Is Airbnb Good for Housing Values?

Jonathan Eng

December 16, 2020

Abstract

This paper analyzes the effect of short-term rentals on neighborhood housing prices. Using a cross section of data between 2013-2015 from *NYC Open Data* and *InsideAirbnb*, we will measure the individual impact of housing factors on its sale price, taking note of the rental concentration variable. We estimate models that include neighborhood fixed effects and other influential variables and find that a 0.165 percentage-point increase in the concentration of short-term rentals leads to a 5.049% increase in the housing sale price.

Introduction

As housing prices continue to increase, renting one's home or room through Airbnb and other short-term rental methods, have become an increasing popular source of income. For example, *House Hacking* has become a common financial strategy for the younger generation, where one would rent a portion of their house to offset the cost of mortgage and other living expenses. Additionally, rental ability is an important factor for those interested in investing in real estate.

Airbnb is effective in giving homeowners the ability to generate revenue but raises concerns about its economic effect on the housing market. In 2014, it was found in New York City that 72% of New York City Airbnb listings were illegal due to violating property use and safety laws (Schneiderman 2014). Due to its simplicity and ability to generate income, it is viewed that Airbnbs often violate of health and safety laws. This leads to unsafe apartments and homes, an increase in traffic in quiet neighborhoods, and an increase in people who do not care about the neighborhood due to their short-term visit. All these factors are predicted to contribute to a decrease in housing values in the neighborhood. On the other hand, if these negatives are controlled through government and state policies, it is expected that the relative effect from the ability to generate additional revenue will increase home prices as the concentration of Airbnb's increases. We will explore the idea of whether an increase in Airbnb concentration has an impact on the sale price of homes, and if so, will it increase or decrease.

In this model, Airbnb concentration represents the concentration of short-term rentals. We attempt to predict selling prices for homes in New York City measured in USD by importing home descriptions, including variables such as neighborhood, residential units, gross square feet, etc., and analyze it using different algorithms to determine how influential the variables are. The raw data is modified to include only relevant data, with missing or insufficient data removed. The model includes linear regression models by incorporating Ordinary Least Squares.

We find that an increase in Airbnb concentration of 0.165 percentage-points leads to a 5.049% increase in housing price, after accounting for different home price influences. In addition, we find Airbnb concentration, square footage, and many of the neighborhoods to be statistically significant, at the 1% level, in determining the sale price of the home, indicating the high importance of the variables.

Literature Review

Airbnb's have increase in popularity and promotes the usage of short-term rentals, leading to an increase in concern in its effects on cities and urban housing markets. Using data from *InsideAirbnb* for New York City, Sheppard and Udell (2016) finds that a 100 percentage-point increase in Airbnb listings is associated with a 17.7% increase in housing values. This is due to the availability for an additional income stream making home ownership more valuable. Public policies to reduce the effectiveness of short-term rentals to increase home affordability appeared to be ineffective and prices continued to increase, however measuring all the consequences of regulatory action can be difficult to quantify.

When observing the Norwegian real estate market between 2016-2019 using data from *AirDNA*, Myrland and Pedersen (2020) conclude that a 1 percentage point increase in Airbnb rentals increase housing sale prices by 0.004%. Though small, it was found to be significant at the 5% level. Instead, the main driving factor in an increase in housing sale price is the household income, with a 1% increase in income raising sale prices by 1-4% within three years.

This is because higher income equals higher purchasing power, resulting in rising prices due to people in general having more money to spend.

Using United States Housing and Airbnb data extracted from *Zillow*, Barron, Kung, and Proserpio (2020) observed that a 1 percentage-point increase in Airbnb listings is associated with an average increase of 0.018% in house prices due to two reasons. Home sharing directly increases the house price because it enables homeowners to generate income from excess housing capacity that would otherwise be wasted, raising the value of owning a home relative to renting. Secondly, as short-term rentals become more popular, homeowners are incentivized to follow the market and switch from long-term rentals to short-term rentals, further increasing the rental-rate concentration which is then funneled into the housing price.

