End-of-Degree Project in Economics

THE AIRBNB EFFECT ON THE RENTAL MARKET: THE CASE OF MADRID

Jorge Luis Casanova Ferrando

Supervisor:

Juan Carlos Salazar Elena

Academic course 2018/2019

May 2019

ABSTRACT

The debate over Airbnb is increasingly gaining attention both in academic and non-academic spheres. However, in specialized literature almost all analyses have ignored the spatial dependence behind it, that is, when landlord's decisions to raise or keep prices are related to each other. In the City of Madrid, non-spatial and spatial regressions were compared over individual rental prices. Results suggest that traditional models were biased and, once contemplating these effects, the impact of Airbnb is no longer significant. The causes could be the lower profitability, lack of legal guarantees and a strong competition of professional hosts. As a result, there are less incentives to displace dwellings from the rental market and push rental prices up.

Keywords: Airbnb, housing rents, endogeneity, spatial dependence.

Table of Contents

1. Introduction

Sharing or "peer-to-peer" companies frequently deal with fragmented buyers which are also highly differentiated one from another. However, all the companies share common aspects such as lower entry costs, short-term contracts, direct transactions and the use of digital media (Fradkin 2017).

When solving these problems, peer-to-peer companies have a trade-off between two objectives: received and use plenty of information efficiently and minimize transaction costs (Einav et al 2016).

Airbnb is a company that defines itself as collaborative[1]. It consists of an intermediation between those who offer a house or room for rent and those who seek it. The entire procedure is done in the website and with the guarantee of Airbnb. The company does not own these dwellings or organize the stays but receives a commission each time a rental is accepted.

In recent years, Airbnb has been a matter of debate for the effect it may have on other markets. However, the literature on the subject is focused on the impact over the hotel industry (Zervas et al. 2014) and over the housing markets (Barron et al. 2017; Horn and Merante 2017; Sheppard and Udell 2018; Segú 2018). Therefore, local governments have regulated this sector under little support in the literature or without enough information.

The objective of this work is to determine the impact that Airbnb has on the rental prices in the City of Madrid. In addition, these effects will be compared between groups of homes depending on the spatial structure of the data.

Those against Airbnb consider that the company has replace owners with short-term hosts. Consequently, housing supply could have been reduced and prices increased afterwards. Additionally, if tenants anticipate the possibility of their apartments being moved to the Airbnb market, they will be willing to accept higher prices and, as a result, push them up.

Other things to be considered against Airbnb are externalities such as noise or crime. These effects could not only decrease rental prices but also harm the coexistence in those neighborhoods.

Finally, the hotel industry has looked with bad eyes how Airbnb have easily acceded this market without minimum health or safety requirements. This problem implies not only a lower market share for the hotels but also the need of innovating to maintain their status.

Those in favor of Airbnb argue that through this company owners could earn an “extra money” while tourists can share a “local experience”. This would imply an increase in the tourism of the city and, therefore, greater economic activity (Kaplan and Nadler 2015). For that reason, in a certain way, collaborative economy platforms can be described as means to achieve efficiency in goods and services that are underutilized (Barron et al. 2017)

This research aims to make three contributions to the current literature. First, an empirical study of the Airbnb effect in Madrid, a city where rental prices have increased 9% annually since 2014[2]. Secondly, this work uses individual data which comes from web scraping techniques on Spanish rental websites and the Airbnb platform. Finally, the main contribution is to incorporate to the current literature the effects of spatial dependence and heterogeneity.

[1] *Airbnb citizen website.* Retrieved January 7,2019 https://www.airbnbcitizen.com/
[2] *Madrid Database 2019.* City Council of Madrid. Retrieved March 13, 2019 http://www-2.munimadrid.es/CSE6/jsps/menuBancoDatos.jsp

The city of Madrid is the capital of Spain and one of the most important tourist centers across Europe[3]. In 2017, Madrid received 9.9 million tourists which generated more than 20.9 million euros[4]. In addition, the increases in the number of tourists in Spain have coincided with a rise in the prices of homes. This have created a debate concerning a possible relation between both markets[5].

Estimating the impact of Airbnb has an important limitation: the lack of official data. In the city of Madrid, there is no rent reference index and the official data of the housing market is not updated (last estimates of the housing stock in Madrid are from 2011[2]).

In turn, companies such as Airbnb have few years in the Spanish market and the data update is not regular. In consequence, data scraping techniques were applied on two rentals websites in Madrid (Fotocasa and Pisos.com) and Airbnb (Inside Airbnb).

Once the information is filtered, this research involved three problems to be solved: how to define Airbnb density, how to avoid endogeneity on the regressions and how to correct for spatial dependence on data.

The database of this work contains geographic information such as latitude and longitude coordinates of each dwelling. This makes possible to create buffer zones around each rental house and find out how many Airbnbs are around it. Subsequently, the characteristics of this Airbnbs will be the most adequate to correct the endogeneity problem in the regression by using 2 Stage Least Squares Estimation (2SLS).

The problems of spatial autocorrelation and heterogeneity are usually common under models with geographic data. In this case, the location and similarity of data across regions are crucial in the rental market. To correct this issue, it will be necessary to include a spillover effect and the creation of clusters on the data.

The structure of this work will be the following. First, section II presents a review of the literature and a brief description of the current regulation of Airbnb in Madrid. Section III describes Airbnb data and the rental market, but also defines the concept of Airbnb density over data. Then, all the methodology is considered under spatial and non-spatial effects, accompanied by a brief review on the theoretical basis of the spatial model. Sections IV, V and VI have a discussion of the results and the conclusions. Finally, sections VII and VIII include the references and annexes of this work.

2. Background

2.1. Airbnb regulation in Madrid

Airbnb, founded in 2008, is a company that intermediates the accommodation through individuals or also called *short-term rental market*. The idea is to offer its platform so that suppliers and customers can value their services. They also guarantee the payment of the accommodation and share reviews and photos from the users.

There are three types of accommodation: entire properties, private rooms and shared rooms. Each of them is referenced by a price per night imposed by the landlords which are called hosts.

[3] The map of the city of Madrid with the 21 districts and the areas under study (Almond Districts) can be seen in Annex 1

[4] *Statistics from Madrid Tourism*. Retrieved March 13, 2019, from https://www.madrid-destino.com/turismo/estadisticas.

[5] Red2red Consultores (2017). Análisis del impacto de las viviendas de uso turístico en el distrito Centro.

Furthermore, each has the option of establish certain conditions such as deposits, additional fees or a minimum number of nights.

Regarding the regulation of this market, until now only the Decree 79/2014 has supervised the tourist apartments and accommodation (called VUT) in the Community of Madrid[6]. This level of regional government has attributed the management of business tourism and, therefore, everything related to it.

In Madrid, tourist apartments must have a first occupation license and deliver a responsible activity statement to the local community. In addition, the VUTs cannot be considered as habitual residences if they have at least three consecutive months active.

As a result, there is no consideration in the law for the so-called collaborative models of Airbnb, but, on the contrary, only those dwellings destined to carry out an economic activity are regulated. In many cases, this occurs with hosts of more than one announces or entirely dedicated to this business, also called professional hosts.[7]

2.2. Literature review

Airbnb has been criticized for its presence in the hotel and residential market. Numerous authors have discussed the so-called "Airbnb effect" on the real estate market and, in most of the debates, Airbnb does seem to affect the real estate prices (Barron et al. 2017; Horn and Merante 2017; Sheppard and Udell 2018; Segú 2018).

The first investigations were focused on hedonic price regressions with differences in differences approximations to assess the impact of Airbnb's appearance. Barron, Kung & Proserpio (2017) studied the impact across zip codes in the United States. The authors also used instrumental variables to control for endogeneity, including the number of searches of the word "Airbnb" on the Internet and the number of commercial premises in a certain region. They determined that a 1% increase in the number of airbnbs increases rents by 0.018% and sale prices by 0.026%.

Subsequent works focused on the analysis of housing units and not geographical areas. Horn and Merante (2017) studied the effect of Airbnb on rental prices in the City of Boston. The Airbnb variable was defined as the percentage of dwellings dedicated to Airbnb in a census tract. Through a panel data analysis, they determined that an increase in the standard deviation over Airbnb density increases the rental price by 0.4%.

Sheppard and Udell (2018) studied the impact of Airbnb's entry on the real estate market in New York City. The variable Airbnb was estimated as the number of publications around each dwelling at the time of its sale and, as for endogeneity, they assumed this cannot be relied upon to measure causal impacts. The conclusion was that by doubling the number of airbnbs 300 meters away, the sale price increases between 6 and 9%.

