Spatial Price Mimicking on Airbnb: Multi-host vs Single Host

David Boto-García
Department of Economics
University of Oviedo
botodavid@uniovi.es

Matías Mayor
Department of Applied Economics
University of Oviedo
mayorf@uniovi.es

Pablo De la Vega
Department of Economics
Complutense University of Madrid
pdelaveg@ucm.es

Abstract:

This paper studies the existence of two different supply operators in the peer-to-peer accommodation rental market for the city of Madrid. We specifically analyse spatial dependencies in price formation and whether the so-called *professional* hosts (i.e. those who have several Airbnb listings) set prices differently from single-property hosts. To this end, hedonic price models are estimated with and without spatial price dependence. Listings' structural characteristics and accessibility measures to transportation hubs and sightseeing spots are considered in the regressions. Our results provide clear evidence that price mimicking is higher among non-professional hosts whereas professional hosts set prices more independently.

Keywords: *Airbnb, spatial hedonic pricing, accessibility, spatial econometrics, professional hosts*

JEL codes: D4, C21, L83, R31, R52, Z3

Cite as:
Boto-García, D., Mayor, M. and DelaVega, P. (2021). Spatial Price mimicking on Airbnb: multi-host vs single-host. *Tourism Management*, 87, 104365. https://doi.org/10.1016/j.tourman.2021.104365

Acknowledgements:

We thank seminar participants at the Universities of Oviedo and Complutense of Madrid for valuable comments to earlier versions of this paper. David Boto-García acknowledges financial support from the Spanish Ministry of Education, Culture and Sport by means of a FPU grant (FPU 16/00031).

1. Introduction

Airbnb is nowadays the leading marketplace for peer-to-peer accommodations and is becoming increasingly demanded by travellers. This platform allows tourists to access to lodging services at generally lower prices than traditional accommodations. Unlike other sharing economy services, Airbnb hosts have the freedom to set their own prices. Assuming that Airbnb is a monopolistic competition market (Gunter et al., 2020), the price is a *markup* over costs, including opportunity ones. Listings are not homogeneous and differ depending on their attributes and location[1]. Given that, each listing would have a theoretical supply price. A common practise is to imitate the price setting of other listings with similar characteristics in the neighbourhood. Therefore, conditional on the hedonic attributes, there might be a spatial price formation process by which hosts mimic the prices of their neighbours. Indeed, Airbnb offers a pricing algorithm that suggests prices to hosts based on location and similarity (Hill, 2015).

Originally, Airbnb started as platform by which people rented spare bedrooms or properties to interested tourists for short periods of time. However, since one of the most argued reasons for hosts' motivation to rent is income (Karlsson and Dolnicar, 2016), nowadays a non-negligible share of hosts behaves close to business intermediaries (Kwok and Xie, 2019; Dogru et al., 2020). This professionalization of the host correlates with the number of listings on property, with those managing multiple properties (henceforth *professionals*) operating de facto as firms and gaining large revenues (Wegmann and Jiao, 2017; Gyódi, 2019). These multi-property hosts have been shown to fix prices more efficiently because they are better able to exploit economies of scale (Li and Srinivasan, 2019). Accordingly, there is a relevant distinction in behaviour between *true peer producers* and *professionals* in the Airbnb rental market (Einav et al., 2016).

There is a growing body of literature on Airbnb hedonic pricing that examines how listing-specific attributes, location and host characteristics contribute to price formation (Gibbs et al., 2018; Deboosere et al., 2019; Falk et al., 2019; Oskam et al., 2019)[2]. Some recent studies go beyond and explore spatial price autocorrelation using spatial econometric models (Önder et al., 2019; Eugenio-Martín et al., 2019; Tang et al., 2019; Lawani et al., 2019; López et al., 2019). All these empirical studies consider that there is a single spatial price dependence process (i.e., the strength of the price autocorrelation is the same for all the listings) and generally find a positive dependence. We, instead, aim to disentangle whether the spatial price dependence by which listings' prices depend on the prevailing prices in the neighbourhood differs depending on the type of host.

Therefore, the main objective of this article is to empirically analyse whether Airbnb spatial price dependence differs between professional and non-professional hosts. In the same spirit of

[1] Hereinafter, we refer to properties listed for rental on Airbnb as 'listings' and the owner of the listing as 'host'. Likewise, we use the term 'professional' for multi-property hosts and 'non-professional' for single-property hosts.
[2] The reader is referred to Sainaghi (2020) for a detailed review of academic research on peer-to-peer accommodations.

tax mimicking studies (Elhorst and Fréret, 2009; Delgado et al., 2015), we propose a two-regime Spatial Lag Model that distinguishes those who practise yardstick competition from those who do not in the Airbnb market. While non-professionals may imitate the prices charged by their neighbours (yardstick competition), professional hosts might fix prices in a more efficient and independent way based on marginal costs and maximising revenues (competitive monopoly equilibrium). This issue is relevant for policy makers, because it would imply that any price regulation in the form of taxes would not have neutral effects in the market.

A secondary goal is to study the role of some accessibility measures on Airbnb prices. There is wide evidence that tourists in cities like to lodge close to the main sightseeing attractions (Varma et al., 2016; Gunter and Önder, 2018; Benítez-Aurioles, 2018a). The housing economics literature has shown that closeness to cultural heritage (Lazrak et al., 2014; Franco and MacDonald, 2018) and transit facilities (Cordera et al., 2019; Yang et al., 2020a; 2020b) suppose important price premiums. Unlike previous studies on Airbnb hedonic pricing that mainly use distance to the city centre (Wang and Nicolau, 2017; Önder et al., 2019; Gibbs et al., 2018; Chica-Olmo et al., 2020), we consider accessibility to a set of monuments, museums and transportation hubs. In this way, we examine which price premiums are larger. Furthermore, the inclusion of accessibility indicators to different amenities in the analysis helps us to better identify the spatial dependence in price formation.

We use data from more than 4,000 Airbnb listings in the central area of Madrid. This city is an interesting case study for several reasons. First, it is the sixth European city in terms of Airbnb listings after Paris, London, Rome, Barcelona and Berlin (Adamiak, 2018). It has experienced a notable increase in its Airbnb supply during recent years, doubling the number of accommodates between 2016 and 2018. In this sense, the economic impact of Airbnb in Madrid is estimated to have been over 780 million euros in 2018 (Airbnb, 2019). Recently, Gil and Sequera (2020) show that Airbnb in Madrid is dominated by professional hosts specialised in the business of renting apartments. Second, Benítez-Aurioles (2018a) provide evidence that the distance elasticity of demand for Airbnb accommodations is three times larger in Madrid than in Barcelona. Therefore, the examination of the role of accessibility to tourism spots seems to be particularly important in this city. Additionally, the increase of rental prices together with the potential displacement of the local population has produced a public debate about whether it has been caused by Airbnb supply (Ardura-Urquiaga et al., 2020). By identifying the differences in price setting between professionals and non-professionals, our study provides policy makers with valuable insights for potential price regulations.