Data

The raw data is from 2013-2015, taken from *NYC Open Data*, a free public data source published by NYC agencies, and *InsideAirbnb*, an independent and non-commercial toolset of Airbnb data. *NYC Open Data* provided details on home sales in NYC and their features. *InsideAirbnb* data was used to count the number of postings within a neighborhood and year. It includes 453,150 observations with 24 descriptive variables for each home sold. This includes important variables such as "*SALE.PRICE*" and "*GROSS.SQUARE.FEET*," however it also includes less important variables such as "*APARTMNET.NUMBER*." These less important features will be ignored. The data also contains both "*NEIGHBORHOOD*" and "*ZIPCODE*" which are equivalent where the neighborhood represents the area name correlating to the numeric zip code. "*ZIPCODE*" will be the feature that we ignore because "*NEIGHBORHOOD*" is more recognizable.

The data had to be altered to contain an easily measurable data type. "*SALE.DATE*" was changed to "*SALE.YEAR*" where the month and day was removed from the dataset and now only contains the year of sale. "*NEIGHBORHOOD*" was changed to a factor variable from a character variable, then separated into individual columns for each neighborhood to serve as neighborhood fixed effects. This process was done to all factor variables to isolate the effects of each factor level. "*SALE.PRICE*" can be easily manipulated in its current state because of the large range of prices of different home types that cannot be directly compared through price, such as apartments and houses. To account for this difference we alter the "*SALE.PRICE*" variable to "*LOG.SALE.PRICE*", where we take the log of the sale price to obtain the value. This helps show the price as a percentage influence of the variables on the sale price. This is also done to "*GROSS.SQUARE.FEET*" due to the large disparity in square footage of the home types.

Some factor variables have an insufficient amount of data to include. An insufficient amount of data could generate more noise in the model and cause extrapolation errors. The chosen minimum to be deemed a sufficient amount of data is 500 observations, though this can be altered to any observation amount desired. Factors that did not meet the observation requirement is removed from the model.

The variable "*SALE.PRICE*" is our Y-variable in the model, meaning it cannot contain any missing data, so all observations with null values for "*SALE.PRICE*" are eliminated.

Some observations contain extreme values compared to the rest of the data and are considered outliers that must be removed. The interquartile range (IQR) was used on the "*SALE.PRICE*" because its maximum was far from the mean and median. To determine what values were outliers, the IQR method eliminates values outside the 1.5*(25th Percentile) and 1.5*(75th Percentile) range.

Short-term rental concentration is contained within the variable “*PERCENT.AIRBNB*.” It is calculated by taking the ratio of the number Airbnb postings and the number homes sold, within the same neighborhood and year, multiplied by 100 to get a percentage. The minimum and maximum values are 0% and 1.32876% respectively, with an average of 0.04958%. Due to the large range, being 6 standard deviations, it is best to display the effects of short-term rental concentration through a 1 standard-deviation increase, or 0.165 percentage points (Table 1).

After cleaning the data, the dataset contains 68,699 observations and 9 variables, including the housing sale price.

Methodology

The data contain 453,150 homes that were sold in New York City, obtained by merging the data of homes sold during the years 2013-2015, with 24 features of each home. After cleaning the data, we are left with 68,699 homes with 9 features of each home. Let *Y* represent the *log of Housing Sale Price (log(Housing_Sale_Price))* for a home in a specified neighborhood in a given year, denoted log(p). The goal of the regressions is to measure the impact of short-term rental concentration on Housing Sale Prices. A popular and easily interpretable method would be a linear regression model, which can be estimated by Ordinary Least Squares (OLS).