The most recent work and related to treat spatial effects is from Ayouba et al. (2019) where they estimated the Airbnb impact over rental rates in France. The dataset was defined by individual apartments and they evaluate the Airbnb density as the percentage of dwelling published on Airbnb in a certain geographical region. The results suggested that, when allowing for

[6] *Community of Madrid*. Retrieved March 17,2019 from http://www.madrid.org/wleg_pub/secure/normativas/contenidoNormativa.jsf?opcion=VerHtml&nmnorma=8631&cdestado=P#no-back-button

[7] The city council of Madrid has recently approved for 2019 the *"Plan Especial de regulación del uso de servicios terciarios en la clase de Hospedaje"* (PEH). The new regulation seeks to control the number of VUTS in the most centric Districts by requiring an independent access and applying for a license in the hole city of Madrid.

heteroscedasticity and spatial error autocorrelation of unknown forms, Airbnb puts upward pressure in some cities.

In Spain, the evidence of the impact of Airbnb on the real estate market is limited. This occurs due to an absence of official data and a scarce historical information.

Segú (2018) is one of the few empirical works on the impact of Airbnb in the city of Barcelona. Airbnb density was defined as the number of publications by neighborhoods under a required level of activity while endogeneity was solved by using instrumental variables. Among them, the distance to the beach and several demographic variables could altogether determine that Airbnb was responsible for a 4% in the increase of the rents.

In the City of Madrid, the only work that refers to the Airbnb effect comes from the Spanish Associations of Housing and Tourist Apartments (Fervitur 2018). They analyzed the impact of VUTs on the housing market using neighborhoods in Madrid during 2016. The results were that the effect on rents was null but, in the case of the Central District, the VUTs were responsible for a 1.19% increase over rates. The report also denies there was a reduction in the supply of housing but, instead, justifies the price increase as changes in demography and the economic recovery.[8]

Despite all the efforts to quantify the impact of Airbnb in the housing market, there is no work in Spain related to spatial econometric or any spatial technic applied to this matter. Nor the spatial dependence or spatial heterogeneity has been treated but only endogeneity as it is a root of bias. Therefore, this investigation seeks to control these problems and compare it with the usual estimation approach on the literature.

3. Methodology

3.1. Data

3.1.1. Airbnb

The Airbnb data comes from Inside Airbnb, a community of activists called Murray Cox. They are responsible for scraping all the information from Airbnb website and publish it every month for different cities around the world, including Madrid.

The website offers features of the hosts and the announcements they make on Airbnb. In the case of Madrid, a sample of 17,008 observations was obtained for the month of April in 2018, which includes entire dwellings, private and shared rooms. For every listing there are aspects such as room characteristics, host id, geographical coordinates and the date of first and last review. This allows me to track the level of activity.

The first assumption of this work is that only entire homes published on Airbnb could affect rental prices. On this, the previous literature maintains the same idea since, among other reasons, there is insufficient data to analyze and differentiate the impact of the rooms on Airbnb over the rental market.

Another of the first decisions when estimating the Airbnb effect is to distinguish those announces that are active and could generate an impact on the rental market from those who are not. On this aspect, there is literature from Zervas et al. (2014), Barron et al. (2017) and Segú. (2018). Under all cases, the level of activity of an advertisement is measured according to the date of the last review on the web and this could be defined by three options:

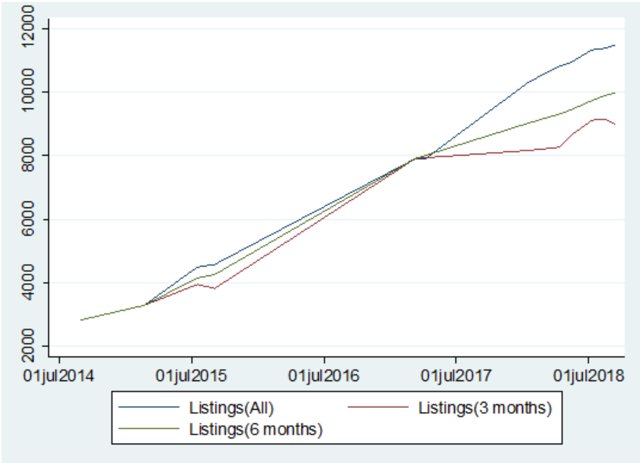
[8] Although this report gets some conclusions, it does not provide detailed information on the methodology used in the study.

The first one is to take the Airbnb first review as a reference and consider it active since then. This can be inconvenient because there are advertisements that once were active, but they stopped updating and could overestimate the Airbnb density.

The second option is to select listings on Airbnb that maintain a level of activity greater than three months since the last review. The third, and last, is to replace the three months by six to give them a greater margin on the level of update.

When observing the changes in the evolution of the announcements of entire homes (Figure 1), it is evident that to consider them all as actives would imply to overestimate the Airbnb supply. Specifically, only in the month of April the difference between those with respect to the minimum three months of activity (8,286) and six months (9,328) indicates a high bias on the data.

Figure 1: Number of listings in Madrid by level of activity



Data comes from InsideAirbnb.

In order to compare the effects by activity levels, the entire analysis will be based on the six months of activity. The goal is to avoid inactive announces and give some margin on errors in the database both on Inside Airbnb and on the Airbnb website.

3.1.2. Rental market

The data was obtained from a web scraping on two well-known rental websites in Spain: pisos.com and fotocasa.es. Each of them is an online real state database where users can post homes for sale or rent. These websites also include a description both physical and geographical, which could be extracted and then analyzed using statistical programs.

The sample obtained includes 5,442 properties published during the month of April in 2018 with the characteristics that appear in Annex 2.

In the first place, almost 60% of the sample is grouped mainly in the Central Districts of the City. These are, in turn, homes that concentrate the highest average rental price and price per square meter, but they are not the largest in terms of square meters or number of bedrooms.

One of the disadvantages of using prices from websites is that they could not be reflecting the final price of rentals agreements in the City[9]. However, since data comes from a web scraping, the price per square feet has been used as a proxy for the rental prices. This procedure has been

[9] In Madrid, the Instituto de la Vivienda de Madrid (IVIMA) currently regulates the obligation of landlords to get into deposits in order to guarantee and keep control over rents. These data exist but are not public by the open data policy of the regional government.

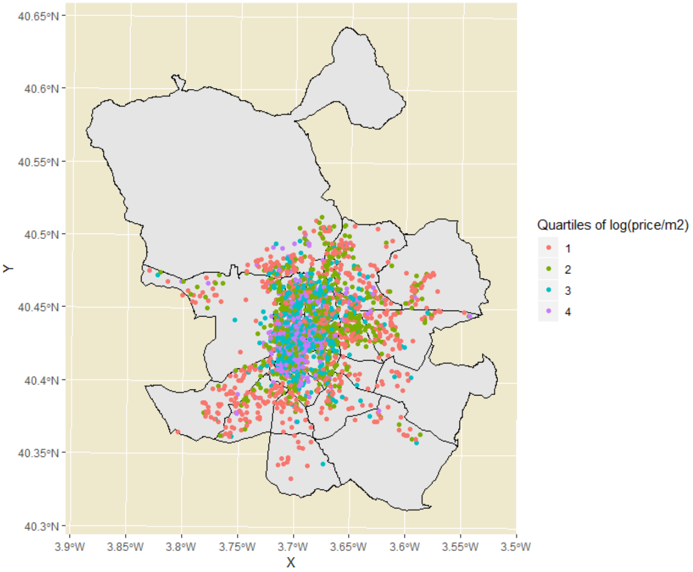
previously used in the literature and can be seen in Cheshire and Sheppard (1998), Orford (2000) and Chasco et al. (2018).

3.1.2.1. Exploratory Spatial Data Analysis (ESDA)

The ESDA consists on observing the tendencies and possible groupings of the data over certain geographical areas. Specifically, if a variable (such as the rental price) presents some of these problems (non-normal distribution, atypical points, spatial autocorrelation or spatial heterogeneity) it is reasonable to think that the residuals of any regression may be altering the results (Chasco 2008).

Figure 2 shows the distribution of the houses according to the price quartile to which they belong. Clearly, the Central Districts (also called Almond Districts) group the largest housing supply of the last quartile and could be a sign of spillover effects, that is, high prices which are surrounded by high prices and vice versa.

