Exploring influence factors of Airbnb penetration in three U.S. cities using spatial regression analysis

BT 4015 Geospatial Analytics

Group 6:
Chen Xinyu A0198779W
Huang Fengjue A0225070U
Ji Xin A0224989J
Wang Yunjie A0225017R

1. Introduction

1.1 Airbnb

Founded in 2018 in San Francisco, it quickly became a global phenomenon contributing significantly to the rise of the sharing economy as a new economic paradigm (Kaplan & Nadler, 2015; Sundararajan, 2014; Gurran & Phibbs, 2017), and now it has over 3,000,000 listings in more than 65,000 cities across the globe (Airbnb About Us. https://www.airbnb.co.uk/about/about-us. Cited August 2017). Airbnb allows people to rent their entire properties or spare rooms by engaging in computer-mediated transactions directly with potential guests. Instead of being based on centralized entities, this example of a peer-to-peer economic model is based on a distributed network of individuals who directly access each other underused assets.

Additionally, Airbnb's rapid growth has also raised a host of questions for cities and indigenous peoples:

Gentrification. When Airbnb infiltrates a city to a certain extent, it causes gentrification. When some surplus housing in the hands of indigenous people becomes a commodity that can be shared for rental, the local housing market prices will rise. Rich people can buy more than one house to rent, which aggravates the contradiction between supply and demand in the local housing market and harms indigenous people's interests (Wachsmuth, 2018).

Quality of local's life. As a result of a large number of cheap and convenient housing external noise pollution. The city's opening has led to a massive influx of people into cities, which is bound to bring more household garbage and noise pollution. To a certain extent, it will also affect the local security situation.

Lack of security regulations. Airbnb lacks uniform standards and rules for all rentals because the platform's houses are mostly private properties. As a result, it is difficult for regulators to require these homeowners to make modifications to meet hotel standards and protect consumers' declared property safety according to security, fire, and sanitation requirements.

1.2 Literature Review

Through literature reading, we find that Airbnb's previous work mainly focuses on the impact of Airbnb on the hospitality industry (Quattrone, 2018).

For Airbnb, scholars have begun to examine its relationship with hospitality's more traditional forms. That has added yet more nuance to the critical debate continuing to surround the platform. Zervas(2017) analyzed Airbnb listings in Texas and found that Airbnb had negatively affected local hotels' revenue: a 1% rise in the number of listings led to a 0.05% loss of hotel revenue. That, however, impacted lower-end hotels mainly and left untouched higher-end ones. They analyze Airbnb's entry into the state of Texas and quantify its impact on the Texas hotel industry over the subsequent decade. In Austin, where Airbnb supply is highest, the causal effect on hotel revenue is in the 8%–10% range; moreover, the impact is nonuniform, with lower-priced hotels and hotels that do not cater to business travelers being the most affected. Quattrone(2018) set up a set of relatively complete explanatory variables. They analyzed Airbnb's spatial distribution in eight U.S. urban areas concerning geographic, socio-demographic, and economic information.

1.3 Motivation

Previous studies on Airbnb mainly focused on a specific city. This study selected three representative cities of Airbnb in the United States for horizontal and vertical comparative analysis.

In previous studies, Airbnb's distribution has not been used as one of the explanatory variables for regression modeling. This study found that Airbnb's distribution has spatial autocorrelation, so we conducted regression modeling after processing the dependent variable.

Our work also goes beyond a descriptive analysis by establishing a global regression model to explore common factors affecting Airbnb penetration among different cities to provide reference suggestions for business owners and the government to formulate policies and regulations. If certain factors that affect Airbnb penetration show high values in certain regions, they can serve as a warning indicator that Airbnb penetration could occur in that region.

1.4 Research Question

In this study, we analyze the spatial penetration of Airbnb in three cities across the U.S., intending to answer three main research questions:

RQ1. What are the factors influencing Airbnb distribution in different cities? They can be a comprehensive dataset of geographic, social, and economic characteristics of urban areas.

RQ2. In the three selected research samples, whether there are common factors that influence the penetration of Airbnb?

RQ3. Whether it can provide a universal basis for the formulation of policies and regulations?

2. Data and metrics

2.1 Cities

In this study, we chose three different cities in the United States: Washington D.C., Los Angeles, and Chicago as the study area. There are many differences between the three cities in different aspects such as population, size, income, consumption level, and so on. In addition to this, we collected the boundary data of these three cities from the most recent U.S. Census Bureau*(https://data.census.gov/cedsci/advanced).*

Table 1 shows the data of these three cities in various social and economic characteristics in 2018, in terms of population, median age, median income, percent of white people, cost of living and Airbnb listings. We chose these cities because among of them have mature Airbnb presence and they vary substantially among each other in many aspects. We want to find whether there are some common influence factors for Airbnb penetration in different cities.

Table 1. *Summary characteristics of the 3 chosen U.S. cities in 2018*

City	Pop	Median Age	Median Income	% of White	Cost of Living	# Airbnb Listings
Washington D.C.	700k	34	$64k	42%	$18k	9123
Los Angeles	3.9m	37	$46k	51%	$17k	28687
Chicago	2.7m	37	$52k	57%	$17k	7555

All the data used in this study is collected by the tracts in the three cities. We chose to operate at the level of tracts, the smallest granularity at which the U.S. Census Bureau collates data because each tract has roughly the same population and we do not need to further normalization. (Quattrone, G, 2018)

After deleting tracts with a population of 0 and missing data, the number of tracts that we studied is 177 in Washington D.C., 1006 in Los Angeles and 797 in Chicago.

2.2 Airbnb data

In this study, we use the number of Airbnb listings in each tract to measure the Airbnb penetration. The Airbnb listing data is collected from InsideAirbnb (http://insideairbnb.com/), which is an independent, non-commercial Airbnb data website that publishes snapshots of Airbnb listings around the world (Quattrone, G, 2018). All of the Airbnb data we used in this study is from August, 2018. The data is in CSV format and contains columns of the coordinates of each Airbnb (longitude, latitude). For Chicago, Los Angeles, and Washington D.C., the total number of Airbnb listings are 7555, 28687, and 9123 respectively.

For the data processing, firstly, we convert the CSV data into spatial point data, then we count the numbers of Airbnb listing points in each tract in the three cities.