2. Data

The city of Madrid constitutes our case study. Specifically, our analysis considers only the six neighbourhoods in the central district because most tourist attractions in Madrid are located in its historic centre (Salas-Olmedo et al., 2018). Our database comes from *DataHippo*, a free-access platform that periodically scraps data from listings in several online accommodation platforms. Web-scraping techniques have become a common practise among scholars (Adamiak, 2018; Benítez-Aurioles, 2018b; López et al., 2019; Oskam et al., 2019). The data gathering procedure is as follows. First, all the listings in a particular area are scraped. Then, the searching algorithm periodically revises all the accommodations and replaces previous data with current information. When a new listing has been added to the platform or changed its price, the database is updated.

Our database has a cross-sectional structure. For each listing, the following information is provided: latitude and longitude coordinates, host ID, number of bedrooms, capacity, room type, total number of reviews, minimum number of nights to stay, and the price per night. To avoid pooling data from different periods, the sample is restricted to those listings that have been updated between April and June 2018. Therefore, our data is representative for 2018Q2. After removing listings with missing values, we have valid information for 4,308 Airbnb accommodations in the central district of Madrid.

Based on the room type, the accommodations can be classified into 3 groups: entire homes, private rooms and shared rooms. In the entire home case, the price reflects the cost of the full accommodation per night, no matter whether the guests equal the total capacity of not. Conversely, in the private and shared room cases, the price is the cost per person per night. Table 1 presents the notation and description of the listings' features along with summary statistics.

Variable	Description	Mean	SD	Min	Max
P	Accommodation price per night (in euros)	97.43	105.32	12	2,405
entire	The listing is an entire home	0.674	0.468	0	1
private	The listing is a private room	0.313	0.464	0	1
shared	The listing is a shared room	0.011	0.108	0	1
bedroom	Number of bedrooms in the accommodation	1.343	1.207	0	10
capacity	Maximum number of accommodates in the listing	3.38	2.100	1	16
min.LOS	Minimum length of the stay	2.652	2.371	1	30
reviews	Total number of reviews	26.598	44.068	0	349
d_justicia	The listing is in *Justicia* neighbourhood	0.153	0.360	0	1
d_embaj	The listing is in *Embajadores* neighbourhood	0.292	0.455	0	1
d_sol	The listing is in *Sol* neighbourhood	0.137	0.344	0	1
d_cortes	The listing is in *Cortés* neighbourhood	0.083	0.276	0	1
d_uni	The listing is in *Universidad* neighbourhood	0.175	0.380	0	1
d_palacio	The listing is in *Palacio* neighbourhood	0.156	0.363	0	1

Table 1.- Summary statistics (N=4,308)

The average price per person and night is 97 euros. Two-thirds (67%) of the listings are entire apartments, with an average capacity of 3.4 people. The minimum length of stay is, on average, 2.6 nights, whereas the mean number of reviews per listing is 26.5. *Embajadores* neighbourhood is the one that concentrates the largest share of the listings (29%).

Figure 1 shows the spatial location of the 4,308 listings within the six neighbourhoods that comprise the central district of Madrid. Black circles represent listings that belong to hosts that own more than 10 listings, whereas red triangles refer to the rest. The listings are densely distributed across the neighbourhoods. Indeed, there are 870 listings per square kilometre in our sample. Regarding spatial price heterogeneity, *Justicia* and *Cortes* are the most expensive neighbourhoods while *Embajadores* and *Sol* the least.

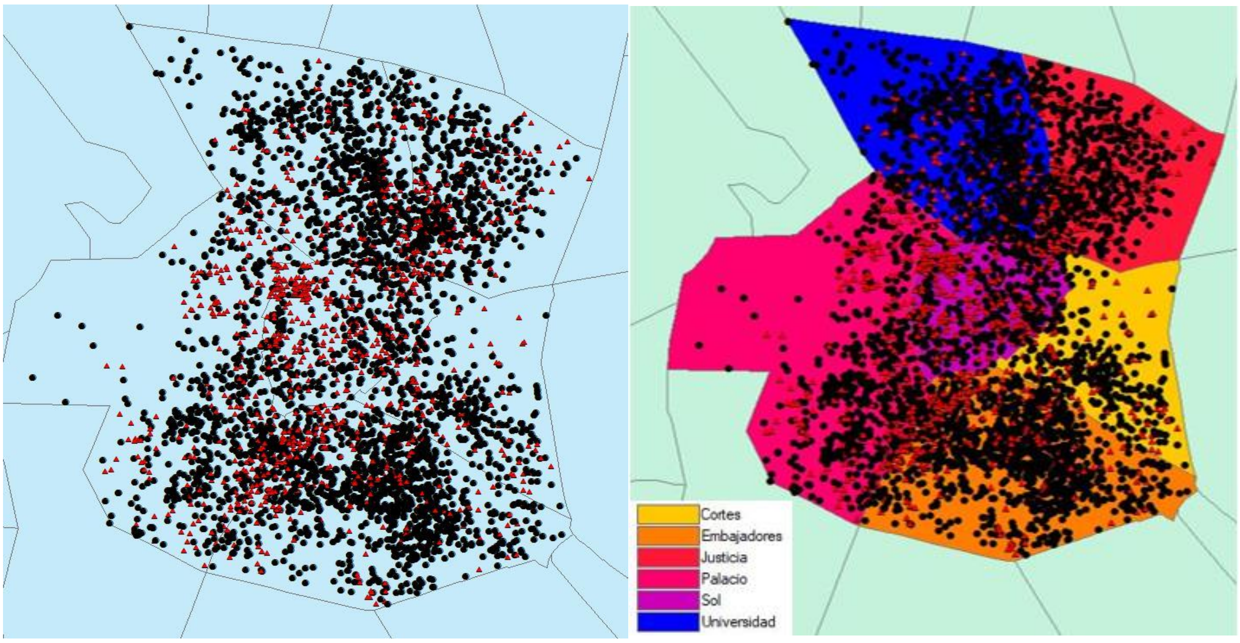

Figure 1.- Spatial distribution of the listings in the central district of Madrid

3. Empirical modelling

In this section we outline our empirical modelling. First, we derive a standard price hedonic model where prices depend on the accommodation attributes and its location. Second, we extend this framework to allow for spatial price dependencies.

3.1.A (standard) hedonic price model

According to the hedonic pricing method (Rosen, 1974), the price of an accommodation (in logs) can be expressed as a function of home-specific characteristics and its location. A micro-foundation for the importance of location characteristics on hedonic housing properties can be found in Li and Brown (1980). If we add an error term, the hedonic model becomes:

$$ln\,P = f\,(Listing\ attributes, Location) + u \qquad (1)$$

where $ln\,P$ is the natural logarithm of the price per night, $Listing\ attributes$ is an array of home-specific characteristics, $Location$ controls for where the listing is placed, and u is a random error term that reflects non-observable features.

