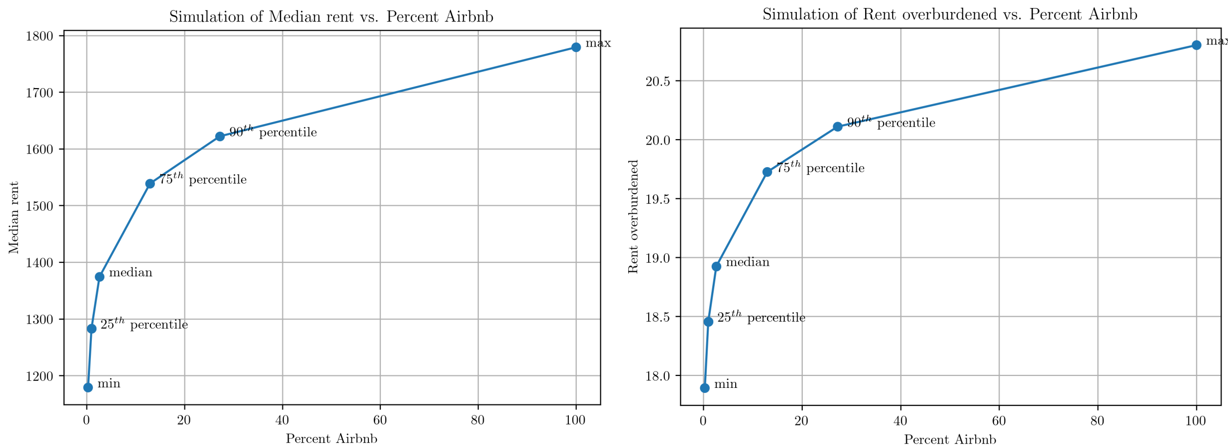


Figure 5: Simulations for the cross-sectional analysis

This would mean that tracts with no or low percentage Airbnb listings will see more households being pushed to a rent burdened category, with similar rise in Airbnb listings.

In the case of median rents, for a typical census tract, a 1% increase in percent Airbnb listings corresponded to a \$12 hike in median gross rents. For census tracts with a lower presence of Airbnb (25^{th} percentile or lower), this number got as high as \$100.

In addition to the cross-sectional study, a longitudinal analysis of panel data was also conducted (but not included in this extract). It is important to note here that the study area for this analysis is the San Francisco City and not the whole MSA. Hence, the unit of analysis is a census tract and the panel data accounts for a four-year period from 2013 to 2016.

- Both Airbnb variables (`percent all rentals` and `percent active (occupied) rental housing`) have positive coefficient for all eight models. Additionally, they are all statistically significant.
- Location variables like `log BART distance` and `log CBD distances` and the time effects variable `trend` in some case are not significant and in others, don't show consistent and expected trends.
- In the case of demographic variables, `percent bachelor's degree`, `percent unemployed` and `median household income` variables show significant and expected signs. However, `percent foreign-born` variable does not.
- The R^2 values for model 1 and 2 are higher and therefore a larger part of the variation in the Y is explained by the independent variables X in these models. The adjusted R^2 values show equivalent results.