2.3 Metrics

In order to explain the Airbnb penetration in Chicago, Los Angeles, and Washington D.C., we used three different groups of variables as the metrics based on the literature review, capturing the geographic, demographic and socio-economic contexts in the three cities. We collects POI of bus stops, hotels, retail buildings, and arts attractions from OpenStreetMap (https://www.openstreetmap.org/) and the demographic data as well as socio-economic data is from U.S. Census Bureau (https://data.census.gov/cedsci/advanced) (*Table 2*).

Table 2. *Summary of the metrics*

Category	Name	Definition
Geographic	Bus Stops	Number of bus stops in an area
	Hotels	Number of hotels in an area
	Retails	Number of retails in and area

Category	Name	Definition
	Arts Attractions	Number of arts and human landscapes in an area
Demographic	Race Diversity Index	Diversity of races in an area
	Proportion of Young People	Proportion of people aged between 20 and 34 to the total population of the area
	Population Density	Density of resident population in an area
Socio-economic	Income Diversity Index	Diversity of income in an area
	Bohemian Index	Proportion of people employed in arts, entertainment and media in an area to the same proportion nationwide
	Talent Index	Proportion of people in an area with degrees higher than an associate degree
	Unemployment Ratio	Proportion of unemployed population to the total population
	Poverty Rate	Proportion of population determined to be poverty to the total population
	Median Household Income	Median estimate of household income in an area
	Median Household Value	Median value of a household in an area
	Proportion of owner-occupied residences	Proportion of dwellings that are owned to those that are occupied

For data processing, we count the number of each kind of POI in each tract, and associate the number of POI, the demographic data and socio-economic data with the correspondent tracts

2.4 Descriptive analysis of the data

Distribution of Airbnb Penetration. Figure 1- 3 show the distribution of Airbnb penetration in the three cities (Figure 1, Figure 2 and Figure 3). The three maps show a clustered patterning of Airbnb penetration in the three cities. And we found all the Airbnb penetration in these three cities cluster in the densely populated city centers.

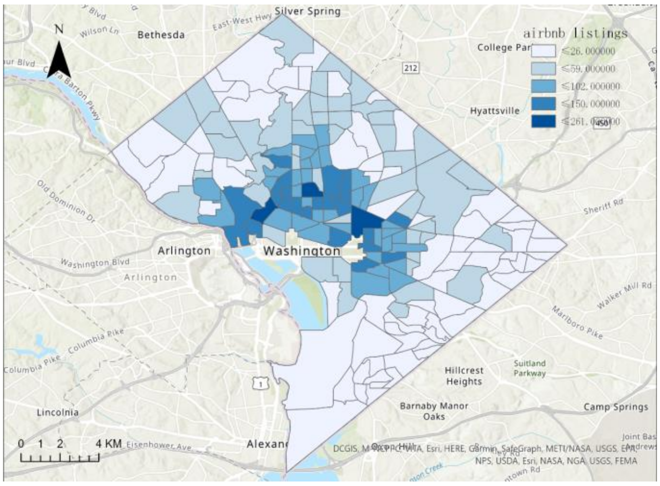


Figure 1. *Airbnb penetration of Washington D.C. in 2018*

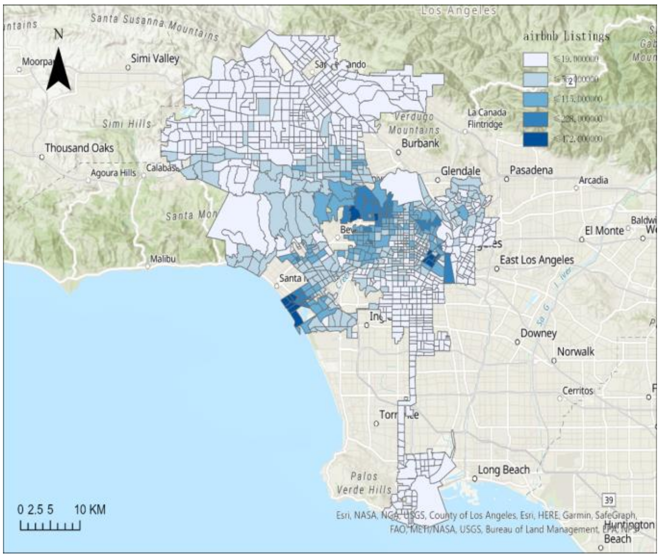


***Figure 2**. Airbnb penetration of Los Angeles in 2018*

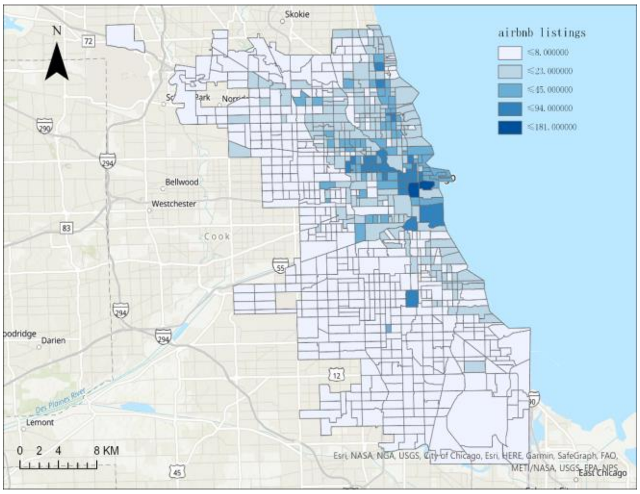


Figure 3. *Airbnb penetration of Chicago in 2018*

Summaries of the metrics. Figure 4-6 are the box plots of the metrics of the three cities. The tracts in each city is split into more and less Airbnb areas based on whether the number of Airbnb listings in the tract is higher than the median. By comparing the two box plots in each city, we can roughly infer which variables might be related to Airbnb penetration. In Washington D.C., Race Diversity, Proportion of Young People, Talent Index and Poverty Rate show great differences between more-Airbnb areas and less-Airbnb areas. In Los Angeles, Talent Index and Median Household Value might be related to Airbnb penetration. In Chicago, Race Diversity, Talent Index and Unemployment Rate might be related to Airbnb penetration.