Figure 2: Quartiles of price per square meters in Madrid



Data comes from Fotocasa and Pisos.com.

For recognizing the presence of spatial autocorrelation, the Moran Test is defined as a coincidence of similar values in nearby locations (Anselin 2000). A positive Moran index indicates positive spatial autocorrelation (high prices-high prices or low prices-low prices) while a negative value would indicate a high-low relation on the variable to be studied. The effect of spatial autocorrelation can be contrasted through the Moran test (Cliff and Ord 1973,1981) which is defined as:

$$I = \frac{n}{S_o} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} (y_i - \bar{y}) (y_j - \bar{y})}{\sum_{j=1}^{n} (y_i - \bar{y})^2} \tag{4}$$

Where y_i is the variable to be studied (in this case the rental price per square meter), and $\bar{y}$ is the mean of the variable y, W is the spatial weights matrix and S_o is the sum of the elements of the matrix W.

The spatial weights matrix W expresses the neighborhood relations existing between the observations (Chasco 2008) and, in this case, it is a square matrix of order n = 5542 in which the elements w_{ij} of the matrix are the spatial weights:

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$$

This matrix reflects the “spatial influence” of unit j on unit i and impose a structure in terms of who are the neighbors for each location.

For this work, I first choose a W contiguity matrix where both rows and columns represent a dwelling in the geographical space. This matrix allows for expressing the relationship of each dwelling with those around it through a binary combination where $w_{ij} = 1$ states i and j are neighbors, and 0 otherwise. By convention, the self-neighbor relation is excluded, so the diagonal elements of W are 0.
W is usually row-standardized (W^S) in such a way that the elements of each row add 1, to facilitate the interpretation of the coefficients. In this case, by multiplying (yW^S) the result is a weighted average of the rental price per square meter of the neighbors of i.
The results show an average of 22 neighbors per dwelling and a complex framework in the Almond Districts, where a large part of the sample is also concentrated (see Annex 3)[10].

From the frequency histogram it can also be observed that the greatest number of connections between the houses are within 5 and 9 links. Therefore, a spatial weights matrix was created based on the idea of the nearest k-neighbor, such that:

$$w_{ij} = \begin{Bmatrix} 1 \text{ if } j \text{ is one of the k nearest neighbors to that of } i \\ 0 \text{ otherwise} \end{Bmatrix}$$

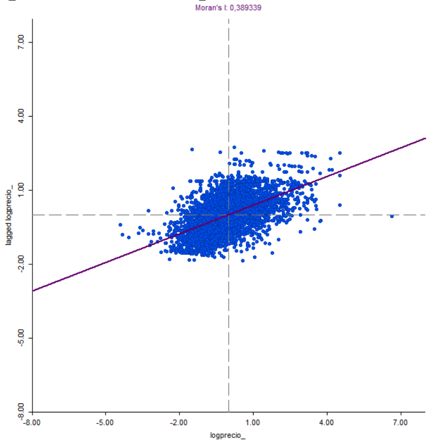
For this work, a matrix of 9-nearest neighbors will be used since it is the highest frequency of relations between dwellings and, in addition, it is consistent with the socioeconomic environment that exists in the City of Madrid.[11] In this way, each house is required to have at least 9 neighbors around him to define a suitable range of approach.

Once the spatial weights matrix is estimated, the absence of spatial autocorrelation can be inspected on the logarithm of the rental price per square feet $\log(price/m2)$.

The results reject the null hypothesis of no spatial autocorrelation (I = 0.38 and p-value = 0.000) and confirm that houses with high / low price levels are grouped together in the City of Madrid.

This test also makes inference about the normality of the distribution as the number of Moran Tests increases. How the information is the same, the statistical program is designed to obtain a permutation of values with a random distribution by assigning different values in each geographical location (Anselin 1995).

Figure 3: Global Spatial Autocorrelation



[10] The results are not consistent with the socioeconomic relationship in the City of Madrid. For instance, an evaluation of the rental market should not consider relations between the Moncloa and Carabanchel in the city of Madrid.
[11] The concept of neighborhood, under the results in the contiguity matrix, is not clear, but the interval defined between [5,9] seems the most appropriate. Therefore, *a priori* it could be thought that the rental market in Madrid operates on a reduced scale.

Even though there is some global space concentration of the rental prices, this phenomenon takes place in non-stationary spatial processes. This means spatial dependence changes with location and, sometimes, there could be small spatial clusters which takes a significant concentration or lack of high values. Consequently, there is global spatial autocorrelation in the variable, but each dwelling contributes differently to it.

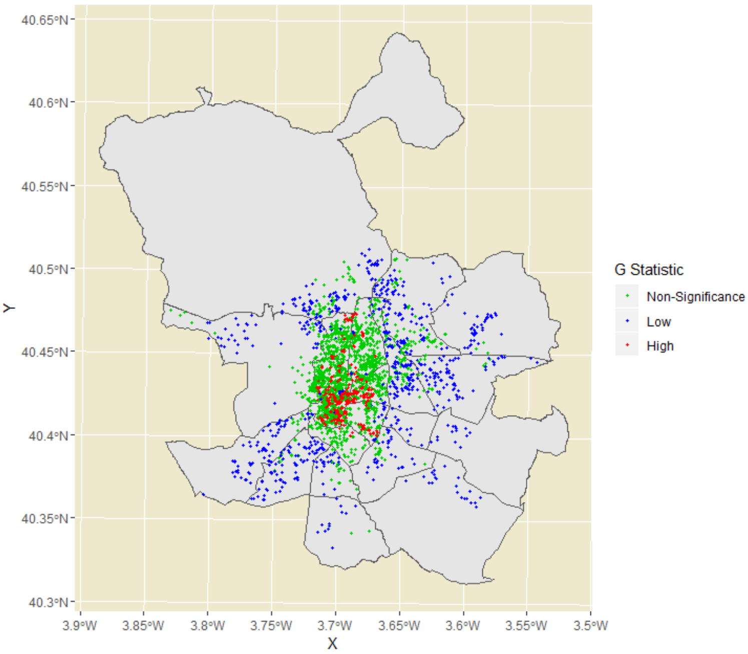
To asses this problem, another way to evaluate spatial autocorrelation is by using Local Spatial Autocorrelation Test (LISA) or the Getis and Ord's local statistics. The latter (Getis and Ord, 1992) measures the degree of association from the concentration of weighted points. As an inferential statistic, the null hypothesis states there is no spatial clustering and is given as:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} x_i x_j}{\sum_{j=1}^{n} x_i x_j} \tag{6}$$

Where x_i and x_j are the variable value for dwellings i and j, and $w_{i,j}$ is the spatial weight matrix between i and j. The result is a Z_G score which points clustering, in this case, of high prices if is more than zero ($Z_G > 0$) and clustering of low prices otherwise ($Z_G < 0$).

When considering the $\log(price/m2)$ variable and the 9- nearest neighbor spatial matrix, the results of the local G indicate that there are signs of clustering in the data (Figure 4). In fact, some houses in the Almond Districts show clusters of higher prices in contrast with the rest of the city. Therefore, this characteristic of neighborhoods in Madrid could explain the differences in rental prices by creating 2 spatial regimes (clusters of higher and lower prices) on the analysis.

Figure 4: Getis and Ord Local Statistic



3.1.3. Influence buffer analysis

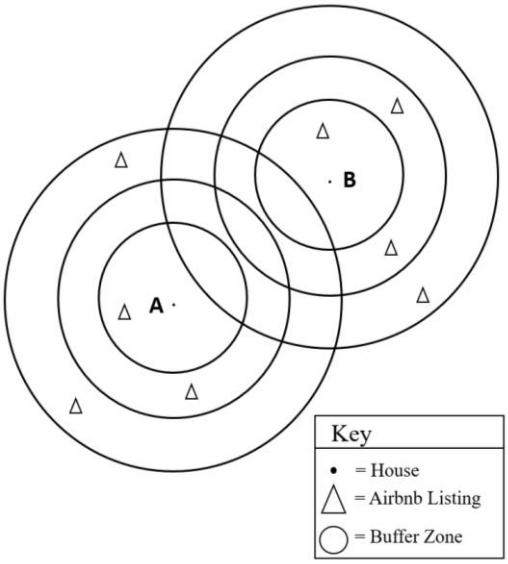
This research focuses on the impact of Airbnb around each house thanks to its geographical information[12]. If, instead, the study defines the Airbnb density as the total of announces on the entire neighborhood, we would be assuming that all Airbnbs affect dwellings by equal and this would lead to a bias in the estimation.