Independent variables (X)		Dependent variable (Y)							
Category	Variable name	Log Rent burdened		Log Rent overburdened		Log median rent		Log median house price	
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Airbnb	Log percent Airbnb all rentals	0.196^{***}		0.003^{**}		0.030^{**}		0.172^{**}	
		(0.014)		(0.008)		(0.018)		(0.074)	
	Log percent Airbnb active rentals		0.193^{***}		0.001^{**}		0.029^{**}		0.037^{*}
			(0.014)		(0.008)		(0.017)		(0.071)
Location	Log BART distance	0.188^{***}	-0.187^{***}	-0.161^{***}	-0.125^{***}	-0.009^{*}	-0.008	0.294^{***}	0.289^{***}
		(0.042)	(0.042)	(0.052)	(0.052)	(0.028)	(0.028)	(0.097)	(0.098)
	Log CBD distance	0.037	0.043	0.042	0.032	0.165^{***}	0.166^{***}	-0.083	-0.084
		(0.050)	(0.050)	(0.061)	(0.061)	(0.033)	(0.033)	(0.115)	(0.115)
	Coastal tracts (dummy)	0.325	0.286	0.302	0.262	-0.028	-0.033	1.128^{**}	-1.195^{**}
		(0.229)	(0.229)	(0.286)	(0.286)	(0.149)	(0.149)	(0.512)	(0.513)
Demographic	Log percent bachelor's degree	-0.040^{**}	-0.035^{***}	-0.122^{**}	-0.102^{**}	0.431^{***}	0.432^{***}	0.462^{***}	0.475^{***}
		(0.059)	(0.059)	(0.057)	(0.057)	(0.048)	(0.048)	(0.173)	(0.174)
	Log percent foreign-born	-0.019	-0.015^{***}	0.012	0.011	0.0003	0.001^{***}	0.574^{***}	0.550^{***}
		(0.015)	(0.015)	(0.009)	(0.009)	(0.017)	(0.018)	(0.069)	(0.070)
	Log percent unemployed	0.086^{**}	0.084^{**}	0.071^{***}	0.065^{***}	-0.016^{***}	-0.016^{***}	-0.309^{**}	-0.324^{***}
		(0.037)	(0.037)	(0.025)	(0.025)	(0.034)	(0.034)	(0.123)	(0.123)
	Log median household income	-0.003^{**}	0.0003^{***}	-0.018^{***}	-0.013^{***}	0.074^{***}	0.074^{***}	0.147^{**}	0.156^{***}
		(0.022)	(0.022)	(0.015)	(0.015)	(0.022)	(0.022)	(0.080)	(0.080)
Time	Trend	0.020	0.013	-0.017^{*}	-0.014^{*}	-0.030	-0.031	-0.248^{***}	-0.138
		(0.016)	(0.016)	(0.009)	(0.009)	(0.020)	(0.021)	(0.085)	(0.086)
Constant	Intercept	3.457^{***}	3.431^{***}	3.887^{***}	3.837^{***}	4.258^{***}	4.251^{***}	8.387^{***}	8.187^{***}
		(0.328)	(0.328)	(0.350)	(0.350)	(0.238)	(0.238)	(0.842)	(0.843)
Tests and statistics	Number of observations	784	784	784	784	784	784	784	784
	R^2	0.459	0.467	0.053	0.049	0.272	0.271	0.174	0.168
	Adjusted R^2	0.452	0.461	0.042	0.032	0.263	0.262	0.164	0.158
	F statistic	72.827^{***}	75.261^{***}	4.825^{***}	3.231^{***}	32.056^{***}	31.937^{***}	18.058^{***}	17.371^{***}

Notes: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 4: Coefficients and Standard errors for Random Effects Linear model

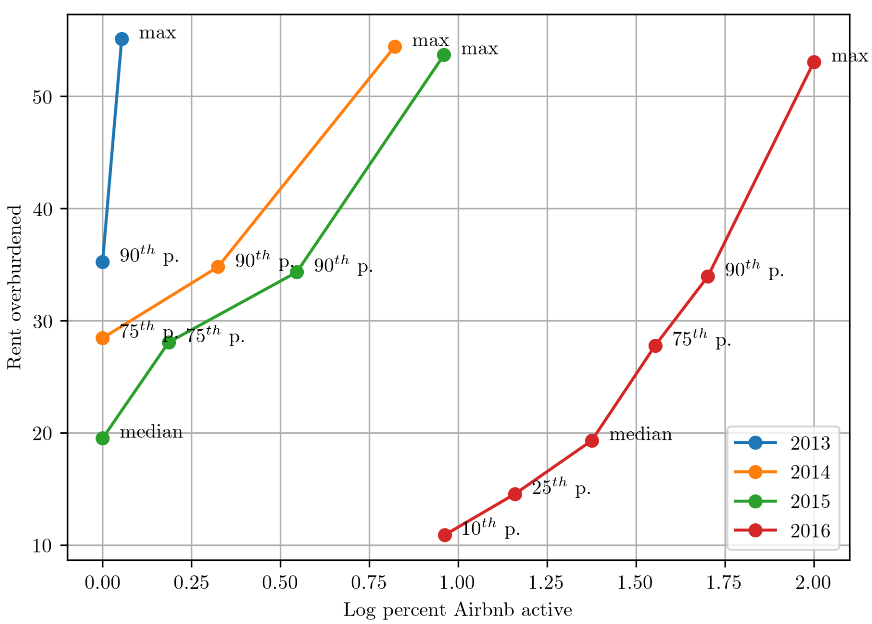


Figure 6: Simulation of rent overburdened households vs. Airbnb listings for 2013–2016

Following observations can be made from the simulation of rent overburdened households and percent Airbnb listings (as a fraction of active/occupied rental units).