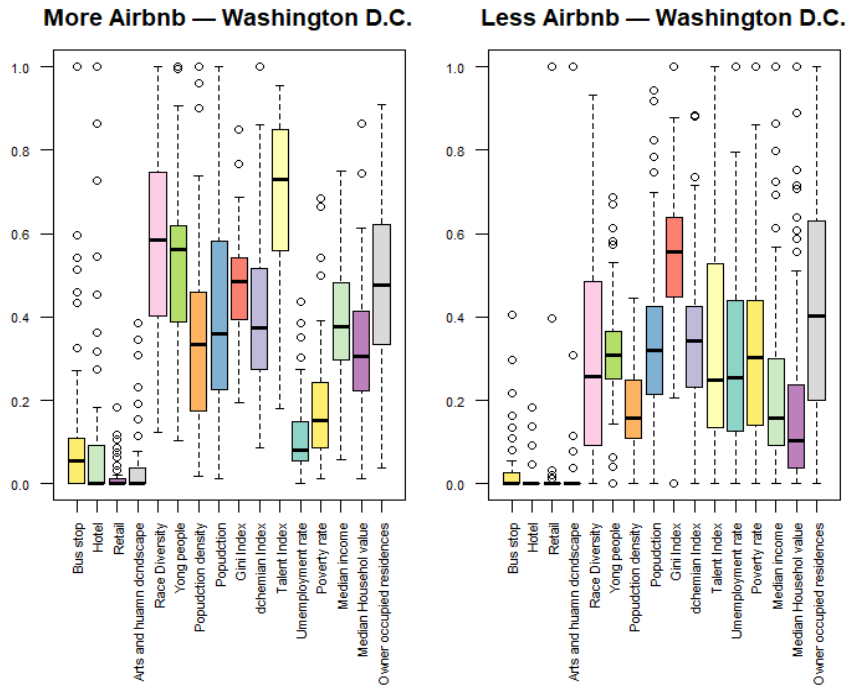


Figure 4. *The box plots of explanatory variables in Washington D.C.*

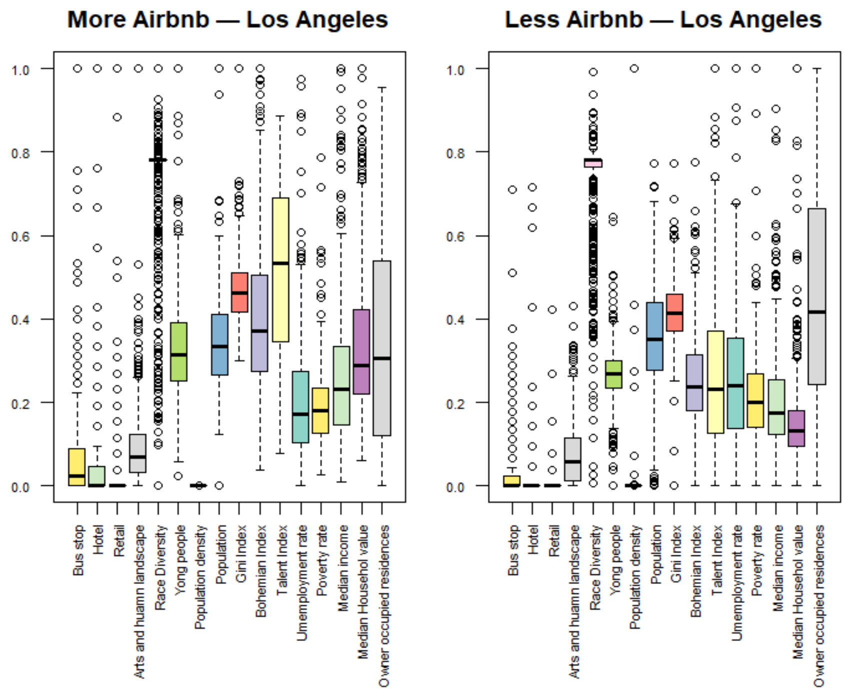


Figure 5. *The box plots of explanatory variables in Los Angeles*

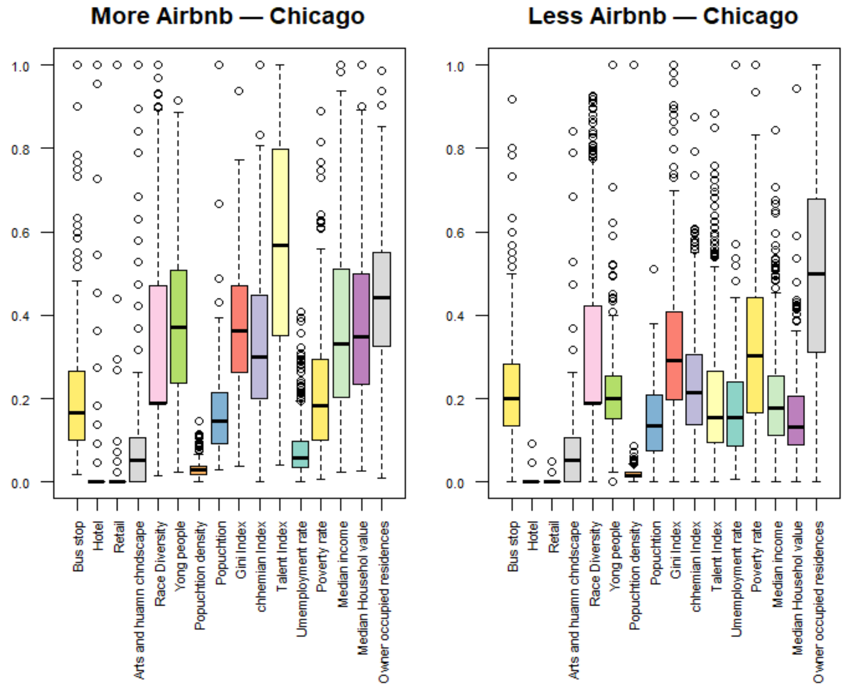


Figure 6. *The box plots of explanatory variables in Chicago*

3. Method

3.1 Workflow of spatial analysis

The workflow of this study is shown in Figure X. In this study, we first use Global Moran/s I to detect the spatial autocorrelation and identify the distribution pattern of Airbnb penetration in the three cities. Then, exploratory analysis is conduced to explore the influence factors of Airbnb penetration in the three cities. Two kinds of regression analysis are used here: we begin by OLS regression analysis and use spatial regression analysis to solve the problem of spatial autocorrelation in the variables. Lagrange Multiplier test is used to choose the more appropriate spatial regression model between spatial error model (SEM) and spatially lagged model (SLM).