[12] Not all publications on Airbnb reflect an exact location but have a margin of error of up to 137 meters. For this analysis, 67.98% of the sample has an exact location.

The objective is to quantify the direct impact of those Airbnbs placed near each rental home. Therefore, buffers zones were created around each rental house, following Sheppard and Udell (2018) proposal.

From 200, 300 and 500 meters around each dwelling, the number and characteristics of the Airbnbs are estimated within that radius of distance. In addition, these buffer zones are defined according to the number of active Airbnbs (in the case of this analysis, with a minimum of 6 months of activity).

Figure 5 Airbnb Buffer Zones



Following Sheppard and Udell (2018) description

The results show, on the one hand, a change of scale in the variables as the radius of influence increases. On the other hand, the most relevant characteristics remain relatively constant at all levels (see Annex 4).

3.2. Spatial Framework

Spatial econometrics is based on dealing with two main spatial effects, namely, spatial dependence and spatial heterogeneity (Anselin 1998).

Spatial dependence occurs when a unit's outcome affects the choices, actions, or decisions of other units (Kirby and Ward, 1987; Ward and O'Loughlin, 2002; Plümper and Neumayer 2010).

Tobler's (1979) "first law of geography" says that "everything is related to everything else, but closer things more so". This implies that things that are closer will be affected by each other and, therefore, no considering can cause bias and inconsistency (Cliff and Ord 1981).

On the other hand, the spatial heterogeneity refers to a variation in relationships over space (Lesage 1999). This occurs when the mean or variance/covariance structure changes over a mapped process. One of the consequences is that the assumption of homoscedasticity cannot be accomplish and a full heteroskedastic error may be assumed. This means there are spatial regimes where different subsets of the data have different model coefficients (Anselin 1990).

Manski (1993) and Elhorst (1993) points out three different types of spatial interaction effects: i) **endogenous effects**: where the decision of a spatial unit to behave depends on the decision taken by other units. ii) **exogenous interaction effects**: where the decision of a spatial unit to behave depends on independent explanatory variables of the decision taken by other units. iii) **correlated effects**: where similar unobserved environmental characteristics result in similar behavior.

In the standard model, incorporating spatial dependence can be done in two ways: with an additional regressor in the form of a spatially lagged dependent variable (Wy), or in the error structure ($\lambda\varepsilon$) where $E(\varepsilon_i\varepsilon_j) \neq 0$.These can be expressed by using Anselin's notation (1988):

$$y = \rho W y + XB + \varepsilon \tag{7}$$

$$\varepsilon = \lambda\varepsilon + u \tag{8}$$

Where y is an $n \; x \; 1$ vector of observations, ρ is spatial autoregressive coefficient of the spatial lag term Wy, ε is a vector of error terms and X are the regressors.

Equation (7), also called spatial lag model or *SAR*, describes a spatial lag model. In this case, a spatial interaction is incorporated to avoid potential biasing influence. Spatial lag dependence in a regression model is like the inclusion of an autoregressive term for the dependent variable y_{t-1} in a time-series context (Anselin and Bera 1998).

Equation (8) is referred to a spatial error model, that is, a special case of a regression with a non-spherical error term (Anselin 1988). Under this scenario, the OLS stills remain unbiased but it is no longer efficient, and the standard errors will do be biased.

The presence of the spatial lag term *Wy* on the right side of Equation (7) will induce a nonzero correlation with the error term. Moreover, the spatial lag for a given observation i is not only correlated with the error term at i, but also with the error terms at all other locations. This will produce endogeneity and, therefore, 2SLS could be an appropriate estimator.[13]

The *SAR* model can be also express with a more general specification, by including additional endogenous variables

$$y = \rho W y + XB + \gamma Y + \varepsilon \tag{9}$$

where Y is an $n \; x \; 1$ matrix of observations on endogenous variables other than the spatially lagged dependent variable.

This model can be estimated through Spatial Two Stage Least Squares or S2SLS (Anselin 1980) that considers an endogenous element (Wy) and a list of instruments. These must have a high degree of correlation with the regressor and be asymptotically uncorrelated with the error term ε (Chasco 2008). The best instruments for the spatially endogenous term are, in fact, the own exogenous variables but spatially lagged (WX) (see Kelejian and Prucha 1998; Lee 2003).

Over the last years, several authors have generalized spatial models with additional endogenous predictors (Kelejian and Prucha 2004; Anselin and Lozano-Gracia 2008; Fingleton and Le Gallo 2008; Betz et al. 2019). The idea is to avoid bias on the estimators which could affect both non-spatial and spatial regressors[14].

So far, the literature has not included any autocorrelation effect on the estimations of the Airbnb effect. And at the same time, it has been proved that ignoring spatial interdependence follows asymptotically biased estimates (Betz et al. 2019).

[13] As a result of endogeneity, the equation can be readily as $= (1-\rho W)^{-1}XB + (1-\rho W)^{-1}u$. This violates the fundamental assumption of uncorrelated error terms $E[x_i u] = 0$

[14] Betz et al. (2019) demonstrated the bias on the OLS estimators can be seen as $\hat{\beta} - \beta = \left[\frac{cov(x,u)}{var(x)}\right] + \rho\left[\frac{cov(x,Wy)}{var(x)}\right]$ where the former refers to non-spatial endogeneity bias and the latter to spatial endogeneity bias

The idea under this investigation is to first study the existence of spatial autocorrelation over the rental market. Once determined, both spatial lagged and Airbnb endogeneity included in the regression must be corrected using instrumental variables.

3.3. Econometric Model

Related to the number of regressors, it is difficult to determine which and how many are the most relevant characteristics for a hedonic price function (Rosen 1974; Butler 1982).
Although there is no consensus, there are three categories on which the independent variables are usually classified: 1) the basic characteristics of the home, 2) the socioeconomic aspects that surround it and 3) the geographical environment.

As a benchmark model, it is first specifying a standard hedonic housing model with a set of explanatory variables: eight are attribute variables, two are geographical characteristics and three are related to the Airbnb buffer zone.

Var.	Description	Units	Source
log_price_m2	Rent price per square meter	log €	Fotocasa and Pisos.com
Structural Characteristics			
hab	Number of bedrooms	numb.	Fotocasa and Pisos.com
hab2[15]	Square of *hab*	numb.	
ban	Number of bathrooms	numb.	
ac	Air Conditioning	0-1	
clo	Closet	0-1	
p_b	Reinforced Door	0-1	
ter	Terrace	0-1	
tras	Storage room	0-1	
Geographical Characteristics			
almendra_c	Almond Districts	0-1	Self-elabor. ArcGIS
central	Central District	0-1	Self-elabor. ArcGIS
Airbnb Characteristics			
airbnb	*1*+Number of Airbnbs	log	Inside Airbnb
occupancy	Avg. occupancy rate in April	numb.	Inside Airbnb
reviews	Avg. number of reviews so far	numb.	Inside Airbnb

The location of an Airbnb, in many cases, is not random and determined by different factors such as amenities, proximity to points of tourist attraction or means of transport.

The exploratory analysis of the data (Section 3.1.2.1.) visually confirms that most of the Airbnb announces are located within the Almond Districts where, in addition, almost all the points of tourist attraction are concentrated (monuments, parks, commercial premises, etc.).[16] Furthermore, they also have a higher level of activity as they are mostly occupied among the year and, as a result, have more reviews compared with the announces on the periphery.

Instrumental variables for $Ln\ (Airbnb)_{if}$ will then be the average number of reviews from the Airbnbs and the average number of days occupied. The idea is intuitive, those listings on Airbnb

[15] A descriptive analysis reveals that the relationship between bedrooms and the price/m2 is not linear. This occurs due to the management of space through the dwelling and is according with the literature (Li & Brown, 1980)
[16] Only in the Central District there are 3,151 active restaurants and bars (15.93% of total in Madrid) according to the City Council of Madrid (2019)

that have more influence will have a higher level of activity and, therefore, more days occupied. The objective is to use the part of the variable Airbnb that allows explaining its behavior, but without being too much correlated with the price of rents.

Finally, regarding that the location of Airbnbs is mostly concentrated on the Central District, a binary variable is defined as 1 if a dwellings unit is in this District and 0 otherwise.