- Since both the independent and dependent variables were log-transformed and fairly-low in magnitude (coefficients), we simulate the results by plotting values estimated by the model at the 10^{th}, 25^{th}, 50^{th} (median), 75^{th}, 90^{th} and 100^{th} (max) percentile of the X variable.
- Each simulation shows trends for Y vs. $\log X$ for years 2013 to 2016.
- Figure 6, overall trend shows that an increase in Airbnb active rentals (Airbnb listings, as a fraction of occupied/active rental housing market) corresponded to an increase in fractions of households which were rent overburdened.
- Furthermore, census tracts with a smaller presence of Airbnb listings (those below 50th percentile) were more sensitive to an increase in Airbnb listings i.e. they saw a higher increase in the rent-overburdened household category as compared to census tracts in the higher percentiles. This trend was consistent across all four years.
- A note of caution – the x-axis for the simulation represents values for log of the variable and care should be taken when reading off numerical values from it.
- Another way to read the simulation graph is to look at changes in rent overburdened households for a fixed percentile mark across all four years. Doing this for all percentiles (except the maximum), we observe no significant change in the fraction of rent overburdened households over the years.
- The maximum value (100^{th} percentile) at first glance seems to be decreasing over the years which is not the case. However, these observations can be interpreted as an indication of a trend wherein over the years, the census tracts with lower rent overburdened populations have seen a larger increase in Airbnb listings. This phenomenon results in a drop in the Y value associated with the census tracts that have maximum percentage of listings which is why the simulation for 2016 shows smaller values of Y as compared to the other years. In effect, these values should not be compared across the years without also accounting for shifting distributions of Airbnb listings.

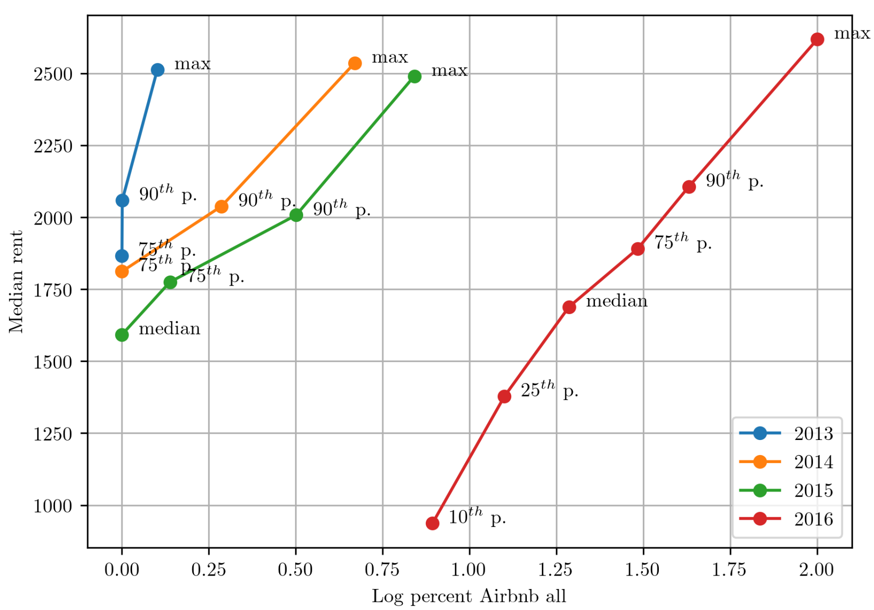


Figure 7: Simulation of median rent vs. Airbnb listings for 2013–2016

Following observations can be made from the simulation of median rents and percent Airbnb listings (as a fraction of all rental units).

- Figure 7, overall trend shows that an increase in Airbnb all rentals (Airbnb listings, as a fraction of all rental housing in the market) corresponded to an increase in the median rent per census tract.