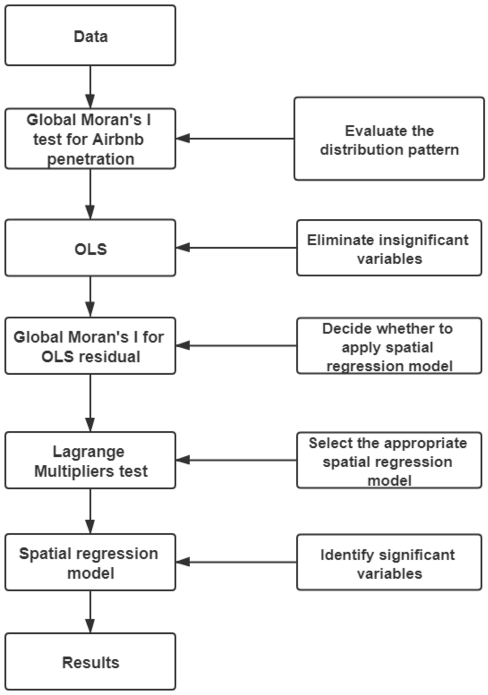


Figure 7. *Workflow of spatial analysis*

3.2 Global Moran's I for Airbnb penetration

The Moran's I statistic is the correlation coefficient for the relationship between a variable and its surrounding values. In this study, Global Moran's I is used to detect the spatial autocorrelation in Airbnb penetration in the three cities, and Queen's method is used as the neighbor selection method.

3.3 Ordinary least squares (OLS) regression analysis

Our research question investigates whether it is possible to explain Airbnb spatial penetration using geographic, social and economic variables. To explore the variables that are significantly correlated with Airbnb penetration and to what extent, we begin by conducting a multivariate linear regression model in the form of Ordinary Least Squares (OLS) (*Formula (1)*).

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon \qquad (1)$$

where y denotes the Airbnb penetration in a given area; $x_1, \ldots, x_k$are the metrics that reflects the geographic, socio-economic, and demographic characteristics of the same area (***Table 2***); $\beta_0, \beta_1 \ldots, \beta_k$ are the unknown parameters; and ϵ is the error term.

Since the Airbnb penetration and many of the metrics are skewed and therefore do not conform the normality assumption of the OLS, we apply a log transformation for both of the dependent variable and independent variables. Besides, since we want to compare the coefficients of different metrics, a z-score normalization is performed to scale the independent variables.

When conducting OLS regression analysis, all of the variables in *Table 2* are used to run the regression first. Then we conduct VIF test to detect the multicollinearity among the variables and remove the variables that are strongly correlated with others. Besides, those variables which are insignificant would also be removed from the regression model. After the variables selection, the model with all of the significant independent variables that are not strongly correlated with each other would be selected as the OLS model for Airbnb penetration in each of the city.

3.4 Global Moran's I test for OLS models

OLS regression assumes that what happens in area X_i is independent of what happens in area X_j. However, since we are dealing with geographic data, this assumption of OLS might be violated because of Tobler's First Law of Geography, meaning that the Airbnb penetration may tend to be geographically clustered. Hence, we would conduct Moran's I test for the OLS models we built in the last step, to diagnose the assumptions for OLS regression. Unlike the global Moran's I for Airbnb penetration in section 3.2, this Moran's I test is for the residuals of OLS model.

3.5 The Lagrange Multipliers test

After detecting the spatial autocorrelation in the OLS model, a spatial regression model would be use to solve the problem of spatial autocorrelation. There are two generally used spatial regression model: spatial error model (SEM) and spatially lagged model (SLM). In this case, the Lagrange Multiplier test is used to decide which one is a more appropriate one.

The workflow of choosing an appropriate model is shown in *Figure 8*. First, we look at the standard Lagrange Multiplier tests (LMerr and LMlag). If both of them are below the 0.05 level, this means we need to have a look at the robust version of these tests (Robust LM). We would choose a spatial lag or a spatial error model according to Robust LM if either of them is significant. However, if both of the Robust LMs are significant, we would choose the one which is orders of magnitude more significant than the other (e.g., p < 0.00000 compared to p < 0.03).

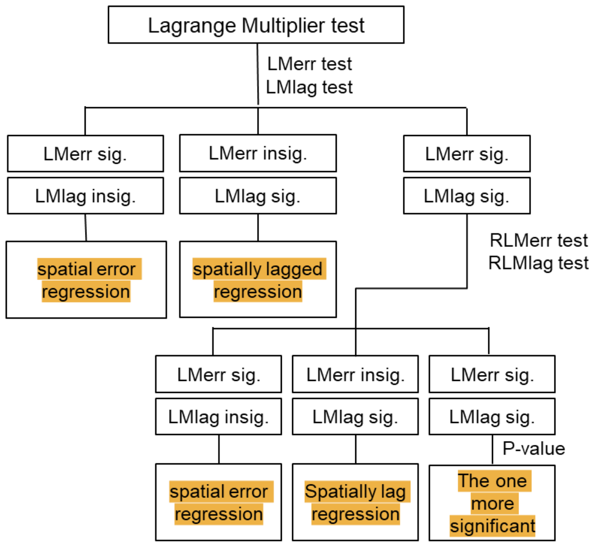


Figure 8. *The workflow of choosing a spatial regression model*

3.6 Spatial regression model

There are two general ways of incorporating spatial dependence in a regression model: a spatially lagged model and a spatial error model. The difference between these two models is both conceptual and technical.

3.6.1 Spatially lagged model

From the conceptual perspective, spatially lagged model treats spatial autocorrelation as a truly spatial process such as the diffusion of behaviour between neighboring units. For example, Airbnb would choose to locate in somewhere because the clustering of Airbnb can bring agglomeration economy and attracts more customers; and the owner of Airbnb might also communicate with each other, which also influence the Airbnb site selection.