The following OLS model will predict log(p):

$$\log(p) = \alpha + \langle \beta_1 \, Percent.Airbnb \rangle + \langle B_2X_{log\_Gross\_Square\_Feet} + B_3X_{Residential\_Units} \rangle + \langle B_4X_{Tax\_Class} + B_5X_{Building\_Class\_Category} \rangle + \langle B_6X_{Year\_built} \rangle + \langle B_7X_{Sale\_Year} \rangle + \langle B_8X_{Neighborhood} \rangle + \varepsilon$$

Each vector indicates a new model that is built upon the previous vector(s). *Percent.Airbnb* represents the percent of homes within a neighborhood that are available for rent. This feature is the most important to us, as it directly relates to discovering the impact of Airbnb concentration on housing prices. A one percentage point increase in Airbnb Listings increases predicted Sales price by (100* β_1) percent. The variable α represents the intercept, which currently is not partially interpretable, and ε represents the error within the model. β_1 and ε are consistently changing as more variables are introduced into the models.

$X_{log\_Gross\_Square\_Feet}$ is the *log of Gross Square Feet (log(Gross_Square_Feet))*, measuring the land size of the building area, and $X_{Residential\_Units}$ represents the number of units within a building. These variables were chosen to highlight the impact of a home's features on its sale price.

X_{Tax_Class} represents the type of building being observed, such as apartment building versus house, and $X_{Building\_Class\_Category}$ represents the infrastructure and technology of the buildings. These variables are factors and are turned into independent variables for each level pre-regression. This creates additional variables equal to the amount of levels within each variable. These variables are able to account for the impact on the type of home on its sale price. A single column is removed from each factor variable and is valued within α, which is now more easily interpretable because of this change.

X_{Year_built} represents the year the home was built, measuring the impact of time appreciation on the home. This variable is a factor and will also have its levels as independent variables pre-regression with a single year removed to be accounted for in the intercept.

X_{Sale_Year} represents the year the home was sold, which allows us to compare the pricings to other homes sold within the same year, as well as the potential house appreciation if sold during a different year. This feature is also a factor and will be given the same adjustments as the previous factors.

$X_{Neighborhood}$ is a factor variable that accounts for the different neighborhoods. It will be given the same adjustments as the previous factors, representing the fixed effects in each neighborhood, such as neighborhood attractions and wealth. This variable will introduce the largest number of dummy variables, representing 69 different neighborhoods, one of which will be removed to be interpreted in the intercept.

Additional models were tested to see if an increase in complexity of the continuous variables would decrease the standard error within the model by allowing for a potential quadratic relationship between the variables and log(p). We have the following two potential models:

$$(1)\ log(p) = \alpha + \beta_1 Poly(Percent.Airbnb, 2) + \beta_2 Poly(X_{log\_Gross\_Square\_Feet}, 2) + \beta_3 Poly(X_{Residential\_Units}, 2) + \beta_4 X_{Tax\_Class} + \beta_5 X_{Building\_Class\_Category} + \beta_6 X_{Year\_built} + \beta_7 X_{Sale\_Year} + \beta_8 X_{Neighborhood} + \varepsilon$$

$$(2)\ log(p) = \alpha + \beta_1 Poly(Percent.Airbnb, 3) + \beta_2 Poly(X_{log\_Gross\_Square\_Feet}, 3) + \beta_3 Poly(X_{Residential\_Units}, 3) + \beta_4 X_{Tax\_Class} + \beta_5 X_{Building\_Class\_Category} + \beta_6 X_{Year\_built} + B_7 X_{Sale\_Year} + \beta_8 X_{Neighborhood} + \varepsilon$$

Poly magnifies the effect of the variables and allows for variable curvature in the event that a linear prediction is not as accurate as a quadratic prediction. The decrease in standard error is too insignificant to bear the negatives of the increasing the model's complexity, as magnifying the variable effects could magnify errors on new data.

Regular Standard Error: 0.6431

Poly 2 Standard Error: 0.6428

Poly 3 Standard Error: 0.6430

Models 1-6 will be used to measure the dependent variables *log(p)*, with each model introducing additional variables to see its relative effect. Coefficient β_1 is important to note for *Percent.Airbnb* to determine if our hypothesis is correct.

Results

It is important to take note that our Y variable is logged, where each β_n coefficient will represent a percentage effect on Y, where a coefficient of 0.10 will indicate a 10% increase in the Housing Sale Price (approximately).