Once the type of relation and the number of variables is defined, the estimated models are the following:

OLS

$$Ln\ (Price/m2)_i = \beta_0 + \beta_1 Ln(Airbnb)_{if} + \beta_2 X_i + \varepsilon_i \tag{10}$$

2SLS

First stage

$$Ln\ (Airbnb)_{if} = \beta_0 + \beta_1 Reviews_{if} + \beta_2 Ocuppancy_{if} + \beta_3 Central_{if} + u_i \tag{11}$$

Second stage

$$Ln\ (Price/m2)_i = \beta_0 + \beta_1 Ln(\widehat{Airbnb})_{if} + \beta_2 X_i + \varepsilon_i \tag{12}$$

Where *i* refers to each property rented and *f* the radius of the buffer zone over which the impact is estimated.

The variable $Ln\ (Price/m2)_i$ is the logarithm of the price per square meter of each property; X_i refers to the set of control variables; $Reviews$ is the average number of reviews that Airbnbs around the property had so far, $Occupancy$ is the number of average days occupied by those Airbnbs around each dwelling in April and $Center$ is a binary variable that has a value of 1 if dwellings are in the Center District and 0 otherwise.

These models do not explicitly consider the spillover effects or any others spatial externalities that are important to explain the rental price. As a result, once the OLS and 2SLS models are estimated, these effects will be evaluated on the residuals to determine the need of including new variables.

4. Results

The results below are related to a 500m distance of each dwelling. This radio indicates a greater influence for Airbnb and, as expected, the biggest impact over the rental market.

In addition, the location of the Airbnbs is not exact and being strict about the radius of the buffer zone can lead to biases in the estimates. By giving a greater margin, estimations seem to be more consistent without misinterpretations. Nonetheless, the results with the rest of the buffer zones can be compared in the Annex 6 and Annex 7 of this work.

In the outputs of Table 1, estimates by both ordinary least squares (OLS) and 2 Stage Least Squares (2SLS) produce significant results in all variables[17]. The idea was to observe the changes before and after correcting endogeneity on the Airbnb variable. In this case, the signs remain constant in both regressions and are consistent with the review of the literature.

[17] The results for all the 2SLS show exogeneity on the instrumental variables and on the Airbnb variable. For more details, see Wooldridge (1995)

The results, moreover, reject the null hypothesis of no association between the Airbnb effect and the rental prices. The variable Airbnb (in this case, number of Airbnbs at 500m around each property) has a positive and significant effect on both OLS and 2SLS regressions. This implies that, while correcting the problem of endogeneity, Airbnb continued to explain the price levels in the City of Madrid.

As for the goodness of fit, the R^2 coefficient was 37.6%. That is to say, the basic characteristics of the property, the number of Airbnbs around and the geographical characteristics of belonging to a certain area explain 38% of the rental prices in Madrid. However, there still some doubts about what could justify the remaining 62% and is not included, such as spatial dependence.

Once the independent terms have been analyzed, each regression is accompanied by some contrasts to evaluate spatial issues (spatial autocorrelation or heteroscedasticity) and non-spatial ones (multicollinearity and normality on the error term).

The Jarque Bera test rejects in both models the null hypothesis of normality in the residuals. This would generate complications since non-normal error terms questions any estimate involving the maximum-likelihood method. Therefore, the contrasts of heteroscedasticity and the Lagrange multiplier test must be interpreted carefully (Anselin 1988).

The Breusch-Pagan test rejects, significantly, the null hypothesis of homoscedasticity in the residuals. Additionally, the results were compared with the White test (White 1980) which also supported the existence of heteroskedasticity and, in fact, are robust under a non-normality scenario.

Possible explanations for heteroskedasticity problems can be diverse. According to Anselin (1995), while the Breusch-Pagan test detects problems in independent terms, the White test can detect heteroskedasticity between groups of data. Therefore, this could be a sign of spatial heterogeneity in the models and, as a result, it would be recommended to include spatial regimes in the regression.

On the spatial dependence, the existence of autocorrelation on the residuals was contrasted with the spatial weights' matrix used in this investigation. As a result, the Moran Test and Anselin and Kelejian test were significant respectively, warming of a possible spillover effect on the regression and verifying what the ESDA demonstrated in section 3.1.2.1.

In the case of Lagrange multipliers (LMERR, R-LMERR, LMLAG, R-LMLAG), the hypothesis of including either spatial autocorrelation by omission of variables (ρW according with the Equation (7)) or spatial autocorrelation on the residuals (λ according with the Equation (8)) are both contrasted. In this case, the results obtained reject the null hypothesis and indicate a problem of spatial dependence both on the error term and the endogenous variable. However, once more these results must be interpreted with precaution as there is no normality on the residuals.[18]

In economic terms, the presence of spatial autocorrelation shows there is a dependence not only on the basic characteristics of a dwelling unit, but also on the rental price of the dwellings around it. This could be creating a spillover effect which must be consider in order to avoid bias and misspecification.

[18] Following the classical Anselin's strategy, the procedure will be to estimate from the specific to the general, that is, beginning with an OLS model and both *SAR* and *SEM* if there are signs of spatial autocorrelation.

Table 1 : Results of the regressions with 500m buffer zone around each dwelling

Variables	OLS	2SLS	SAR I	SAR II [b]
Airbnbs at 500m (6 months active)	0.017***	0.023***	0.012*	AA 0.009 BB 0.001
Bedrooms	-0.176***	-0.175***	-0.166***	
Bedrooms²	0.011***	0.011***	0.010***	
Bathrooms	0.044***	0.044***	0.039***	
Air Conditioning	0.056***	0.057***	0.056***	
Closet	-0.024***	-0.024***	-0.019*	
Reinforced Door	-0.041***	-0.041***	-.035***	
Garage	-0.0323***	-0.0308***	-0.030***	
Terrace	-0.051***	-0.050***	-0.042***	
Storage room	-0.031***	-0.031***	-0.029***	
Almond Districts	0.265***	0.249***	0.167***	
Constant	2.811***	2.80***	2.05***	
ρW	-	-	0.300***	0.336***
Observations	5542	5542	5542	5542
R^2	0.3761	0.3766 [a]	0.4300 [a]	0.4419 [a]
Test				
Jarque Bera	1428.863***	1909.3***	-	
Breusch-Pagan	327.278***	369.975***	-	
White	298.778***	158.446***	-	
I Moran	25.321***	24.818***	-	
LMERR	630.475***	-	-	
R-LMERR	64.821***	-	-	
LMLAG	590.226***	-	-	
R-LMLAG	24.572***	-	-	
Anselin-Kalejian		443.091***	13.007***	3.679
Chow Test for Spatial Regimes – Airbnb	4.334	36.071***	-	0.357
Chow Test for Spatial Regimes – Global	187.315***	216.635***	-	59.052***

* p<0.05, ** p<0.01, *** p<0.001

[a] Pseudo R^2. Under 2SLS and SAR models, this is estimated as the squared correlation between observed and predicted values of the dependent variable (Anselin and Le Gallo 2006; Wooldridge 2015).

[b] This model has been also run using larger spatial weight matrices. In the case of the AA group, the Airbnb effect continued to be not significant . As for the BB group, the coefficient was negative and significant in the *22-nearest neighbors* , warning of possible externalities of Airbnb over the rental market.

The new model, called the spatial autoregressive model or *SAR I* (see section 3.2), includes a spatial lag on the 2SLS to estimate if the spillover effect can absorb all the spatial dependence or if, on the contrary, it is necessary to include more complex specifications. The expression would then be the following:

$$Ln\ (Price/m2)_i = \beta_0 + \rho W(Price/m2)_i + \beta_1 Ln(\widehat{Airbnb})_{if} + \beta_2 X_i + \varepsilon_i \tag{13}$$

where $W(Price/m2)_i$ is the spatially lagged endogenous variable of matrix n x 1 and the rest of the variables that remain constant from the previous model (Equation (12)). This means that the price of a house depends on its own characteristics, the Airbnbs that surround it and the average price of its neighbors. In addition, to correct the endogeneity problems over $W(Price/m2)_i$ and $Ln(Airbnb)_{if}$ I use the same instruments of Equation (11) plus the spatially lagged independent terms X_i (see Section 3.3).

The objective is, on the one hand, to include the spatial effect on the rental prices. On the other hand, to avoid the endogeneity of the Airbnb variable which generates biases on the estimates (OLS model or Equation (10)).

The estimation of the *SAR I* model was performed by using the Spatial 2 Stages Least Square method (S2SLS). Although Anselin (1988) proposes the maximum likelihood method to estimate it, it also appeals to the S2SLS as the most appropriate when the hypothesis of normality in the error term is not supported (Chasco, 2008).