From the technical perspective, spatially lagged model incorporates spatial dependence explicitly by adding a "spatially lagged" variable y on the right-hand side of the regression equation (Formula 3). Its distinctive characteristic is that it includes a spatially lagged "dependent" variable among the explanatory factors, saying that the values of y in the neighboring areas of observation n_i is an important predictor of y on each individual area n_i.

$$y = \rho W_y + X\beta + \epsilon \qquad (2)$$

where y denotes the Airbnb penetration in a given area. W_y is a spatial weight matrix of Airbnb; ϵ is the error term. β and ρ are the unknown parameters

3.6.2 Spatial error model

From the conceptual perspective, the spatial error model treats the spatial autocorrelation as a nuisance that needs to be dealt with. It implies that the spatial dependence observed in the data does not reflect a truly spatial process, but merely the geographical clustering of the sources of the behaviors of interest. For example, Airbnbs cluster together not because they interact with each other, but because the Airbnbs usually want to locate in places with certain kinds of characteristics, for example, high population density, convenient transportation, and high median income.

From the technical perspective, instead of assuming that a spatial lag influences the dependent variable, it is a model that relaxes the standard regression model assumption about the need for the errors to be independent (*formula 3 - 4*).

$$y = X\beta + \epsilon \qquad (3)$$

With

$$\epsilon = \lambda W_\epsilon + u \qquad (4)$$

where y denotes the Airbnb penetration in a given area. W_ϵ is a spatial weight matrix of error term; ϵ is the error term. β , u and λ are the unknown parameters

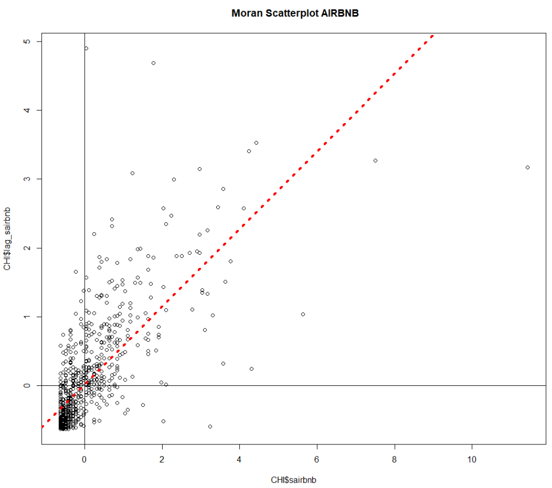
4. Results

4.1 Global Moran's I for Airbnb penetration

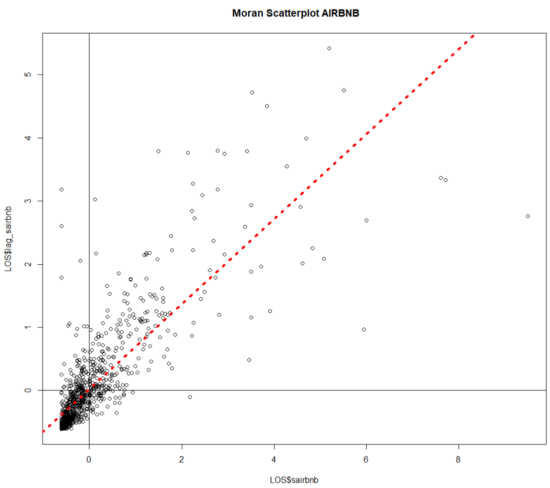
Table 3. *Moran's I test for Airbnb penetration under randomization*

City	Standard Deviate	P-value	Moran's I
Chicago	28.701	< 2.2e-16	0.564376481
Los Angeles	36.087	< 2.2e-16	0.673660941
Washington D.C.	14.156	< 2.2e-16	0.607409271

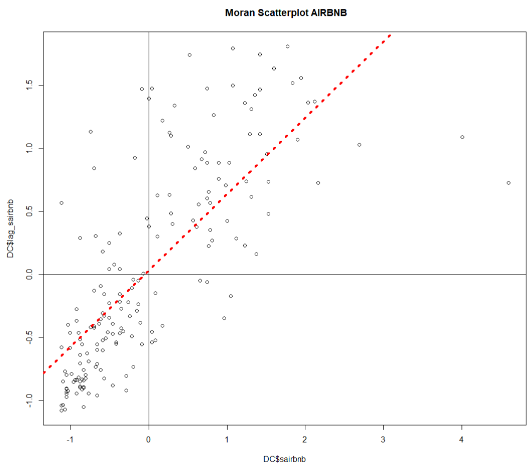
Data: Airbnb listings in each tract (Chicago, Los Angeles, Washington D.C.)



(a) Chicago



(b) Los Angeles



(c) Washington D.C.

Figure 9. *Moran's I scatter plot*

We test the Moran's I in each of the three cities. From the Moran's I test and scatter plot, we can see that the Airbnb distribution in the three cities selected in the study presents the phenomenon of spatial autocorrelation. Therefore, Airbnb's distribution should also be considered as one of the explanatory variables in regression modeling. It is the model constructed from this that is explanatory.

4.2 Ordinary least squares (OLS) regression analysis

After detecting the spatial autocorrelation of Airbnb penetration in the three cities, we begin by conducting ordinary least square (OLS) regression analysis to explore the influencing factors of Airbnb penetration in the three cities. First, we use all of the potential explanatory variables in *table 4* to conduct OLS regression analyses for the three cities. Then we test the multicollinearity in the three models and remove the variables which are strongly correlated with others and the variables which are not significant from the models. After the variable selections, we obtain the three OLS regression models, showing the variables associated with Airbnb penetration in the three cities (*table 4*). Table X shows that each of the variable in the three models has a p-value under 0.05 and a VIF under 4, and the adjusted R^2 of the three models are higher than 0.6. However, the results may not be significant if they present evident spatial autocorrelation. Hence, we would like to diagnose the assumptions for OLS regression by exploring the residuals.

Table 4. *OLS regression analysis for Airbnb penetration in the three cities*

(a)Chicago

Variables	β	P-value	VIF
Bus Stop	0.064	0.012*	1.38
Hotel	0.074	0.001**	1.16
Bohemian Index	0.104	1.58e-05 ***	1.25
Talent Index	0.360	<2e-16 ***	2.94
Owner Occupied Residences (%)	-0.219	<2e-16 ***	1.35
Median Household Value	0.299	1.82e-15 ***	2.97
Population	0.088	0.001***	1.39
Young people (%)	0.158	2.76e-07***	2.03
Adjusted R-squared		**0.643**	

(b) Los Angeles

Variables	β	P-value	VIF
Bus Stop	0.064	0.012*	1.38
Hotel	0.074	0.001**	1.16
Bohemian Index	0.104	1.58e-05 ***	1.25
Talent Index	0.360	<2e-16 ***	2.94
Owner Occupied Residences (%)	-0.219	<2e-16 ***	1.35
Median Household Value	0.299	1.82e-15 ***	2.97
Population	0.088	0.001***	1.39
Young people (%)	0.158	2.76e-07***	2.03
Adjusted R-squared		**0.643**	

(c) Washington D.C.