Regarding home-specific attributes, the literature agrees that price is positively related with accommodation size (Benítez-Aurioles, 2018b; Wang and Nicolau, 2017). Accordingly, entire homes are expected to be more expensive. We define a dummy variable for whether the listing is an entire home *(entire*), with private/shared rooms acting as the reference category. Unfortunately, we lack data about the size of the listing. We proxy it by the maximum number of accommodates (*capacity*), which has been shown to be positively related with price (e.g. Chattopadhyay and Mitra, 2019). The number of bedrooms (*bedroom*) and the minimum length of the stay (*min.LOS*) are also included as other relevant attributes. Finally, several studies have acknowledged the relevance of the number of reviews (Wang and Nicolau, 2017; Gibbs et al., 2018; Benítez-Aurioles, 2018b; Oskam et al., 2019) for explaining listing prices, so this variable is also considered (*reviews*).

Concerning location, a large body of literature documents that house prices increase with accessibility to cultural heritage (Lazrak et al., 2014; Franco and MacDonald, 2018), transportation hubs (Cordera et al., 2019) and transit facilities (Yang et al., 2020a; 2020b). To explore their impact on Airbnb rental market prices, we define the following variables:

- Accessibility to sightseeing spots: from the large number of sightseeing spots in Madrid, we selected 19 places of interest (13 museums and 6 monuments, see Appendix A). For each listing, we compute the average distance (in kilometres) to each sightseeing spot (denoted as *dist.museums* and *dist.monum*, respectively). Accessibility is then calculated as the number of museums and monuments in the listings' vicinity (within a radius of 500 metres). These two accessibility measures are labelled *mus_500* and *mon_500*. We expect prices to be higher as the number of monuments and museums in the vicinity increases.
- Accessibility to *Atocha* train station: *Atocha* is the most important train station in Madrid. Potential guests might be willing to pay a price premium to be close to this transportation hub, but at the same time this can impose some nuisance in the form of unattractive landscape or noise (Yang et al., 2020b). To explore this, we first compute the distance between each listing and *Atocha* (*dist.atocha*). Subsequently, we define a dummy variable equal to 1 if *Atocha* is within a 500-metre radius (*Atocha.500*).
- Accessibility to the subway: given that Madrid is a subway-dependent city, listings that are close to a subway stop might exhibit higher prices due to the accessibility advantage it conveys. Out of the existing 740 entrances to subway stations, we selected the ones that are, on average, on a 1.25 kilometres radius from the listings (49 subway entrances). We then compute the distance (in kilometres) from each listing to each subway entrance. Next, we define two alternative measures for accessibility: 1) the distance to the closest subway entrance (*min.dist.sub*), and 2) the number of subway station entrances within a radius of 750 metres (*num.sub.less750*).

Descriptive statistics of the above defined accessibility indicators are presented in Table 2.

Variable	Description	Mean	SD	Min	Max
dist.mus	Average distance to museums	1.784	0.285	1.181	2.808
dist.mon	Average distance to monuments	3.063	0.290	2.544	3.977
dist.Atocha	Distance to Atocha station	1.361	0.512	0.010	3.051
mus.500	Number of museums within 500 metres	0.491	0.984	0	7
mon.500	Number of monuments within 500 metres	0.605	0.793	0	3
Atocha.500	Indicator variable for whether Atocha station in within a 500-metre radius.	0.048	0.215	0	1
min.dist.sub	Minimum distance to the nearest subway stop	0.199	0.118	0.001	1.033
num.sub.less750	Number of subway stops within 750 metres	17.980	9.627	0	37

Table 2.- Summary statistics of the accessibility indicators

Listings owned by hosts with several listings are expected to exhibit a different price setting. The literature has shown that multi-property hosts are more economically-focused when hosting (Farmaki and Kaniadakis, 2020; Gil and Sequera, 2020). To account for this, we first define a count variable for the number of listings per host in the city centre (*num.list.centre*). There are 2,636 distinct hosts in our sample, who own 1.69 properties on average. About 80% only own one listing whereas there is a host with 170 accommodations. Figure 2A presents a scatterplot of the price per listing in euros against the number of listings owned by its host. Figure 2B shows a similar scatterplot but of the mean price set by each host (average of the prices of their properties) against the number of listings.

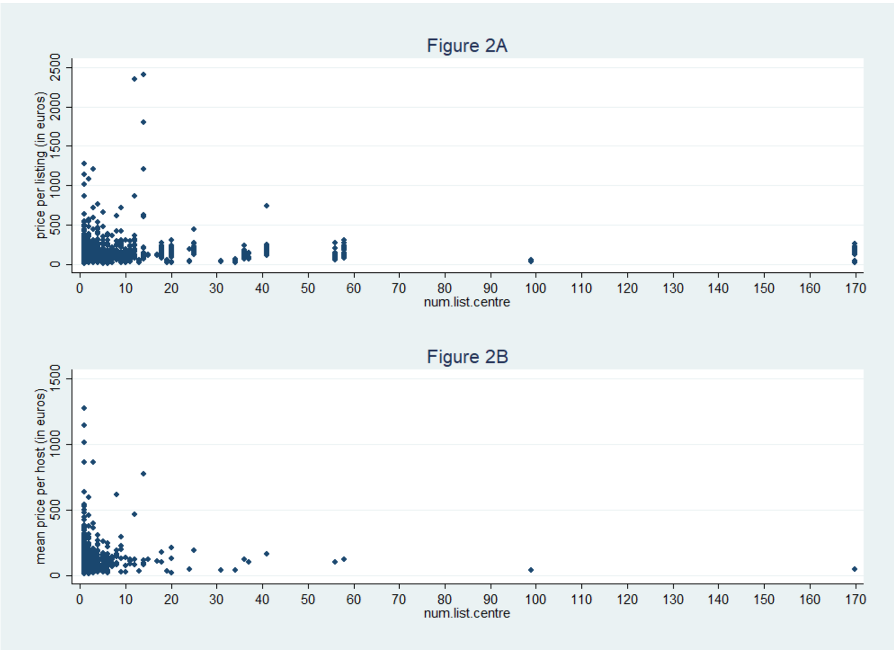

Figure 2.- Scatterplot of prices per listing and mean price per host against *num.list.centre*

According to Figure 2, it seems that there is a negative relationship between prices and the number of listings on property. This would mean that multi-host owners charge lower prices than single-property hosts. To check this, we expand equation (2) by including *num.list.centre* in the model. Nonetheless, the differences in price setting might become relevant up to a certain threshold. Therefore, we add to the model a dummy variable for whether the host has 10 listings or more (*mhost*).

The definition of the threshold point for considering a host to be a professional is subject to debate. Gibbs et al. (2018), Falk et al. (2019) and Gunter et al. (2020) consider a host as multi-host if she owns two or more listings. Conversely, in Ayouba et al. (2019) multi-hosts are those with several listings or who rent a property for more than 120 days per year. We choose 10 as the cut-off point following Deboosere et al. (2019), who report a differential effect in hosts' behaviour from ten properties onwards. Nonetheless, the robustness of this threshold point is examined later.