- As before, for the early years (2013-14), we do not include data points for the lower medians in the simulation in order to avoid clustering of percentiles around zero. This is a consequence of fewer tracts having Airbnb listings during that time.
- From the graph, we observe that census tracts with a smaller presence of Airbnb listings (those below the 50^{th} percentile) were more sensitive to an increase in Airbnb listings i.e. they saw a higher increase in the median rent per tract as compared to census tracts in the higher percentiles. This trend was consistent across all four years.
- A note of caution – the x-axis for the simulation represents values for log of the variable and care should be taken when reading off numerical values from it.
- Another way to read the simulation graph is to look at changes in median rent for a fixed percentile mark across all four years. Doing this for all percentiles, we observe overall increases in the median rent over the years for each of the 50^{th} (median), 75^{th}, 90^{th} and 100^{th} (maximum) percentiles.
- Since both the y- and x-axes plot medians of the variables, one should be extra-cautious while interpreting these results as they are sensitive to the time-dependent probability distributions for these variables.

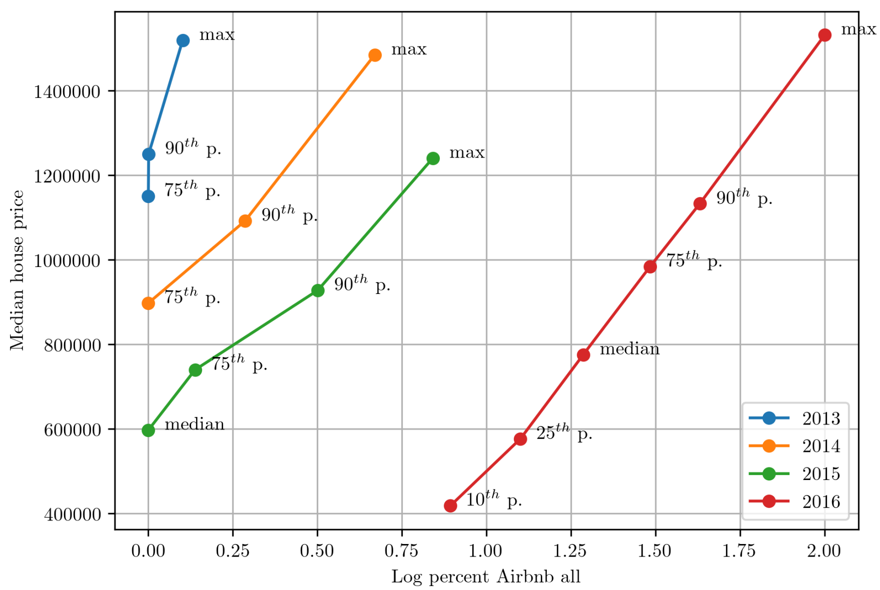


Figure 8: Simulation of median house price vs Airbnb Listings for 2013–2016

Following observations can be made from the simulation of median house prices and percent Airbnb listings (as a fraction of all rental units) —

- Figure 8, the overall trend shows that an increase in Airbnb all rentals (Airbnb listings, as a fraction of all rental housing in the market) corresponded to an increase in the median house prices in census tracts.
- As before, for the early years (2013–14), we do not include data points for the lower medians in the simulation in order to avoid clustering of percentiles around zero. This is a consequence of fewer tracts having Airbnb listings during that time.
- From the graph, we observe that census tracts with a smaller presence of Airbnb listings (those below the 50^{th} percentile) were more sensitive to an increase in Airbnb listings i.e. they saw a higher increase in the median house price per tract as compared to census tracts in the higher percentiles. This trend was consistent across all four years.
- A note of caution – the x-axis for the simulation represents values for log of the variable and care should be taken when reading off numerical values from it.

- Another way to read the simulation graph is to look at changes in median house price for a fixed percentile mark across all four years. Doing this for all percentiles, we observe a sharp and unexpected decrease in the median house price (corresponding to these percentiles) from 2013–2015. This does not mean that house prices were decreasing in any way.
- These observations can be interpreted as an indication of a trend wherein over the years, the census tracts with lower house prices have seen a larger increase in Airbnb listings. This phenomenon results in a drop in the Y value for each of these percentiles which is why the simulation for 2013–15 shows a decreasing trend in Y. In effect, these values should not be compared across the years without also accounting for the time-dependent probability distributions for these variables since both the y- and x-axes plot percentiles of the variables.