The results of the *SAR I* model, considering the k-nearest neighbor matrix, show that once the spatial lagged is incorporated, the Airbnb effect is lower over the rental prices. Therefore, it is reliable to think that under previous models this variable was picking up part of the spillover effect and not its real effect.[19]

The autoregressive spatial coefficient (ρ) has a high level of significance and a positive sign. The rest of the variables maintain the coefficients in some cases and in others are reduced. This implies that variables such as "Almond Districts" lose explanation since the spatially lagged rental price already describes the differences on data.

To study the spatial dependence on the residuals of the model, normally a S2SLS estimation uses the Anselin and Kelejian test (1997) in which the null hypothesis implies absence of spatial dependence. [20] In this case, the result of the *SAR* model rejects the null hypothesis and, therefore, assume there continues to be a problem of misspecification. One reason for this spatial dependence is precisely the spatial heterogeneity by the differences between clusters on rental prices. To solve this issue, two spatial regimes that were observed in the ESDA are then incorporated over the *SAR I* model (see Section 3.1.2.1).

In this way, the *SAR II* model will have the following structure:

$$Ln\ (Price/m2)_i = \beta_0 + {\beta_1}^{AA} Ln(\widehat{Airbnb})_{if} + {\beta_2}^{AA} X_i + {\beta_1}^{BB} Ln(\widehat{Airbnb})_{if} + {\beta_2}^{BB} X_i + \rho W(Price/m2) + \varepsilon_i \tag{14}$$

where AA and BB are the spatial regimes presented in Figure 4: AA would be dwellings with high prices surrounded by high ones and BB the rest of the sample.

[19] The results are consistent with the rest of the distances whereas Airbnb maintains a positive sign (see Annex).

[20] Anselin and Kelejian (1998) demonstrated that to obtain an appropriate result on the Moran test, the residues are required to be normally distributed. Also, as being a test based on OLS models it will tend to under-reject the null when a problem is present. They proposed an extended test to residuals called "Anselin and Kelejian test" for a two stage least squares (2SLS) regression estimation.

Once again, the model is estimated by using S2SLS where the endogenous variables are the spatially lagged price $W(Price/m2)$ and the number of Airbnbs $Ln(\widehat{Airbnb})_{if}$.

The *SAR II* model shows there are differences in how Airbnb affects rental prices in Madrid. First, those dwellings contained in a cluster of high prices surround it by high prices (AA) will be more affected by the number of Airbnb (0.009) than the rest (0.001), but under any scenario the impact is no significant.

Secondly, the Anselin and Kelejian test allows to assume there is no spatial autocorrelation on the residuals. This confirms the idea that the relation between heterogeneity and spatial dependence could be, sometimes, confusing.

The results of the Chow test indicate that, despite obtaining homogeneity on the Airbnb variable, at a global level it has not been possible to absorb all the effect of spatial heterogeneity. This demonstrates the need to incorporate more complex structures that allow explaining the rental princes in a more appropriate way.

5. Discussion

The houses located in the Almond Districts are not only characterized by high prices but also for having a greater number of Airbnbs around it. In addition, instrumental variables allow recognizing those homes surrounded by a higher level of activity in the Airbnb market.

The traditional models that were first analyzed indicated a biased effect on the dependent variable. What was happening is that the Airbnb variable was absorbing part of the spillover effect on the rental price. However, once the problem is solved the effect remains positive but without power of explanation and with a lower coefficient.

The non-traditional regressions, like the *SAR II* model, divided the sample into AA and BB houses. Those of the AA group are located within a high Airbnb influence. These are, in turn, houses with high prices in areas with a high level of economic and tourist activity. In contrast, BB group classify houses located in the rest of the central and the periphery of the City where the only feature in common could be in some cases the lower rental prices. Consequently, the Airbnb coefficients on the *SAR II* model have different values because they describe two types of clusters, but they are still relatively small.

As for the reasons of these results, one of the main argues against Airbnb is that landlords are being displace from the real estate market to the Airbnb market. This generates a shortage of housing supply and, as a result, push rental prices up. However, Madrid seems to be a city where the displacement does not occur or at least not in a significant way. The causes are several and could include the followings.

First, the profitability of Airbnb may be lower compared with the real state market. The income of an announce depends on how many days it has been rented and at which price. In the case of Airbnb, despite of having higher prices the number of days is lower than what is expected (Annex 5). In fact, during the month of April only 14 days where occupied on average, that is, almost 50% of the time. But it is also interesting to notice that for non-professional hosts (those with only one announce on Airbnb) the average income is very similar to what could be earned in the long-term market. What is different, and does not appear on the data, is when we distinguish between the gross and net income.

In the long-term market, part of the incomes for renting a house could be exempt of paying taxes in Spain[21]. This gives more stability when it comes to have a formal contract or receive rents every month. On the contrary, having an announce on Airbnb has not fiscal advantages and may imply bigger costs for changing of tenants every month. In addition, the regulation only tried to supervise the so-called professional hosts without allowing truly sharing economies on the short-term rental market.

Second, non-professional hosts must compete against professional ones (many times companies) by setting lower prices. This gives them more days occupied but also a lower income at the end of the month (Annex 5). In contrast, professional hosts gain more per announce because prices are usually higher and, in many cases, they offer a different type of lodgment or services. Furthermore, the minimum number required is different and gives them more flexibility on the market.

For these reasons, the incentives to displace dwellings to the Airbnb market might not be enough and even riskier in both legal and economic terms.

Finally, the Airbnb density is low and not uniform if data is entirely analyzed throughout the City of Madrid. On the one hand, Airbnb listings are clearly concentrated in the Central District, while the rest is mostly in part of the Almond Districts. Hence, the result of a higher impact on this region goes in line with what is expected.

[21] Art. 23.2 on the Personal Income Tax Law (IRPF). Retrieved April 22,2019. https://www.boe.es/eli/es/l/2006/11/28/35/con

6. Conclusion

The short-term rental market, as a part of a sharing economy, have functioned with a different dynamic from the usual market. Airbnb, one of the most important companies in the sector, has been object of numerous naggings. The main allege in cities like Madrid has been the impact on the real estate market. By displacing owners from the long-term to the short-term rental market, prices have increased without any concern from local governments.

Through an empirical study, it was determined that, after correcting the endogeneity and spatial dependence on traditional models, Airbnb does not explain the prices of rents in the City of Madrid.

In order to define the geographical space of study, it was first decided to divide the sample according to the concept of "Almond Districts". Later, based on the literature some buffers zones around each property were created to define the Airbnb density and observe their characteristics. From this way, the problem of proximity was corrected as best as possible, but always assuming the margin of error on the exact location of the Airbnbs.

Endogeneity was one of the most relevant problems and a reason for debate in previous studies. For this reason, the number of reviews and the average number of days occupied by the Airbnbs were found to be adequate instruments. When estimating the model, it was observed that under Ordinary Least Squares (OLS) the coefficient was biased.

Being a geographical issue and not considering the problems of dependence and spatial heterogeneity generates biases on the coefficients. Therefore, it was decided to include a spillover effect on prices and separate the sample into clusters (high and low prices) to evaluate if there were changes in the estimates. The result was that, under both clusters, the Airbnb effect was no longer significant, although those houses with high prices received a greater impact with respect to the rest.

Reasons to justify this lack of effect are the unfavorable regulation for Airbnb hosts in Madrid, the sparse average number of days occupied and the competitive relation between professional and non-professional hosts.

The results of the study, despite denying that Airbnb does affect the rental price, do not reflect other externalities that have been generated in Madrid. However, this study has political implications on how local government could properly justify future regulations. Finally, it also opens a new debate on which measures should be taken as this market have considerably increased in terms of professional hosts.

7. References

Airbnb (2018), 'About Us', Accessed via https://www.airbnb.com/about/about-us.

Anselin, L. (1980). Estimation methods for spatial autoregressive structures. Regional Science Dissertation and Monograph Series #8.

Anselin, L. (1988). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical analysis, 20(1), 1-17.

Anselin, L. (1990). Spatial dependence and spatial structural instability in applied regression analysis. Journal of Regional science, 30(2), 185-207.

Anselin, L. (1995). Local indicators of spatial association—LISA. Geographical analysis, 27(2), 93-115.

L. Anselin and H. H. Kelejian, Testing for spatial error autocorrelation in the presence of endogenous regressor, Int. Regional Sci. Rev. 20 (1997) 153-182

Anselin, L. (1998). Exploratory spatial data analysis in a geocomputational environment. Geocomputation, a primer. Wiley, New York, 77-94.