Therefore, the empirical model becomes:

$$ln\, P_i = \alpha + \beta_1\, entire_i + \beta_2\, bedrooms_i + \beta_3\, capacity_i + \beta_4\, min.LOS_i + \beta_5\, reviews_i + \gamma_1\, mus.500_i + \gamma_2\, mon.500_i + \gamma_3\, Atocha.500_i + \gamma_4\, num.sub.less750_i + \delta_1\, num.list.centre_i + \delta_2\, mhost_i + u_i \tag{2}$$

where subscript i stands for the listing and u_i is a random error term normally distributed with zero mean and constant variance.

3.2. A price hedonic model with spatial dependence

As it happens in the hotel industry, Airbnb listing prices might also depend on the prices of the other listings in the vicinity due to competitive rivalry. Since accommodation supply is more concentrated in cities than in other destinations (Eugenio-Martín et al., 2019), dependencies in price formation are especially likely to hold in Madrid. Moran I statistic under different spatial weigh matrixes and semi-variograms using different distance widths and spatial lags indicate there is indeed spatial price autocorrelation in Airbnb listings (available upon request).

There are several explanations for this spatial dependence. Airbnb market can be understood as a chain-linked competitive monopoly (Chamberlin, 1933), where each firm provides a different product based on its characteristics and price differentiation takes place (Eugenio-Martín et al., 2019; Gunter et al., 2020). When searching for an accommodation, potential guests can filter by the attributes they are looking for and the geographic area they want to stay. Hence, Airbnb listings compete for demand catching with other listings in the neighbourhood with similar characteristics. Furthermore, hosts have the possibility to ask the platform for a price suggestion based on the prevailing prices in the same neighbourhood. Price setting thus appears to be the result of a spatial reaction function (e.g. Padovano and Petrarca, 2014) that incorporates the decisions made by other agents in the market given characteristics and subject to random noise as follows:

$$ln\, P_i = f\,(ln\, P_j, Listing_i, Location_i) + u_i \tag{3}$$

where P_j denotes the price levels of other listings in the neighbourhood for $j \neq i$.

Empirically, assuming the spatial dependence in price arises due to price mimicking, we propose a Spatial Lag Model (SAR) as follows:

$$ln\,P_i = \alpha + \rho \sum_{j=1}^{J} w_{ij} \ln P_j + \beta_1\, entire_i + \beta_2\, bedrooms_i + \beta_3\, capacity_i\, + \beta_4\, min.LOS_i \\ + \beta_5\, reviews_i + \gamma_1\, mus.500_i + \gamma_2\, mon.500_i\, + \gamma_3\, Atocha.500_i \\ + \gamma_4\, num.sub.less750_i + \delta_1\, num.list.centre_i + \delta_2\, mhost_i + u_i \tag{4}$$

with $W = \sum_{j=1}^{N} w_{ij}$; $w_{ii} = 0$; $w_{ij} = w_{ji}\ \forall i,j$ for $i,j = 1,\dots,N$

where ρ is the spatial autocorrelation coefficient that measures the strength of the spatial dependence, W is a *NxN* spatial weight matrix in which w_{ij} denote the (i,j)th element following LeSage and Pace (2009), and the rest of variables are the same as introduced before. Due to the correlation of the spatial lag term $WlnP$ and u, OLS estimates are biased and inconsistent. Therefore, the model in (4) is estimated by Maximum Likelihood (ML).

In its reduced form, the SAR model conceptualizes prices as a function of the characteristics of the listing itself and the prices of the neighbouring properties subject to a distance decay operator (Anselin and Lozano-García, 2008). This is similar to the models that explain strategic interactions in fiscal policy (e.g. Delgado et al., 2015). Remarkably, contrary to other modelling alternatives, the SAR specification considers *global spillovers* in the sense that a shock in prices in listing *j* would impact the price of listing *i* even though they are not direct neighbours.

The spatial weight matrix *W* reflects the neighbouring relationship in the data. Different alternatives are used in empirical applications: contiguity-based, K-nearest neighbours and inverse-related distance-based, with and without a cut-off point. The latter has been the most widely used in related applications. We consider a listing *j* to be a neighbour of listing *i* if it is within a given radius. The following distance thresholds are examined: 100, 150, 200, 250, 300, 350, 400, 450 and 500 metres. Among them, the 350-metre threshold is the one that provides the best fit based on the log likelihood and information criteria (see Appendix B). Accordingly, we take 350 as the distance threshold that defines the neighbouring relationship. Formally, the elements of the weight matrix (w_{ij}) are defined as follows:

$$w_{ij} = \begin{cases} 1\ if\ d_{ij}\ \leq 350\ m \\ 0\ if\ d_{ij} > 350\ m \end{cases} \tag{5}$$

The weight matrix is row standardized. The average number of links is 446 so that each listing is connected, on average, with 10% of the sample. Although quite sparse, there is some evidence that parameter estimates of a SAR are biased when using denser weighting matrices (Smith, 2009).

3.3 A Spatial Price Hedonic model with two regimes

As introduced before, hosts' price mimicking might be heterogeneous. Single-property hosts that rent spare space might have limited information about equilibrium prices. They might be *rationally ignorant* so that they take the prevailing prices in the neighbourhood as a benchmark. Contrariwise, hosts with several listings (*professionals*) might be more commercially oriented and adjust prices to ensure demand. They have more experience and devote more effort to the optimal price setting. Since they might have lower costs due to economies of scale (Li and Srinivasan, 2019), their marginal costs might decrease with the number of listings and they might be able to set lower prices. Under this reasoning, single hosts adopt the so-called yardstick competition (Shleifer, 1985) and imitate the prices set by neighbours, whereas multi-host operate more independently with lower price dependence.

To explore this, we propose a SAR model with two regimes following Rietveld and Wintershoven (1998), Allers and Elhorst (2005), Elhorst and Fréret (2009), Delgado and Mayor (2011) and Delgado et al. (2015). These spatial regimes are identified with a dummy variable *mhost* that takes value one for professionals (>10 listings) and zero for non-professionals (<10 listings). The model considers two different intercepts ($\alpha_{1|mhost=0}$ and $\alpha_{2|mhost=1}$) and two different spatial autocorrelation coefficients ($\rho_{1|mhost=0}$ and $\rho_{2|mhost=1}$) for each type of host as follows:

$$
\begin{aligned}
ln\ P_i = {} & \rho_{1|mhost=0}\,(1-\delta_i)\sum_{j=1}^{J} w_{ij}\, ln\ P_j + \rho_{2|mhost=1}\,\delta_i \sum_{j=1}^{J} w_{ij}\, ln\ P_j + \alpha_{1|mhost=0} \\
& + \alpha_{2|mhost=1} + \rho\, W\, ln\ P_j + \beta_1\, entire_i + \beta_2\, bedrooms_i + \beta_3\, capacity_i \\
& + \beta_4\, min.LOS_i + \beta_5\, reviews_i + \gamma_1\, mus.500_i + \gamma_2\, mon.500_i \\
& + \gamma_3\, Atocha.500_i + \gamma_4\, num.sub.less750_i + \delta_1\, num.list.centre_i \\
& + \delta_2\, mhost_i + u_i
\end{aligned} \tag{6}
$$

where δ_i is a binary variable that takes value 1 when $mhost = 1$ and 0 otherwise (Allers and Elhorst, 2005). Note that $(1 - \delta_i)\sum_{j=1}^{J} w_{ij}\, ln\ P_j$ and $\delta_i \sum_{j=1}^{J} w_{ij}\, ln\ P_j$ denote the different spatial effects of the neighbouring prices depending on the multi-property regime. Since $mhost$ is used to define the two regimes, it is omitted from the specification. Equation in (7) is estimated by Maximum Likelihood using adapted MATLAB code from Elhorst and Fréret (2009).