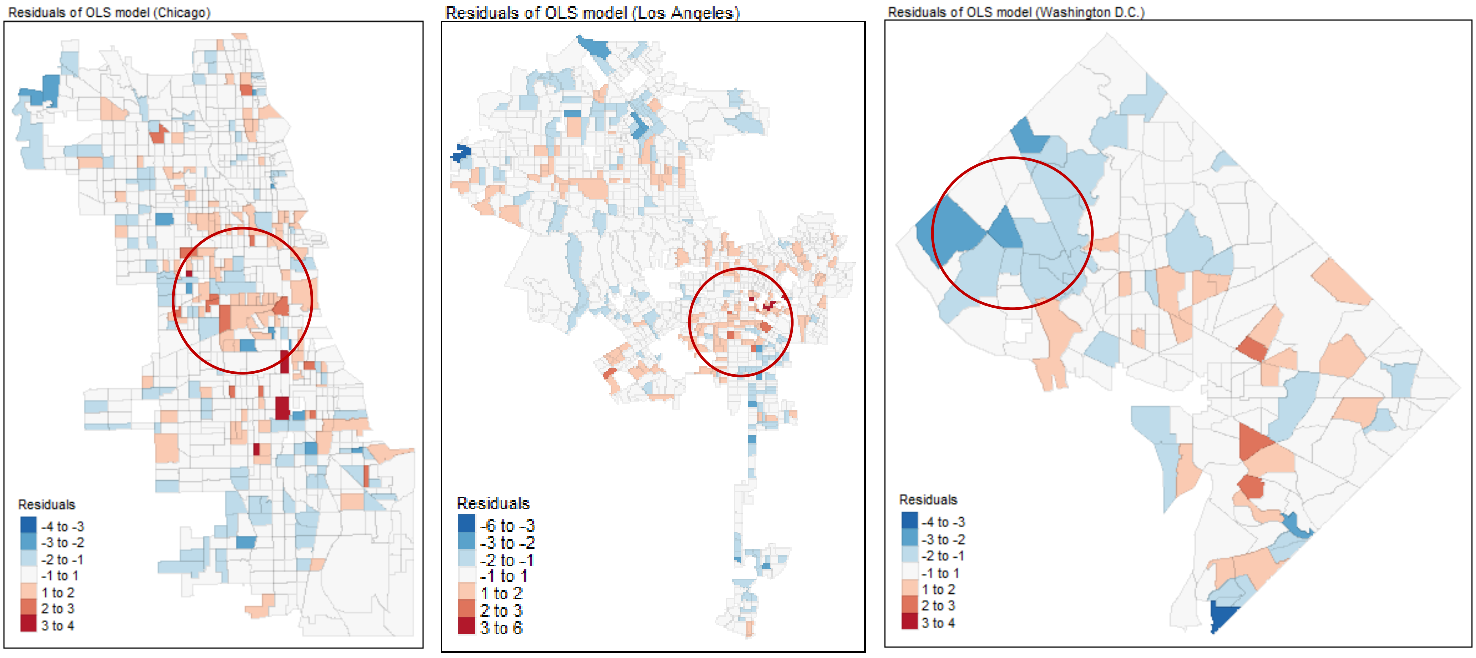
Variables	β	P-value	VIF
Bus Stop	0.095	0.025*	1.28
Unemployment Rate	-0.167	0.007**	2.73
Talen Index	0.302	8.08e-05 ***	3.93
Owner Occupied Residences (%)	0.190	0.000***	1.93
Population	0.108	0.005**	1.07
Young people (%)	0.376	5.77e-12 ***	1.88
Adjusted R-squared		**0.7092**	

Geographic indicators   Socio-economic indicators   Demographic indicators

4.3 Moran's I test for OLS model

We produce three choropleth maps for the residuals of the three models in table 4 (*figure* 10). We use standard deviations as the classification method to show how much the residuals differs from the mean of the residuals, and each standard deviation becomes a class in the choropleth maps. The negative residuals (blue tones) show that the area is over predicted, while the positive residuals (red tones) show that the area is under predicted. In Figure 10, we notice that there is spatial patterning exist in the three choropleth maps, for example, some over-predicted/under- predicted cluster together in the red circles marked on *figure 10*. This shows that the spatial autocorrelation may be present here, and we would require a more formal test to diagnose the model.

Therefore, we conduct Moran's I test for residual spatial autocorrelation. From *table* 5, we obtain statistically significant values for Moran's I for the residuals of the three models, which indicates that the spatial autocorrelation is an issue in the three OLS models we constructed before. To solve this problem, we would conduct spatial autoregression model, which take the spatial autocorrelation into consideration.



(a)Chicago *(b) Los Angeles* *(c) Washington D.C.*

Figure 10. *Choropleth maps of the residuals of the three OLS models. The red circles shows the areas with a clustered patterning of residuals.*

Table 5. *Moran's I test for residual spatial autocorrelation*

(a) Chicago

Moran's I statistic standard deviated = 8.3406, **p-value < 2.2e-16**			
Sample estimates	Observed Moran's I	Expectation	Variance
	0.1608214189	-0.0061206353	0.0004006254

(b) Los Angeles

Moran I statistic standard deviate = 14.873, **p-value < 2.2e-16**			
Sample estimates	Observed Moran's I	Expectation	Variance
	0.2991553428	-0.0042227007	0.0004160967

(c) Washington D.C.