Anselin, L., and Bera, A. (1998). Spatial dependence in linear regression models with an application to spatial econometrics. Handbook of Applied Economics Statistics, Springer-Verlag, Berlin, 21, 74.

Anselin, L. (2000). Computing environments for spatial data analysis. Journal of Geographical Systems, 2(3), 201-220.

Anselin, L. and Le Gallo, J. (2006). Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects, Spatial Economic Analysis, 1:1, 31-52

Anselin, L., and Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. Empirical economics, 34(1), 5-34.

Anselin, L. (2014). Modern spatial econometrics in practice: A guide to GeoDa, GeoDaSpace and PySAL. GeoDa Press.

Ayouba, K., Breuillé, M.-L., Grivault, C., and Le Gallo, J. (2019). Does Airbnb Disrupt the Private Rental Market? An Empirical Analysis for French Cities. International Regional Science Review.

Betz, T., Cook, S. J., and Hollenbach, F. M. (2018). Spatial interdependence and instrumental variable models. Political Science Research and Methods, 1-16.

Barron, K., Kung, E., and Proserpio, D. (2018). The sharing economy and housing affordability: Evidence from Airbnb.

Chasco, C. (2008). Geografía y precio de la vivienda en los municipios urbanos de España. Revista de Economía de Castilla-La Mancha, 11, 243-272.

Chasco, C., Le Gallo, J., and López, F. A. (2018). A scan test for spatial groupwise heteroscedasticity in cross-sectional models with an application on houses prices in Madrid. Regional Science and Urban Economics, 68, 226-238.

Cliff, A. D., and Ord, J. K. (1981). Spatial processes: models and applications. Taylor and Francis.

Elhorst, J.P., 2010. Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis 5,9-28

Einav, L., Farronato, C., and Levin, J. (2016). Peer-to-peer markets. Annual Review of Economics, 8, 615-635.

Fevitur (2018). El impacto de las viviendas de uso turístico (VUTs) en el mercado de viviendas en alquiler de Madrid

Fradkin, A. (2017). Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb.

Fingleton, B., and Le Gallo, J. (2008). Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: finite sample properties. Papers in Regional Science, 87(3), 319-339.

Getis, A. Cliff, A.D. and Ord, J.K. (1973). Spatial autocorrelation. London: Pion. Progress in Human Geography, 19(2), 245–249.

Getis, A. and Ord, J.K. (1992) The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis, 24, 189-206.

Gutierrez, J., Garcia-Palomares, J. C., Romanillos, G., and Salas-Olmedo, M. H. (2016). Airbnb in tourist cities: comparing spatial patterns of hotels and peer-to-peer accommodation. arXiv preprint arXiv:1606.07138.

Horn, K., and Merante, M. (2017). Is home sharing driving up rents? Evidence from Airbnb in Boston. Journal of housing economics, 38, 14-24.

Kaplan, R. A., and Nadler, M. L. (2015). Airbnb: A case study in occupancy regulation and taxation. U. Chi. L. Rev. Dialogue, 82, 103.

Kelejian, H. and D. Robinson (1993), "A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a country expenditure model". Papers in Regional Science 72; pp. 297-312.

Kelejian, H. H., and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics, 17(1), 99-121.

Kelejian, H. H., and Prucha, I. R. (2004). Estimation of simultaneous systems of spatially interrelated cross-sectional equations. Journal of econometrics, 118(1-2), 27-50.

Kirby, A. M., and Ward, M. D. (1987). The spatial analysis of peace and war. Comparative Political Studies, 20(3), 293-313.

Lee, L. F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews, 22(4), 307-335.

LeSage, J. P. (1999). Spatial econometrics (p. 23). Morgantown, WV: Regional Research Institute, West Virginia University.

Li, M. M., and Brown, H. J. (1980). Micro-neighborhood externalities and hedonic housing prices. Land economics, 56(2).

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. The review of economic studies, 60(3), 531-542.

Plümper, T., and Neumayer, E. (2010). Model specification in the analysis of spatial dependence. European Journal of Political Research, 49(3), 418-442.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. Journal of political economy, 82(1), 34-55.

Sheppard, S., and Udell, A. (2018). Do Airbnb properties affect house prices. Williams College Department of Economics Working Papers, 3.

Segú, M. (2018). Do short-term rent platforms affect rents? Evidence from Airbnb in Barcelona.

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. Journal of the American Statistical Association, 74(367), 519-530.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. econometrica, 48(4), 817-838.

Ward, M. D., and O'Loughlin, J. (2002). Spatial processes and political methodology: Introduction to the special issue. Political Analysis, 10(3), 211-216.

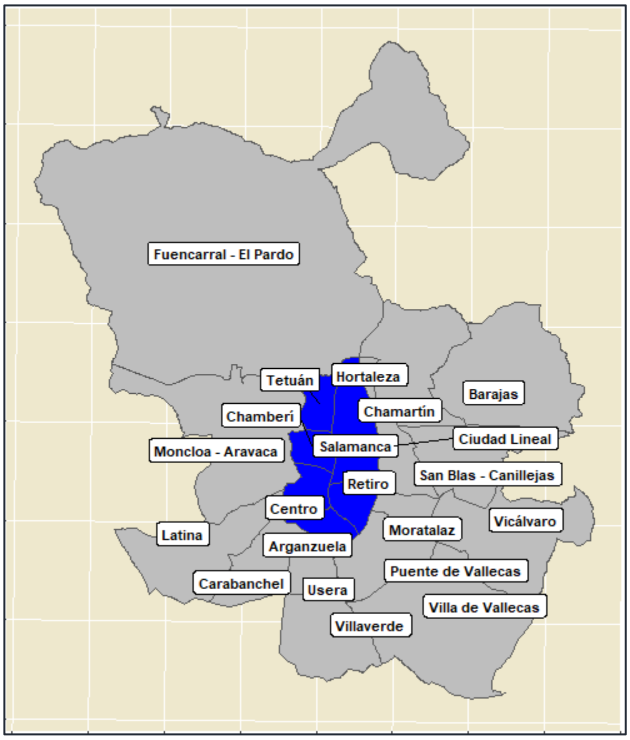
Wooldridge, J. M. 1995. Score diagnostics for linear models estimated by two stage least squares. In Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao, ed. G. S. Maddala, P. C. B. Phillips, and T. N. Srinivasan, 66–87. Oxford: Blackwell.

Wooldridge, J. M. (2015). Introductory econometrics: A modern approach. Nelson Education.

Zervas, G., Proserpio, D., and 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.

8. Annexes

Annex 1. Political map of Madrid and the Almond Districts

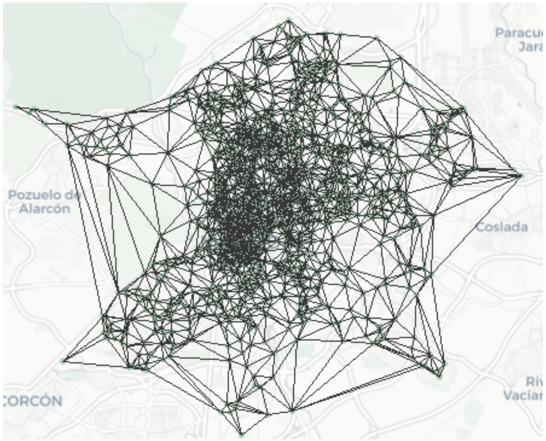


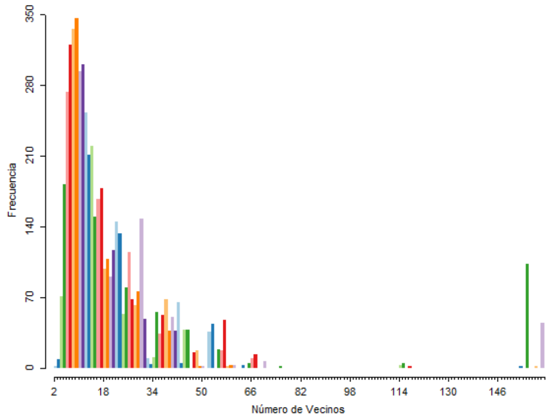
Blue regions belong to the Almond Districts and gray otherwise