Chapter 5: Discussion and Conclusion

Using spatial regression analyses of cross-sectional and longitudinal data specifically focusing on census tract level location, demographic, neighborhood level, job accessibility and the main – Airbnb variables across the San Francisco MSA and the SF Francisco city, we aimed to address the following: Do short-term rentals in the form of Airbnb rentals impact the rent affordability of the study area and if so then to what extent? Overall, we find a significant correlation between the indicators Airbnb's presence (percent Airbnb listings as a fraction of total rental housing available and weighted Airbnb listings) and various rental affordability measures (rent-burdened households, rent overburdened households, median rents, and median house prices). While the correlation does not mean causation, the relative magnitudes of coefficients can be simulated for better understanding. The simulations from the spatial lag models (cross-sectional study) provide useful insights about the relationship between Airbnb and rental affordability. These also reveal that the tracts with lower percentages of Airbnb listings are more at risk of having low rental affordability as the presence of Airbnb increases.

This research also finds its motivation in connecting its findings with the regulatory debate. While a more comprehensive and deeper analysis is warranted for estimating a regulatory response (if any) to the platform, it is important to ascertain its effects both positive and negative. One critical dissection that should be acknowledged is that of the casual and commercial hosts. While the casual hosts stay true to the spirit of home-sharing i.e. utilization of underutilized/latent resources to support their incomes, the billion-dollar company seems to find its major revenue source in the commercial hosts (sometimes the super hosts) who own multiple properties and have entire house listings available throughout the year. Such hosts could have been landlords as part of the long-term rental housing stock. In a bid to investigate the revenue potential of Airbnb hosts as opposed to becoming a landlord in the same area, one can compare certain numbers available. While this analysis is not extensive, the main purpose of it is to understand how lucrative it is for the hosts to rent with Airbnb, rather than putting their property for long-term renting. The data shown is collected from AirDNA, (a paid service that provides Airbnb data analytics) mainly created to assist hosts to decide on their listing prices and better understand revenue trends. This study used the free data points that are available on the website i.e. county-wise average revenue earned by Airbnb hosts monthly (for 2017). The vacancy rates assumed for these estimates were not disclosed by the service. These revenue amounts can be compared with the gross median rent in that county to understand the difference between renting for long term vs short term. Table 5 shows the difference between renting with Airbnb (short-term) and renting lease based for long-term.

County	**Avg. monthly revenue**	**Median gross rent**	**Ratio (revenue/rent)**	**Percent entire home listings**
San Francisco	$ 3,107	$ 1,784	1.74	59%
Alameda	$ 2,155	$ 1,622	1.33	59%
San Mateo	$ 2,375	$ 2,114	1.12	42%
Marin	$ 2,298	$ 1,921	1.20	60%
Contra Costa	$ 1,466	$ 1,692	0.87	37%

Table 5: Comparison of revenue earned as an Airbnb host to that earned as landlord.

It can be observed that except Contra Costa County, all other counties present lucrative options for the Airbnb hosts to rent short-term instead of long-term. However, it should also be noted that the highest earning hosts are the commercial or the super hosts who host multiple entire home listings.

Hence, an important consideration for regulating Airbnb and like platforms are targeted policies ensuring proliferation of casual hosts who make tourism more affordable, benefit local business and empower homeowners by providing extra income but keeping a check on the commercial hosts who can potentially skirt hotel taxes and regulations by participating in the 'sharing' economy model. This presents as an opportunity for planners to evaluate their policies and development controls to better respond to Airbnb and the sharing economy. These can include better zoning provisions, extensive research and analysis of neighborhoods with high Airbnb listing concentrations, keeping an eye out for gentrification indicators and affordability concerns etc. Incentives can be promoted for the casual hosts who in way stay true to the idea of sharing economy. One major challenge in this process is to establish data sharing between Airbnb and planning agencies to better gauge the impacts and extent of the model. This leaves a lot of scope for research in terms of assessing neighborhood level impacts, negative and positive externalities of increase in Airbnb listings in a certain area and the need and type of regulation needed to accommodate Airbnb and sharing economy in general.