4. Estimation results

4.1. Standard hedonic price modelling

We first estimate the standard hedonic price model in (2) by Ordinary Least Squares (OLS). Because some hosts have more than one listing, we cluster standard errors at the host level to control for potential cross-sectional dependence following Xie and Mao (2017) and Kwok and

Xie (2019). By doing so, the model estimates are robust to any common shared unobserved factor at the host level.

Recent studies by Wang and Nicolau (2017), Falk et al. (2019) and Moreno-Izquierdo et al. (2020) have acknowledged the importance of examining the effect of the hedonic attributes on the mean price but also on the conditional distribution. Therefore, we also run quantile regression for the 10th, 25th, 50th, 75th and 90th percentiles. Again, standard errors are clustered at the host level following Parente and Santos-Silva (2016).

Entire apartments vs shared/private rooms can be understood as separate products within the same market (Sainaghi et al., 2021). Even though level differences are controlled for through the dummy *entire*, it could be the case that the listing and location characteristics have a different effect on prices depending on the accommodation type. To allow for heterogeneous effects, we initially interacted all the variables in (2) with *entire*, both for the linear and the quantile regressions (see Appendix C Table A3). Since only three of the eleven interactions are significant (*capacity x entire*, *min.LOS x entire* and *mhost x entire*), we opt for a parsimonious specification with only these three interaction terms. Table 3 presents the estimation results for the OLS and quantile hedonic regressions. As shown by Parente-Santos Silva tests, clustering is relevant here to avoid biased standard errors.

Starting with the listing attributes, there is a significant price premium for full apartments (*entire*). This matches earlier results obtained by Wang and Nicolau (2017), Ert and Fleischer (2019), Deboosere et al. (2019) and Moreno-Izquierdo et al. (2020). Specifically, full apartments are on average 107% ((exp (0.728)-1)*100) more expensive[3]. Similarly, the higher the number of bedrooms and the maximum capacity of the listing, the higher the price. This is consistent with Chen and Xie (2017), Gibbs et al. (2018), Lawani et al. (2019) and Chattopadhyay and Mitra (2019). Worthy of note, the positive contribution of capacity to price mainly holds for entire apartments according to the significance of the interaction term. Interestingly, a higher minimum number of nights required is associated with a lower price on average (-2.6%), but positively impact prices for entire apartments (+3.6%). Concerning the number of reviews, this variable is negatively related to prices (-0.2% per review), in line with Benítez-Aurioles (2018b), Gibbs et al. (2018), Lawani et al. (2019) and Moreno-Izquierdo et al. (2020)[4].

[3] In a log-linear model, the price premium for a dummy variable *D* is computed as follows (Halvorsen and Palmquist, 1980): $\frac{\Delta p}{\Delta D} = \exp\left(\frac{\Delta \ln p}{\Delta D}\right) - 1$.

[4] Since the number of reviews could be understood as a proxy of demand (quantity of accommodates), there could be an endogeneity problem here. Because there are no valid instruments available, we test for endogeneity using the Lewbel's approach (Lewbel, 2012). The estimates are presented in Appendix C Table A5. Anderson LM test for under-identification and Cragg-Donald Wald F statistic for first-stage regression suggest the generated instruments are sufficiently correlated with *reviews* (relevance condition). The Sargan test indicates the overidentifying restrictions are valid (exogeneity condition). Finally, the Durbin-Wu-Hausman test does not reject the null hypothesis of exogeneity.

Dependent variable: *Ln P*	OLS	Quantile regression				
Explanatory variables	Coeff.	Coeff. (q=0.1)	Coeff. (q=0.25)	Coeff. (q=0.5)	Coeff. (q=0.75)	Coeff. (q=0.9)
entire	0.305***	0.366***	0.349***	0.482***	0.602***	0.665***
	(0.077)	(0.050)	(0.089)	(0.107)	(0.095)	(0.134)
bedrooms	0.027*	0.038***	0.035**	0.017	0.017	0.013
	(0.015)	(0.013)	(0.014)	(0.019)	(0.014)	(0.013)
capacity	0.046	-0.021	-0.018	0.094*	0.225***	0.342***
	(0.032)	(0.013)	(0.052)	(0.054)	(0.043)	(0.064)
capacity x entire	0.065**	0.118***	0.112**	0.014	-0.101**	-0.197***
	(0.033)	(0.016)	(0.050)	(0.055)	(0.044)	(0.065)
min.LOS	-0.026***	-0.019***	-0.014	-0.016***	-0.024***	-0.020***
	(0.007)	(0.004)	(0.010)	(0.005)	(0.004)	(0.006)
min.LOS x entire	0.036***	0.029***	0.023*	0.022**	0.030***	0.041**
	(0.010)	(0.006)	(0.013)	(0.009)	(0.010)	(0.016)
reviews	-0.002***	-4.0e-04	-2.5e-04	-0.001***	-0.002***	-0.003***
	(0.000)	(2.3e-04)	(1.7e-04)	(1.6e-04)	(1.4e-04)	(2.5e-04)
num.list.centre	-0.001	0.001	-0.001	-2.3e-04	-4.4e-04	-0.001*
	(0.001)	(0.001)	(0.001)	(3.9e-04)	(4.8e-04)	(0.001)
mhost	-0.230**	-0.153	-0.075	-0.167***	-0.218***	-0.230*
	(0.097)	(0.202)	(0.099)	(0.058)	(0.081)	(0.122)
mhost x entire	0.421***	0.359**	0.294***	0.326***	0.374***	0.405***
	(0.098)	(0.177)	(0.098)	(0.074)	(0.085)	(0.119)
mus.500	0.030***	0.025*	0.034***	0.036***	0.036***	0.037***
	(0.009)	(0.015)	(0.010)	(0.010)	(0.012)	(0.013)
mon.500	0.028**	0.018	0.042***	0.051***	0.037***	0.013
	(0.013)	(0.019)	(0.012)	(0.011)	(0.013)	(0.018)
Atocha.500	-0.049	0.007	-0.034	-0.010	-0.029	-0.084
	(0.035)	(0.030)	(0.036)	(0.037)	(0.040)	(0.061)
num.sub.less750	0.003**	0.002	0.003**	0.003***	0.003***	0.003
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Constant	3.727***	3.225***	3.432***	3.511***	3.602***	3.717***
	(0.075)	(0.048)	(0.082)	(0.100)	(0.088)	(0.114)
Parente-Santos Silva test for intra-cluster correlation [p-value]		22.899 [0.00]	25.136 [0.00]	26.796 [0.00]	23.387 [0.00]	18.875 [0.00]
Observations	4,308	4,308	4,308	4,308	4,308	4,308
R-squared	0.573	0.547	0.553	0.566	0.545	0.499

Table 3.- OLS and quantile regression parameter estimates
Standard errors adjusted for 2,636 clusters in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The number of listings the host has in the centre of Madrid (*num.list.centre*) is not significantly associated with price. However, hosts with more than 10 listings charge significantly lower prices, on average (-20%). This finding is consistent with Deboosere et al. (2019) and Oskam et al. (2019) but contradicts Wang and Nicolau (2017), Kwok and Xie (2019), Gibbs et al. (2018) and Chica-Olmo et al. (2020). More importantly, the estimates show that multi-hosts behave differently depending on the type of property. The owners of entire listings with more than 10 properties charge significantly higher prices (+52%). Therefore, the mixed evidence found in the literature about the effect of the number of listings on property can be partially explained by: i) non-linearities in multi-host behaviour so that differences become relevant from a certain threshold onwards, and ii) the heterogeneity in hosts' price setting depending on the type of property.