Annex 2: Characteristics of rental dwelling by Districts in Madrid

District	Price per square meter (€/m2)	Price (€)	Bedrooms	Square meters	Observations
Centro	21	1695	2	93	16.9%
Arganzuela	16	1207	2	79	2.7%
Retiro	18	2006	3	122	5.3%
Salamanca	19	2320	2	130	12.5%
Chamartin	18	2446	3	144	13.9%
Tetuan	17	1528	2	94	6.5%
Chamberi	19	1777	2	106	9.8%
Fuencarral-el pardo	13	1586	3	135	3.3%
Moncloa-aravaca	16	2245	3	174	5.4%
Latina	11	842	2	77	1.9%
Carabanchel	12	869	2	77	2.4%
Usera	12	896	2	79	1.2%
Puente de vallecas	13	788	2	68	1.0%
Moratalaz	12	924	3	76	0.4%
Ciudad lineal	14	1438	3	112	6.0%
Hortaleza	13	2084	3	177	5.5%
Villaverde	12	921	2	79	0.5%
Villa de vallecas	12	833	2	76	0.6%
Vicalvaro	11	1060	3	112	0.3%
San blas-canillejas	12	1199	2	103	2.6%
Barajas	12	1637	3	157	1.3%

Annex 3. Contiguity Matrix from the rental dwellings database

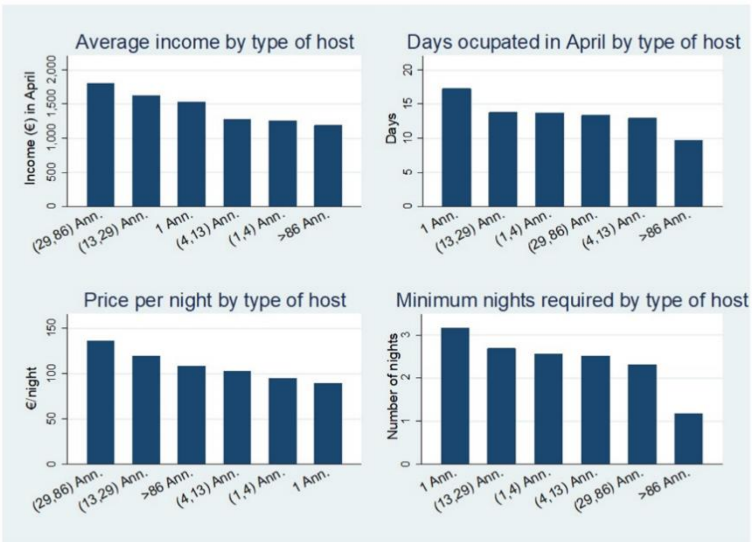




Annex 4. Airbnbs characteristcs by Districts of Madrid

District	Number of reviews			Number of days ocupated in April			Number of Listings		
	200m	300m	500m	200m	300m	500m	200m	300m	500m
Centro	50	53	51	15	15	15	159	342	880
Arganzuela	27	32	33	16	18	18	10	23	74
Retiro	28	27	27	15	16	16	12	27	77
Salamanca	17	18	18	16	15	15	19	42	110
Chamartin	14	16	17	13	14	14	5	10	29
Tetuan	19	19	19	16	16	16	9	20	50
Chamberi	23	25	25	15	15	16	15	34	97
Fuencarral-el pardo	6	10	12	8	12	16	0	1	3
Moncloa-aravaca	14	16	17	10	12	13	13	31	92
Latina	16	18	20	9	10	12	2	5	11
Carabanchel	14	16	18	10	11	14	2	4	9
Usera	24	21	23	16	17	19	2	4	11
Puente de vallecas	22	18	20	10	15	17	3	6	16
Moratalaz	11	18	24	6	4	11	1	1	2
Ciudad lineal	13	15	23	13	17	16	2	3	10
Hortaleza	9	12	18	8	14	14	1	2	3
Villaverde	3	7	8	6	15	16	1	1	1
Villa de vallecas	6	14	16	2	7	9	0	1	1
Vicalvaro	10	14	26	1	1	2	0	0	0
San blas-canillejas	19	21	23	6	9	12	1	1	4
Barajas	16	25	36	7	11	14	1	1	3

Annex 5. Differences between professional and non-professional hosts (nº announces)



Source: Inside Airbnb; Estimations can be biased due to some failures in the web scraping

Annex 6. Results of the regressions with 300m buffer zone around each dwelling

Variables		OLS	2SLS	SAR I	SAR II[b]
Airbnbs at 300m (6 months active)		0.002***	0.022***	0.010*	AA 0.013 BB -0.0006
Bedrooms		-0.175***	-0.174***	-0.165***	
Bedrooms²		0.010***	0.010***	0.010***	
Bathrooms		0.044***	0.044***	0.039***	
Air Conditioning		0.058***	0.058***	0.056***	
Closet		-0.023***	-0.024***	-0.019*	
Reinforced Door		-0.041***	-0.041***	-0.035***	
Garage		-0.031***	-0.0308***	-0.031***	
Terrace		-0.050***	-0.050***	-0.042***	
Storage room		-0.031***	-0.031***	-0.029***	
Almond Districts		0.261***	0.258***	0.170***	
Constant		2.815***	2.81***	2.03***	
ρW		-	-	0.311***	0.335***
Observations		5542	5542	5542	5542
R^2		0.3772	0.3784 [a]	0.4317 [a]	0.4418 [a]
Test					
Jarque Bera		1877.400***	-	-	
Breusch-Pagan		335.250***	358.039***	-	
White		298.792***	154.203***	-	
I Moran		25.283***	-	-	
LMERR		628.339***	-	-	
R-LMERR		69.902***	-	-	
LMLAG		578.154***	-	-	
R-LMLAG		19.717***	-	-	
Anselin-Kalejian		-	407.010***	12.015***	3.691
Chow Test for Spatial Regimes	Airbnb	6.024	13.752***	-	1.08
	Global	185.578***	192.971***	-	60.291***

* p<0.05, ** p<0.01, *** p<0.001

[a] Pseudo R^2. Under 2SLS and SAR models, this is estimated as the squared correlation between observed and predicted values of the dependent variable (Anselin and Le Gallo 2006; Wooldridge 2015).

[b] This model has been also run using larger spatial weight matrices. In the case of the AA group, the Airbnb effect continued to be not significant . As for the BB group, the coefficient was negative and significant in the 22-nearest neighbors, warning of possible externalities of Airbnb over the rental market.

Annex 7. Results of the regressions with 200m buffer zone around each dwelling

Variables	OLS	2SLS	SAR I	SAR II[b]
Airbnbs at 200m (6 months active)	0.022***	0.027***	0.012***	AA 0.01
Bedrooms	-0.174***	-0.173***	-0.164***	BB 0.003
Bedrooms^2	0.010***	0.010***	0.010***	
Bathrooms	0.045***	0.045***	0.039***	
Air Conditioning	0.059***	0.059***	0.056***	
Closet	-0.024***	-0.024***	-0.019***	
Reinforced Door	-0.042***	-0.042***	-0.035***	
Garage	-0.031***	-0.032***	-0.030***	
Terrace	-0.051***	-0.051***	-0.043***	
Storage room	-0.031***	-0.031***	-0.029***	
Almond Districts	0.266***	0.257***	0.162***	
Constant	2.820***	2.815***	1.98***	
ρW	-	-	0.331***	
Observations	5542	5542	5542	5542
R^2	0.3769	0.3778 [a]	0.4335 [a]	0.4442 [a]
Test				
Jarque Bera	1841.312***	1909.3***	-	
Breusch-Pagan	339.408***	353.045***	-	
White	295.702***	153.080***	-	
I Moran	25.630***	24.818***	-	
LMERR	645.614***	-	-	
R-LMERR	74.691***	-	-	
LMLAG	587.573***	-	-	
R-LMLAG	16.651***	-	-	
Anselin-Kalejian		542.562***	9.504***	1.926
Chow Test for Spatial Regimes – Airbnb	10.101*	26.333***	-	0.264
Chow Test for Spatial Regimes – Global	149.782***	204.118***	-	57.975***

* p<0.05, ** p<0.01, *** p<0.001

[a] Pseudo R^2. Under 2SLS and SAR models, this is estimated as the squared correlation between observed and predicted values of the dependent variable (Anselin and Le Gallo 2006; Wooldridge 2015).

[b] This model has been also run using larger spatial weight matrices. In the case of the AA group, the Airbnb effect continued to be not significant . As for the BB group, the coefficient was negative and significant in the 22-nearest neighbors, warning of possible externalities of Airbnb over the rental market.