Table 2 represents the regression results from models 1-6, where each column indicates a new model with each model number correlating to the column number, and each row indicates a variable's effect on log(p) for a 1 unit increase of the variables measurement type.

In model 1, *Percent.Airbnb* has a coefficient of 0.652, which concludes for every percentage-point increase in *Percent.Airbnb*, the *HousingSalePrice* will increase by 65.2%. *Percent.Airbnb* is statistically significant at the 1% level as well. Although appearing to have a large impact on log(p), the maximum Airbnb concentration is 1.33%, and the average is 0.04%. Due to a 1 percentage-point being much larger than the average *Percent.Airbnb*, it may be preferable to use a 1 standard deviation of Airbnb concentration to display its relative effect on log(p). However, we are only on model 1 of 6 and it is possible our estimation can become more accurate with additional variables. The results will be adjusted to show the effects of a standard deviation increment in the final model.

Model 2 introduces $X_{log\_Gross\_Square\_Feet}$ and $X_{Residential\_Units}$. The coefficient for *Percent.Airbnb*, decreases to 0.377, but remains within the 1% level of significance. $X_{log\_Gross\_Square\_Feet}$ has a coefficient of 0.50, indicating a 1% increase in *Gross_Square_Feet*, or in other words 1 unit increase in *log(Gross_Square_Feet)*, will result in a 50% increase in the *HousingSalePrice.* For every 1 unit increase in the amount of *ResidentialUnits* within a building, there is found to be a negative correlation to the sale price, with a coefficient of -0.01. This coefficient appears to be incorrect, as additional units should increase a home's sale price, so we must continue to introduce more variables to prevent this estimation error. $X_{log\_Gross\_Square\_Feet}$ is within the 1% level of significance, however $X_{Residential\_Units}$ is within the 5% level of significance, which is still considered a trustworthy level of confidence.

Model 3 introduces X_{Tax_Class} and $X_{Building\_Class\_Category}$. The coefficient and level of significance varies for each factor level variable from X_{Tax_Class} and $X_{Building\_Class\_Category}$. Looking at the table, one can see the relative impact of the different variables. *Percent.Airbnb,* maintained a similar coefficient of 0.383, despite the introduction of more variables.

Model 4 introduces X_{Year_built} where majority of the years are significant at the 1% level. Interestingly, the coefficient for *Percent.Airbnb* increased to 0.431 and remains in the 1% level of significance.

Model 5 introduces X_{Sale_Year} which is significant at the 1% level for each dummy variable. X_{Sale_Year} drastically drops the coefficient of *Percent.Airbnb* to 0.324, but it still remains significant at the 1% level.

Model 6 introduces $X_{\text{Neighborhood}}$ which adds variables for every neighborhood to account for the neighborhood fixed effects. It drops the coefficient of $X_{PercentAirbnb}$ to 0.306, but it still remains significant at the 1% level. $X_{\text{Residential\_Units}}$ also switches to a positive coefficient from a negative coefficient. It is likely that properties with more units were located in less wealthy neighborhoods, and by accounting for unobserved neighborhood characteristics that are relatively constant over time, such as amenities associated with location in the fixed effects of each neighborhood, an increase in *ResidentalUnits* now shows a positive influence on the *HousingSalePrice*, which makes more sense than a negative correlation.

Model 6 is the final model and has the greatest number of variables that measure the *HousingSalePrice*. We observed that for every 1 unit increase in *Perecnt.Airbnb*, the *HousingSalePrice* increases by 30.6%. To represent the effect of *Percent.Airbnb* in a more sensible manner, we use the relative impact of a 1 standard deviation increase in *Percent.Airbnb*. The standard deviation for *Percent.Airbnb* is 0.165 percentage-points (Table 1). Using the values from model 6, we can conclude a 1 standard deviation increase in *Percent.Airbnb* of 0.165 percentage-points, increases *HousingSalePrice* by 5.049%, a better representation of the mean of *Percent.Airbnb*.