Concerning the accessibility measures, listings that are close to the main museums and monuments have a positive price premium. A marginal increase in the number of museums (*mus.500*) and monuments (*mon.500*) within a 500-metre radius raises average prices by 3% and 2.8%, respectively. This result is robust to the threshold considered. By contrast, there are no significant price differences depending on being within a 500-metre radius from Atocha station (*Atocha.500*). This finding is consistent with Önder et al. (2019). On the one hand, being close to the main train station could be highly priced due to the accessibility advantage it conveys. However, because Atocha station is usually highly crowded, tourists might dislike being close to it to avoid noise and congestion, which has been shown to be negatively related with Airbnb prices (Chica-Olmo et al., 2020). The latter is consistent with fair evidence about trade-offs between the benefits and disadvantages of being located closed to transportation hubs (Li and Brown, 1980). As for accessibility to the subway, the number of stops within a 750-metre radius from the listing (*num.sub.less750*) is positively related with prices (+0.3%).

Moving to the quantile regression, the price premium for full apartments (*entire*) is larger for high-priced accommodations than for cheap ones. By contrast, additional bedrooms only increase prices in the low-price segment, being non-significant for listings with prices above the median. Regarding *capacity*, the implicit price of an additional guest increases as far as we move to more expensive accommodations, in line with Wang and Nicolau (2017). Regarding the different behaviour of multi-hosts, we document an interesting finding. On average, hosts with more than ten properties (*mhost*) charge relatively lower prices in the high-price segment, whereas multi-hosts of entire apartments charge significantly higher prices across the whole distribution. This indicates that listings owned by professionals are relatively more expensive for entire apartments (especially as we move to the right tail of the price distribution) and relatively cheaper for shared/private rooms in the high-price segment. This result is partially in line with Wang and Nicolau (2017), who find that multi-hosts charge relative higher prices as we move to the upper part of the price distribution.

5.2. Robustness checks

We have conducted several robustness checks. First, we have re-estimated the model including a set of neighbourhood fixed effects (NFE) to see whether our estimates are affected by environmental factors (aesthetic attributes, public facilities, security, cultural diversity, etc.). Accessibility variables are excluded to avoid multicollinearity problems. Results are shown in the first column of Table A6 in Appendix C. The parameter estimates for the listing attributes are very similar with and without the SFE. Prices are (marginally) higher in *Justicia* neighbourhood and lower in *Embajadores* and *Palacio* neighbourhoods in comparison to *Sol* (reference category).

Second, we have replaced the dummy *mhost* defined as having more than 10 listings by a less strict alternative that takes value 1 if the host has two or more listings in the city centre (*mhost.1*). Results are shown in the second column of Table A6 in Appendix C. This dummy is now not significant. The same result is reported in Kakar et al. (2018). This indicates that 10 is a better threshold point to discriminate professionals from non-professionals. Consequently, host professionalization appears to operate from 10 listings onwards in our case study. Third, instead of considering the number of listings in the city centre, we replace it by the number of listings she has in the Community of Madrid (*num.list.CM*). The estimates are shown in the third column of Table A6 in Appendix C. Results are roughly the same.

Finally, alternative definitions for the accessibility measures were tested. We used the average distance to monuments (*dist.mus*) and museums (*dist.mon*), the minimum distance to the closest subway station (*min.dist.sub*) and the distance to Atocha station (*dist.Atocha*). Results are shown in the fourth column of Table A6 in Appendix C. Consistent with our main findings, prices are negatively related with the average distance to museums and monuments. Prices linearly increase as we move away from *Atocha* station. However, prices are higher as the minimum distance to the closest subway entrance increases, which appears to be counterintuitive. Note that by using *min.dist.sub* we impose that all the existing entrances provide the same level of accessibility to the subway, which might not be the case. Indeed, being really close to a subway entrance seems to have a price penalty, possibly because of negative externalities in the form of noise. Accordingly, prices increase with the number of different subway entrances in the neighbourhood (variety) but decrease when the listing is really close to a given entrance.

5.3. Spatial price hedonic modelling

Table 4 presents the Maximum Likelihood parameter estimates for the SAR model in equation (5) expanded with the three interaction terms. As before, standard errors are clustered at the host level.

Dependent variable: *Ln P*	
Explanatory variables	SAR
entire	0.369***
	(0.054)
bedrooms	0.025***
	(0.007)
capacity	0.052**
	(0.024)
capacity x entire	0.058**
	(0.025)
min.LOS	-5.8e-05***
	(2.9e-06)
min.LOS x entire	0.010**
	(0.005)
reviews	-0.001***
	(1.5e-04)
num.list.centre	-8.8e-04***
	(2.21e-04)
mhost	-0.255***
	(0.040)
mhost x entire	0.448***
	(0.039)
mus.500	-0.001
	(0.007)
mon.500	0.046***
	(0.009)
Atocha.500	0.008
	(0.028)
num.sub.less750	0.001
	(8.6e-04)
Constant	1.177***
	(0.050)
ρ	0.582***
	(0.055)
Observations	4,308
Log Likelihood	-2,518.14
AIC	5,070.3

Table 4.- SAR hedonic model parameter estimates
Standard errors adjusted for 2,636 clusters in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The spatial lag coefficient is positive and statistically significant, confirming our expectations of positive price dependence over space. The magnitude and statistical significance of the parameter estimates for the explanatory variables are similar to the OLS model in Table 3, although with some important differences. First, both bedrooms and capacity are now statistically significant at conventional levels. Second, prices linearly decrease as the host owns more listings. Third, once considering the spatial price dependence, the number of museums in the vicinity (*mus_500*) and the number of subway entrances (*num_sub_less750*) are not significant.