Moran I statistic standard deviate = 6.7505, **p-value = 1.473e-11**			
Sample estimates	Observed Moran's I	Expectation	Variance
	0.283467005	-0.015603764	0.001962773

4.4 The Lagrange Multiplier test

Between the two generally used spatial regression models: spatial error model and spatially lagged model, we conduct Lagrange Multiplier test of the OLS models we conducted before, to choose the more appropriate spatial regression model. For all of the three cities, the standard Lagrange Multiplier test for the error (LMerr) and the lagged (LMlag) models are below the 0.05 level (*table 6*), which means the robust version of Lagrange Multiplier test should be used to choose the more appropriate model. The results of Robust Lagrange Multiplier test for both error (RLMerr) and lagged (RLMlag) models are also significant for the three cities, however, RLMlag is orders of magnitude more significant than the other: for Chicago, <2.2e-16 compared to 0.003975; for Los Angeles, <2.2e-16 compared to 0.8635; for Washington D.C., 0.004004 compared to 0.01886 (*table 6*). In this situation, spatially lagged model is the more appropriate one to solve the spatial autocorrelation problem in the models of Airbnb penetration in the three cities.

Table 6. *Lagrange Multiplier test to choose the more appropriate spatial regression model*

(a) Chicago

LMerr = 62.182	p-value = 3.109e -15
LMlag = 135.86	p-value < 2.2e -16
RLMerr = 8.2953	p-value = 0.003975
RLMlag = 81.97	**p-value < 2.2e -16**

(b) Los Angeles

LMerr = 211.44	p-value < 2.2e -16
LMlag = 362.17	p-value < 2.2e -16
RLMerr = 0.029577	p-value = 0.8635
RLMlag = 150.76	**p-value < 2.2e -16**

(c) Washington D.C.

LMerr = 37.761	p-value = 7.998e-10
LMlag = 40.528	p-value = 1.938e-10
RLMerr = 5.5148	p-value = 0.01886
RLMlag = 8.2822	**p-value = 0.004004**

4.5 Spatial regression model: spatially lagged model (SLM)

Unlike OLS regression model, of which the coefficients for any of the explanatory variables (X) measure the absolute impact of these variables on the dependent variable (Y), the effect of an explanatory variables is the sum of two particular effects that X exerts on Y in spatially lagged model. One part of the effect is a direct, local effect of the covariate in that unit ($X_i \rightarrow Y_i$), and the other part of the effect is an indirect, spillover effect due to the spatial lag ($X_j \rightarrow Y_j \rightarrow Y_i$). The coefficients in table X only measure the direct, local effect of Xi on Yi, rather than the net effect, while the Rho parameter is the spatial lag, which measures the indirect effects (the effect of Airbnb penetration in surrounding tracts on Airbnb penetration in this tract).

Table X shows that all of the Rho parameters in the models of the three cities are significant, which adds further evidence that spatially lagged regression model (SLM) is a better model than the OLS regression model in studying Airbnb penetration. Besides, the positive coefficients of the Rho parameters in the three models indicates that when the Airbnb penetration in surrounding tracts increases, so does the Airbnb penetration in each tract.

***Chicago**. Table 7-(a)* shows that the significant influence factors of Aribnb penetration in Chicago is *Bus Stop*, *Bohemian Index*, *Talent Index*, *Proportion of Owner-occupied Residences*, *Median Household Value*, *Population*, and *Proportion of Young People*. Compared to the OLS model for Chicago, the significant variable *Hotel* in OLS model becomes insignificant in SLM, while the rest of variables are still significant. The coefficients of most of the variables decrease, and the coefficient of the spatial lag parameter Rho is 0.51817. According to the coefficients of the variables, the most influential factors of Airbnb penetration in Chicago is *Talent Index*, and the second most influential one is Median Household Value. It is worth mentioning that *Proportion of Owner-occupied Residences* is the only negative significant influence factor in the case of Chicago.

Los Angeles. *Table 7-(b)* shows that the significant influence factors of Airbnb penetration in Los Angeles is *Bus Stop*, Arts Attractions, *Bohemian Index*, *Talent Index*, *Median Household Value*, *Population*, and *Proportion of Young People*. Compared to the OLS model for Los Angeles, all of the factors keep being significant, and the coefficient of the spatial lag parameter Rho is 0.59349. *Population* and Median Household Value have strongest influence on the Airbnb penetration in Los Angeles according to the coefficients.

Washington D.C. *Table 7-(c)* shows that the significant influence factors of Airbnb penetration in Los Angeles is *Bus Stop*, Unemployment Rate, *Proportion of Owner-occupied Residences*, *Population*, and *Proportion of Young People*. Compared to the OLS model for Washington D.C., *Talent Index* becomes insignificant, and the coefficient of the spatial lag parameter Rho is 0.48089. *Proportion of Young People* is the strongest influence factor in the case of Washington, and Unemployment Rate is the only negative significant influence factor. In contrast to the situation in Chicago, the *Proportion of Owner-occupied Residences* is a positive factor here.

Table 7. *Spatially lagged analysis for Airbnb penetration in the three cities*

(a) Chicago

Variables	β	P-value
Bus Stop	0.055	0.014*
Hotel	0.034	0.103
Bohemian Index	0.050	0.021*
Talent Index	0.183	2.370e-07***
Owner Occupied Residences (%)	-0.101	1.475e-05***
Median Household Value	0.136	7.673e-05***
Population	0.088	1.018e-06***
Young people (%)	0.083	0.002**
Rho: 0.51817, LR test value: 132.26, p-value: < 2.22e-16		
Asymptotic standard error: 0.038177 z-value: 13.573, p-value: < 2.22e-16		
AIC: 1303.3, (AIC for lm: 1433.5)		

(b) Los Angeles

Variables	β	P-value
Bus Stop	0.047	0.003**
Arts attractions	0.089	1.714e-07***
Bohemian Index	0.089	8.046e-06***
Talent Index	0.154	8.514e-09***
Median Household Value	0.181	1.961e-11***
Population	0.249	4.194e-13***
Young people (%)	0.092	5.231e-06***
Rho: 0.59349, LR test value: 132.26, p-value: =2.7862e-10		
Asymptotic standard error: 0.066259 z-value: 7.2576, p-value: =3.9391e-13		
AIC: 217.24, (AIC for lm: 255.06)		

(c) Washington D.C.

Variables	β	P-value
Bus Stop	0.085	0.018*
Unemployment Rate	-0.109	0.039*
Talent Index	0.134	0.061
Owner Occupied Residences (%)	0.194	1.867e-05***
Population	0.129	8.473e-05***
Young people (%)	0.220	8.774e-07***
Rho: 0.48089, LR test value: 39.819, p-value: 2.22e-16		
Asymptotic standard error: 0.038177 z-value: 13.573, p-value: < 2.22e-16		
AIC: 1226.5, (AIC for lm: 1570.7)		

5. Discussion

5.1 Main findings

The study finds that there is strong spatial dependence in Airbnb penetration in all of the three cities by conducting Global Moran's I test and the Lagrange Multiplier test. Rather than a merely geographical clustering of the sources of the behaviour of interest, Airbnb actually interact with each other and might tend to cluster together because of the agglomeration effect that attracts more customers.