Percent.Airbnb remains significant at the 1% level for all models, indicating that short-term rental concentration plays a significant role in determining the sale price for homes. Additionally, we can observe that *Gross_Square_Feet* and *Neighborhood* are also important variables in determining *HousingSalePrice* due to their larger coefficients and significance at the 1% level.

Comparing my prediction to the literature used to aid in creation of our model, there is a large range of predictions among us due to the difference in variable choice and methodology, where my prediction concludes the largest impact for Airbnb percentage with a 0.165 percentage-point increase in Airbnb concentration leading to a 5.049% increase in housing price. By adjusting the literature results to be more comparable to increment of 0.165 percentage-points for *Percent.Airbnb*, we see the following observations: Sheppard and Udell (2016) found that a 0.1 percentage-point increase of Airbnb listings leads to a 0.0177% increase in housing value. Myrland and Pedersen (2020) conclude that a 0.1 percentage point increase in Airbnb rentals increase housing sale prices by 0.0004%. And Barron, Kung, and Proserpio (2020) observed that a 0.1 percentage-point increase in Airbnb listings is associated with an average increase of 0.0018% in house prices.

Although the predictions are different from each other, and much different from the results we concluded, it important to note we unanimously concluded a positive relationship between short-term rental concentration and housing value.

Conclusion

In this paper, we explain the impact of short-term rental concentration on housing sale price in New York City (NYC) by creating a model that predicts the selling prices of NYC homes and incorporating "*PERCENT.AIRBNB"* to represent short term rental concentration.

We found a strong correlation between "*SALE.PRICE*" and "*PERCENT.AIRBNB*." As we include additional variables in the model, "*PERCENT.AIRBNB*" remains significant at the 1% level. We also found *"LOG_GROSS.SQUARE.FEET"* and many "*NEIGHBORHOOD"* variables to be significant at the 1% level of significance. This allows us to conclude that Airbnb percentage, home square footage, and neighborhood are all important variables in determining the housing sale price.

In the final model, we find that an increase in *Percent.Airbnb* of 0.165 percentage-points is concluded to increase *HousingSalePrice* by 5.049%. The data gained from this paper answers the debate on whether short-term concentrations such as Airbnb has an impact on housing prices. An increase in Airbnb's concentration positively correlates to an increase in housing prices, but the extent of the positive impact remains debatable due to the different findings from the literature and our own observations.

The concerns for negative impacts for an increase in Airbnb concentration were not directly addressed, leaving out external influences such as visitors not caring for the neighborhoods and consequently decreasing housing prices. We can assume this feature is within the neighborhood fixed effects, but this may be a large assumption.

We know that increasing Airbnb concentration increases housing prices, so with government and state intervention to ensure health and safety laws are followed, and careful screening of Airbnb customers and hosts, the negatives can be reduced significantly enough to highlight the positive impact. Therefore, New York City should first create and police regulations that protect the neighborhood and its people from inappropriate uses of Airbnb, then encourage Airbnb rentals rather than reduce them, as this will appreciate housing prices in the neighborhood.

References

Barron, Kyle, et al. The Effect of Home-Sharing on House Prices and Rents: Evidence from Airbnb. 2020.

Myrland, Ørjan N., and Cathrine Pedersen. The Effect of Airbnb on Real Estate Prices. UIiT The Arctic University of Norway, 2020.

Schneiderman, E. T. (2014), 'Airbnb in the City', New York State Office of the Attorney General

Sheppard, Stephen, and Andrew Udell. Do Airbnb properties affect house prices? Williams College Department of Economics, 2016.