A proper interpretation of the effects of the explanatory variables on prices in the SAR model needs to consider not only their direct effects but also the indirect ones (Anselin, 2003). Following LeSage and Pace (2009) and Halleck-Vega and Elhorst (2015), the partial derivatives take the form of a NxN matrix for each explanatory variable *k* as follows:

$$\begin{bmatrix} \frac{\partial E(\ln P_1)}{\partial x_{1k}} & \cdots & \frac{\partial E(\ln P_1)}{\partial x_{Nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(\ln P_N)}{\partial x_{1k}} & \cdots & \frac{\partial E(\ln P_N)}{\partial x_{Nk}} \end{bmatrix} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \begin{pmatrix} \beta_k & \cdots & 0 \\ \cdots & \beta_k & \cdots \\ 0 & \cdots & \beta_k \end{pmatrix} = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} (\boldsymbol{I} \beta_k) = (\boldsymbol{I} - \rho \boldsymbol{W})^{-1} \beta_k \tag{8}$$

being $ln\ P$ a vector of dimension N, I the identity matrix of dimension NxN, ρ the lag coefficient, W the weight matrix of dimension NxN and β_k the parameter for variable *k*. Therefore, the dimension of the right-hand side of (8) is NxN.[5] A change in variable *k* in listing *j* exerts an effect in the price of listing *j* (direct effect) but also on the rest of listings (indirect effect) through the spatial dependence. The direct effect is given by the diagonal elements in (8) while the indirect effect is represented by the off-diagonal elements. Importantly, a change in attribute *k* in listing *j* will impact the price of listing *m* even through *j* and *m* are not neighbours. Intuitively, this happens due to shared neighbours that spread a shock over the full network.

Since the diagonal elements of $(I - \rho W)^{-1}$ vary per listing, each Airbnb has different direct and indirect effects. To facilitate interpretation, LeSage and Pace (2009) propose a way to summarize them. The direct effects can be measured by the average of the derivatives in the main diagonal of (8). The indirect effect will be computed as the average row sums of the off-diagonal elements. Accordingly, the indirect effect can be interpreted as the effect on the price of a change in variable *k* in all other listings on listing *j*. The total effect will be simply the sum of the two.

Columns 1-3 in Table 5 presents the direct and indirect effects obtained from the SAR estimates. For simplicity, we only report the ones that are significant at 95% level. To facilitate interpretation, columns 4-5 show the direct and indirect price premiums (in percentage). The direct effects of the attributes on prices are slightly lower than the ones obtained in the baseline OLS. However, the magnitude of the indirect effects is sizeable. This appears to confirm that price setting is heavily dependent on the prevailing prices in the neighbourhood. In this sense, conclusions obtained from a model that neglects this spatial dependence could be severely misleading.

Interestingly, the ratio between the direct and the indirect effect is about 0.58, which implies that the indirect effect is twice as large as the direct effect. For example, whereas the direct price premium for *entire* is 97%, the indirect effect amounts to 157%. This means that a shock by which the share of surrounding entire listings increased would rise prices by 157% through the price mimicking process. Similarly, whereas the direct effects for the number of bedrooms and the capacity are 2.5% and 8.8%, the indirect effects

[5] Note that the product of any matrix (in this case, $(I - \rho W)^{-1}$) and a scalar (β_k) is equivalent to the product of $(I - \rho W)^{-1}$ and $I\beta_k$.

amount to 3.5% and 12.2%, respectively. Again, neglecting the indirect effects in price setting would lead to deceitful implications.

Dependent variable: *ln P*	Estimated values			Price premium (%)	
Explanatory variables	Direct	Indirect	Total	Direct	Indirect
entire	0.370***	0.514***	0.884***	97	157
	(0.031)	(0.135)	(0.148)		
bedrooms	0.025***	0.035***	0.060***	2.5	3.5
	(0.007)	(0.013)	(0.019)		
capacity	0.052***	0.072***	0.125***	8.8	12.2
	(0.010)	(0.023)	(0.030)		
capacity x entire	0.058***	0.081***	0.140***		
	(0.011)	(0.026)	(0.035)		
reviews	-0.001***	-0.002***	-0.003***	-0.1	-0.2
	(1.6e-04)	(5.3e-04)	(6.1e-04)		
num.list.centre	-8.8e-04***	-0.001**	-0.002***	-0.08	-0.1
	(2.7e-04)	(5.1e-04)	(7.4e-04)		
mhost	-0.255***	-0.355***	-0.611***	3.9	5.5
	(0.036)	(0.098)	(0.118)		
mhost x entire	0.448***	0.623***	1.072***		
	(0.038)	(0.159)	(0.172)		
mon.500	0.046***	0.064**	0.110***	4.6	6.4
	(0.009)	(0.023)	(0.030)		

Table 5.- Direct, indirect and total effects

*** p<0.01, ** p<0.05, * p<0.1

5.4. Two-regime spatial hedonic model

Finally, Table 6 presents the parameter estimates for the two-regime Spatial Lag Model. In the first column we consider 10 listings as the cut-off point for distinguishing professional from non-professional hosts. In the second column, we are less strict and consider professional hosts those who own more than one listing. The parameter estimates for the structural and accessibility variables remain almost unchanged compared to Table 4, which provides further robustness to our results.

Note that this model uses the full sample and the two lag coefficients (ρ_1 and ρ_2) gather the effect of existing prices in the surrounding area on the prices charge by non-professionals and professionals, respectively. When professional status is defined as those who own more than 10 properties, the lag coefficient for non-professionals rises to 0.67 whereas the corresponding one for professionals is 0.25. A two-sided t-test (t=3.46, p-value<0.001) rejects the null hypothesis that the two parameters are equal, providing evidence on the different price dependence between professionals and non-professionals.

Dependent variable: *ln P*	Two-regime SAR	
Explanatory variables	mhost if num.list.centre>10	mhost if num.list.centre>1
entire	0.370***	0.395***
	(0.033)	(0.036)
bedrooms	0.024***	0.015**
	(0.006)	(0.006)
capacity	0.053***	0.069**
	(0.010)	(0.010)
capacity x entire	0.057***	0.050***
	(0.011)	(0.011)
min.LOS	-5.6e-05***	-5.7e-05***
	(4.3e-05)	(4.4e-05)
min.LOS x entire	0.010**	0.008*
	(0.004)	(0.004)
reviews	-0.001***	-0.001***
	(1.5e-04)	(1.5e-04)
num.list.centre	-0.001***	-0.001***
	(2.6e-04)	(2.1e-04)
mhost x entire	-0.461***	
	(0.037)	
mhost.1 x entire		0.151***
		(0.030)
mus.500	-2.5e-04	-0.003
	(0.008)	(0.006)
mon.500	0.043***	0.036***
	(0,009)	(0.009)
Atocha.500	0.004	0.017
	(0.033)	(0.033)
num.sub.less750	8.6e-04	3.5e-04
	(8.1e-04)	(8.2e-04)
alpha1	0.774***	0.629
	(0.282)	(0.384)
alpha2	2.351***	1.260***
	(0.457)	(0.313)
Rho1 (non-professional)	0.676***	0.706***
	(0.065)	(0.089)
Rho2 (professional)	0.251**	0.543***
	(0.106)	(0.072)
Observations	4,308	4,308
Log L	-2,512.16	-2,588.91
R-squared	0.572	0.561

Table 6.- Two-regime SAR hedonic model parameter estimates
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Interestingly, when the definition of multi-host is less strict, the lag price coefficients are more similar (0.70 and 0.54), although the spillover effect is still greater for non-professionals. However, in this case a two-sided t-test (p-value=0.15) does not reject the null of parameter equality. The latter result seems to indicate that the 10 listings is a better threshold for distinguishing professionals from non-professionals. The estimates suggest that non-professional operate under yardstick competition, by which they strongly imitate the prices in the neighbouring area. Conversely, professionals behave in a much more independent way so that they can be assumed to truly operate under monopolistic competition. As a result, a shock in the mean characteristics in the neighbourhood would translate into a larger shift in the same direction for non-professionals than for professionals.