By conducting explanatory analysis using OLS regression and spatial regression, this study also finds that there are some common influencing factors of Airbnb penetration among the three different cities (Chicago, Los Angeles and Washington D.C.), including population, proportion of young people (under 34 years old), and number of bus stops (*table 8*). There are also some other factors such as Bohemian Index, Talent Index, Proportion of owner-occupied residences, and Median Household Value have significant influences on Airbnb penetration in two of the cities (*table 8*). These common factors indicate that Airbnb penetration is more likely to happen in the areas which are populous, vibrant, and with a higher proportion of young people and creative class. We also notice that proportion of owner-occupied residences has negative influences in Chicago but has positive influence in Washington D.C.. Quattrone et al. (2008) found that in London there is there is a statically significant negative relation between the proportion of owner-occupied residences and Airbnb penetration, because Airbnb hosts tend to rent rather than own the property. Our analysis only partially confirms this result, and the reasons of the positive significant influence in Washington D.C. would be explored further.

***Table 8**. Significant influencing factors of Airbnb penetration in the three cities*

Categories	Variables	β (Chicago)	β (Los Angeles)	β (Washington D.C.)
Geographic	Bus Stop	0.055	0.047	0.085
	Hotel	-	-	-
	Arts attraction	-	0.089	-
Socio-economic	Unemployment Rate	-	-	-0.109
	Bohemian Index	0.050	0.089	-
	Talent Index	0.183	0.154	-
	Owner Occupied Residences (%)	-0.101	-	0.194
	Median Household Value	0.136	0.181	-
Demographic	Population	0.088	0.249	0.129
	Young people (%)	0.083	0.092	0.220
Rho (spatial autoregressive parameter)		0.51817	0.59349	0.48089

5.2 Limitations

This study only studies the influence factors of Airbnb penetration in three cities, which is not comprehensive enough. Future work should replicate this study upon more cities, in order to generate a generalized prediction model of Airbnb penetration.

Besides, this study mainly focuses on global situation of Airbnb penetration in these three cities, not taking spatial heterogeneity into consideration. Future work can focus on local situation of Airbnb penetration in each of the city, using local regression model such as Geographic Weighted regression.

5.3 Practical Implications

One of the main findings of this study is that there are some common factors of Airbnb penetration in the three cities. By extending this study to more cities, we would obtain a generalized model of Airbnb penetration in cities of different nature. This suggests that, to a certain degree, the model could be applied to a city that has not been previously analyzed, to identify areas that tend to be have strong Airbnb penetration, and plan interventions to mitigate the negative impacts of Airbnb on local neighborhoods.

Reference

An Introduction to Spatial Econometrics in R .(n.d.). Retrieved April 23, 2016, from http://www.econ.uiuc.edu/~lab/workshop/Spatial_in_R.html#testing-for-spatial-autocorrelation

Anselin, Luc, and Anil K Bera. 1998. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." Statistics Textbooks and Monographs 155. MARCEL DEKKER AG: 237–90.

Anselin, Luc. 2003. "An Introduction to Spatial Regression Analysis in R." *Available at: Https://geodacenter.asu.edu/drupal_files/spdepintro.pdf*.

Gurran, N., Phibbs, P. (2017). When tourists move in: how should urban planners respond to airbnb? J Am Plan Assoc 83(1):80–92

Hughes, R., & MacKenzie, D. (2016). Transportation network company wait times in greater seattle, and relationship to socioeconomic indicators. Journal of Transport Geography, 56, 36-44. doi:10.1016/j.jtrangeo.2016.08.014

Kaplan, RA., Nadler, ML. (2015). Airbnb: a case study in occupancy regulation and taxation. U Chi L Rev Dialogue 82:103–115

Medina, J. and Solymosi, R. (2019, April 2). Crime Mapping in R from https://maczokni.github.io/crimemapping_textbook_bookdown/index.html#

Miller, S. R. (2016). First principles for regulating the sharing economy. Harvard Journal on Legislation, 53(1), 147.

Quattrone, G., Greatorex, A., Quercia, D., Capra, L., & Musolesi, M. (2018). Analyzing and predicting the spatial penetration of airbnb in U.S. cities. EPJ Data Science, 7(1), 1-24. doi:10.1140/epjds/s13688-018-0156-6

Sundararajan, A. (2014). What Airbnb gets about culture that uber doesn't. Harvard Business Review 11

Wachsmuth, D., & Weisler, A. (2018). Airbnb and the rent gap: Gentrification through the sharing economy. Environment and Planning. A, 50(6), 1147-1170. doi:10.1177/0308518x18778038

Zervas, G., Proserpio, D., & Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of airbnb on the hotel industry. Journal of Marketing Research, 54(5), 687-705. doi:10.1509/jmr.15.0204

Appendix:

Some important codes

#OLS regression analysis

```
> fit_1 <- lm(airbnb ~ retail+bus+hotel+art+BI+IDI+MHI+MHV+UR+PR+TP+PD+RDI+young+POOR+TI, data=ncovr_sf)

> summary(fit_1)

> vif(fit_1)
```

 # Moran's I test for OLS model

```
# coerce the sf object into a new sp object

> ncovr_sp <- as(ncovr_sf, "Spatial")

# create a list of neighbours using the Queen criteria

> w <- poly2nb(ncovr_sp, row.names=ncovr_sp$FIPSNO)

> summary(w)

> wm <- nb2mat(w, style='B')

> rwm <- mat2listw(wm)

# Moran's I test for OLS

> lm.morantest(fit_1, rwm, alternative="two.sided")
```

 # The Lagrange Multipliers test

```
> lm.LMtests(fit_2, rwm_n, test = c("LMerr","LMlag","RLMerr","RLMlag","SARMA"))
```

#Spatially Lagged Model (SLM)

```
> fit_2_lag <- lagsarlm(airbnb ~ bus+hotel+BI+MHV+TP+young+POOR+TI, data=ncovr_n_sf,rwm_n)

summary(fit_2_lag)
```