Appendix

Table 1: Descriptive Statistics:

Variable	OBS	MEAN	STD. DEV	MIN	MAX
NEIGHBORHOOD:		N/A	N/A	N/A	N/A
BEDFORD-STUYVESANT	3388				
FLUSHING-NORTH	3253				
BUSHWICK	2139				
…	…				
BUILDING.CLASS.CATEGORY:		N/A	N/A	N/A	N/A
ONE FAMILY DWELLINGS	28729				
TWO FAMILY DWELLINGS	28103				
THREE FAMILY DWELLINGS	7365				
…	…				
TAX CLASS:		N/A	N/A	N/A	N/A
1	64164				
2	344				
2A	3480				
2B	711				
RESIDENTAIL.UNITS:	68699	1.96	1.74	1	66
LOG_GROSS.SQUARE.FEET	68699	7.593	0.447	5.864	10.893
YEAR.BUILT:		N/A	N/A	N/A	N/A
1920	10254				
1930	8771				
1925	8455				
…	…				
LOG_SQUARE.PRICE	68699	13.14	0.5605	11.21	14.96
SALE.YEAR:		N/A	N/A	N/A	N/A
2013	27108				
2014	27478				
2015	14113				
PERCENT.AIRBNB	68699	0.04958	0.1639565	0	1.32876

Table 2: Regression Results of Multiple Models
("*Year.Built*" & "*Neighborhood*" Excluded from View)

Home Sale Price versus Short Term Rental Concentration

	Dependent variable:					
	LOG_SALE.PRICE					
	(1)	(2)	(3)	(4)	(5)	(6)
PERCENT.AIRBNB	0.652***	0.377***	0.383***	0.431***	0.324***	0.306***
	(0.012)	(0.012)	(0.011)	(0.012)	(0.013)	(0.012)
LOG_GROSS.SQUARE.FEET		0.504***	0.528***	0.545***	0.563***	0.386***
		(0.006)	(0.006)	(0.006)	(0.006)	(0.005)
RESIDENTIAL.UNITS		-0.012***	-0.031***	-0.038***	-0.038***	0.013**
		(0.002)	(0.008)	(0.008)	(0.008)	(0.006)
TAX.CLASS.AT.PRESENT2A			-0.061	-0.043	-0.032	-0.153***
			(0.063)	(0.063)	(0.062)	(0.048)
TAX.CLASS.AT.PRESENT2B			0.118	0.169**	0.185**	-0.054
			(0.074)	(0.073)	(0.073)	(0.057)
BUILDING.CLASS.CATEGORY02 TWO FAMILY DWELLINGS			-0.031***	-0.006	-0.012	0.007
			(0.009)	(0.009)	(0.009)	(0.007)
BUILDING.CLASS.CATEGORY03 THREE FAMILY DWELLINGS			-0.066***	-0.015	-0.023	-0.020
			(0.016)	(0.016)	(0.016)	(0.013)
BUILDING.CLASS.CATEGORY07 RENTALS - WALKUP APARTMENTS			0.149**	0.153**	0.136**	0.233***
			(0.063)	(0.062)	(0.062)	(0.048)
BUILDING.CLASS.CATEGORY14 RENTALS - 4-10 UNIT			0.139**	0.138**	0.116*	0.293***
			(0.063)	(0.063)	(0.062)	(0.049)
SALE.YEAR2014					0.066***	0.083***
					(0.004)	(0.003)
SALE.YEAR2015					0.174***	0.189***
					(0.005)	(0.004)
Constant	13.113***	9.329***	9.192***	9.019***	8.852***	9.829***
	(0.002)	(0.041)	(0.043)	(0.046)	(0.045)	(0.052)
Observations	67,385	67,385	67,385	67,385	67,385	67,385
R^2	0.041	0.188	0.193	0.202	0.216	0.524
Adjusted R^2	0.041	0.187	0.193	0.202	0.216	0.524
Residual Std. Error	0.512 (df = 67383)	0.471 (df = 67381)	0.470 (df = 67375)	0.467 (df = 67355)	0.463 (df = 67353)	0.361 (df = 67285)
F Statistic	2,893.304*** (df = 1; 67383)	5,183.342*** (df = 3; 67381)	1,788.179*** (df = 9; 67375)	588.522*** (df = 29; 67355)	598.918*** (df = 31; 67353)	748.907*** (df = 99; 67285)
Note:						*p<0.1; **p<0.05; ***p<0.01