References

1. Anselin, L. (2017), *Spatial Data Science.* [Video file] Retrieved from `www.youtube.com/channel/UCzvhOfSmJpRsFRF2Pgrv-Wg`
2. Aurand, A. (2016). Out of Reach 2016: No Refuge for Low Income Renters. *National Low-Income Housing Coalition.* Retrieved from: `http://nlihc.org/sites/default/files/oor/OOR_2016.pdf`
3. Barron, K., Kung, E. and Proserpio, D. (2018) *The Sharing Economy and Housing Affordability: Evidence from Airbnb.* Retrieved from SSRN: `https://ssrn.com/abstract=3006832`
4. Budget and legislative analyst (2016) Policy Analysis Report, Board of Supervisors, *City and County of San Francisco.* Retrieved from: `https://sfbos.org/sites/default/files/FileCenter/Documents/55575-BLA.ShortTermRentals%20040716.pdf`
5. Bensinger, G. (2017) Airbnb Valued at $31 Billion After New Funding Round; Home-sharing site adds $1 billion cash cushion to stave off IPO, *Wall Street Journal* (Online); New York, N.Y. Retrieved from: `https://www.wsj.com/articles/airbnb-valued-at-31-billion-after-new-funding-round-1489086240`
6. Codagnone, C. and Martens, B. (2016). Scoping the Sharing Economy: Origins, Definitions, Impact and Regulatory Issues. *Institute for Prospective Technological Studies Digital Economy Working Paper 2016/01.* DOI: JRC100369
7. Codagnone, C., Biagi, F., and Abadie, F. : The Passions and the Interests: Unpacking the 'Sharing Economy'. *Institute for Prospective Technological Studies, JRC Science for Policy Report EUR 27914 EN,* DOI:10.2791/474555
8. Gutierrez, J., Garcia, J., Romanillos, G. and Olmedo. M. (2016). The eruption of airbnb in tourist cities: comparing spatial patterns of hotels and peer-to-peer accommodation. *Tourirm Management, Elsevier Ltd.* Retrieved from: `http://isiarticles.com/bundles/Article/pre/pdf/148944.pdf`
9. Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, 10(12), 1192-1217.
10. Hooijer, P. (2016). The Relationship between Airbnb and the Hotel Revenue: Evidence from The Netherlands. *University of Amsterdam.* Retrieved from: `http://www.scriptiesonline.uba.uva.nl/608470`
11. Kaplan, R. A. (2015). Airbnb: A Case Study in Occupancy Regulation and Taxation. *The University of Chicago Law Review Dialogue*, p103-115.
12. Lee, D. (2015). How Airbnb Short-Term Rentals Exacerbate Los Angeles's Affordable Housing Crisis: Analysis and Policy Recommendations. 10, 229-253.
13. Malhotra, A., and Van Alstyne, M. (2014). The dark side of the sharing economy and how to lighten it. *Communications of the ACM*, 57, 24-27.
14. Meni, D. (2017, March 21). Are Airbnbs driving up your rent? Retrieved from: `https://ggwash.org/view/62774/are-airbnbs-driving-up-your-rent`

15. Gurran, N. and Phibbs, P. (2017) When Tourists Move In: How Should Urban Planners Respond to Airbnb?, *Journal of the American Planning Association,* 83:1, 80-92, DOI: 10.1080/01944363.2016.1249011

16. O'Sullivan, F. (2015, December 23). Tourist-Heavy Barcelona Is Cracking Down on Airbnb. Retrieved from: `https://www.citylab.com/equity/2015/12/barcelona-airbnb-tourism/421788/`

17. PricewaterhouseCoopers (PwC) report. (2014). The sharing economy: How will it disrupt your business? Retrieved from – `http://www.pwc.co.uk/economic-policy/index.jhtml`

18. Quattrone, G., Proserpio, D., Quercia, D., Capra, L., and Musolesi, M. (2016). Who Benefits from the "Sharing" Economy of Airbnb? *Cornel University Library* Retrieved from: `https://arxiv.org/pdf/1602.02238`

19. Shaheen, S., Sperling, D., and Wagner, C. (1999). A Short History of Carsharing in the 90's. *Journal of World Transport Policy and Practice*, 5(3), 18-40.

20. Shephard, S. and Udell, A. (2016). Do Airbnb properties affect house prices? *New York State Office of the Attorney General,* supra note 42 at 2, 10-11. Retrieved from: `https://ag.ny.gov/pdfs/AIRBNB%20REPORT.pdf`

21. Streitfeld, D. (2014, October 15). Airbnb Listings Mostly Illegal, New York State Contends. *New York Times.* Retrieved from: `https://www.nytimes.com/2014/10/16/business/airbnb-listings-mostly-illegal-state-contends.html`

22. Stulberg, A. (2016). Airbnb Probably Isn't Driving Rents Up Much, At Least Not Yet. Retrieved from: `https://fivethirtyeight.com/features/airbnb-probably-isnt-driving-rents-up-much-at-least-not-yet/`

23. Zervas, G., Byers, J. and Proserpio, D. (2016). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. Retrieved from: `http://people.bu.edu/zg/publications/airbnb.pdf`