5. Discussion and conclusions

Consistent with the recent literature that has put forward the existence of spatial price dependence in Airbnb rental market, this paper studies the potential existence of two different spatial price processes. Based on host professionalization status, we have examined whether yardstick price competition is more prevalent among non-professional hosts. We have specifically hypothesized that those who truly share their underutilised lodgings in exchange of a fee may set prices in a *rationally ignorant* way (i.e. mimicking the prevailing prices in the vicinity). However, professional hosts who operate as formal accommodation providers might seek to maximize revenues and set prices based on competitive monopoly equilibrium conditions.

Using web-scraped data from the city centre of Madrid, we have conducted standard and spatial hedonic regressions. We have first estimated a baseline OLS model with listing-specific features and accessibility measures to sightseeing spots and transportation hubs. To assess the potential different effect of the listing characteristics over the price distribution, we have also run quantile regression. The results from this baseline analysis indicate that entire apartments exhibit a price premium of about 107%. Prices increase with capacity and the number of bedrooms but are negatively related to the number of reviews. Interestingly, average prices for professional hosts are lower in general but higher for entire apartments. Regarding location, the number of museums and monuments in a 500-metre radius positively impacts prices, while closeness to Atocha train station does not significantly impact prices, possibly due to nuisance effects. Accessibility to subway stops is positively valued.

Subsequently, we have estimated a Spatial Hedonic Lag Model that enables us to separate the direct from the indirect effects of the explanatory variables through price dependence. The estimates show that the indirect effects are twice as large as the direct ones. This suggests that any shock in the mean characteristics in the vicinity exerts important *spillover* effects. Therefore, ignoring spatial price dependences in hedonic price modelling would produce biased estimates and could lead to misleading results. To disentangle the differences in spatial price dependencies between professional and non-professional hosts, we have estimated a two-regime SAR model. When we consider more than 10 listings as the cut-off point, we have shown that the price dependence is significantly different between professionals and non-professionals. While in the former case hosts seem to be less affected by the prices in the neighbourhood, in the latter they appear to strongly imitate the prevailing price levels in the area.

According to this evidence, it seems that the Airbnb rental market is composed of two types of owners that set prices differently. The cut-off point that distinguishes professional from non-professional hosts is around ten listings. Importantly, when we are less strict and define multi-host as those who have more than one listing, the two coefficients for spatial dependence are not statistically different. This implies that the more-than-one but less-than-ten segment behaves close to *rationally ignorant* firms in the sense that their price setting is more similar to non-professional hosts. Professionalization thus seems to emerge from ten listings onwards. We speculate that this threshold point might be related with the exploitation of economies of scale.

Our study has some relevant policy implications. There is nowadays a public debate about whether Airbnb services should be regulated and in which ways they should be. Given the important gentrification it is taking place in the city centre of Madrid as a consequence of Airbnb growth (Ardura-Urquiaga, 2020), policy makers are developing different strategies to cut the continuous growth in Airbnb supply. This policy debate focuses on how to regulate commercially oriented multi-unit hosts. Following the practise of other European cities like Amsterdam, Paris or London, the council of Madrid has recently passed a new plan that aims to prevent hosts from renting their listings for more than 90 days per year. Another policy regulation could be the implementation of a tourism tax to Airbnb rentals managed by professionals. In the light of our findings, policy makers should be aware that price regulations that pursue professional hosts would not be neutral in the sense that they would spread over the neighbouring listings. Indeed, since non-professionals imitate prices in the vicinity more strongly, a tax aimed to discourage commercially oriented supply could translate into larger prices in the ‘true sharing’ segment that in the professional one.

REFERENCES

Adamiak, C. (2018). Mapping Airbnb supply in European cities. *Annals of Tourism Research*, 71, 67-71. https://doi.org/10.1016/j.annals.2018.02.008

Airbnb (2019). *Informe de actividad económica de Airbnb en la ciudad de Madrid*. Retrieved from https://press.Airbnb.com/wp-content/uploads/sites/4/2019/03/2019-Madrid-Economic-Activity-Report.pdf

Ardura Urquiaga, A., Lorente-Riverola, I. and Ruiz Sánchez, J. (2020). Platform-mediated short-term rental and gentrification in Madrid. *Urban Studies*, https://doi.org/10.1177/0042098020918154.

Allers, M.A. and Elhorst, J.P. (2005). Tax Mimicking and Yarstick Competition Among Local Governments in the Netherlands. *International Tax and Public Finance*, 12, 493-513. https://doi.org/10.1007/s10797-005-1500-x

Anselin, L. (2003). Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review*, 26, 153–166. https://doi.org/10.1177/0160017602250972

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, 5-34. https://doi.org/10.1007/s00181-007-0152-3

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*, https://doi.org/10.1177/0160017618821428

Benítez-Aurioles, B. (2018a). The role of distance in the peer-to-peer market for tourist accommodation. *Tourism Economics*, 24(3), 237-250. https://doi.org/10.1177/1354816617726211

Benítez-Aurioles, B. (2018b). Why are flexible booking policies priced negatively? *Tourism Management*, 67, 312-325. https://doi.org/10.1016/j.tourman.2018.02.008

Chamberlin, E. (1933). *Theory of Monopolistic Competition*. Cambridge: Harvard University Press.

Chattopadhyay, M. and Mitra, S.K. (2019). Do Airbnb host listing attributes influence room pricing homogeneously? *International Journal of Hospitality Management*, 81, 54-64. https://doi.org/10.1016/j.ijhm.2019.03.008

Chen, Y., and Xie, K. (2017). Consumer valuation of Airbnb listings: A hedonic pricing approach. *International Journal of Contemporary Hospitality Management*, 29(9), 2405-2424. https://doi.org/10.1108/ijchm-10-2016-0606

Chica-Olmo, J., González-Morales, J.G. and Zafra-Gómez, J.L. (2020). Effects of location on Airbnb apartment pricing in Malaga. *Tourism Management*, https://doi.org/10.1016/j.tourman.2019.103981

Cordera, R., Coppola, P., dell'Olio, L. and Ibeas, A. (2019). The impact of accessibility by public transport on real estate values: a comparison between the cities of Rome and Santander. *Transportation Research Part A*, 125, 308-319. https://doi.org/10.1016/j.tra.2018.07.015

Deboosere, R., Kerrigan, D.J., Wachsmuth, D. and El-Geneidy, A. (2019). Location, location and professionalization: a multilevel hedonic analysis of Airbnb listing prices and revenue. *Regional Studies, Regional Science*, 6(1), 143-156. https://doi.org/10.1080/21681376.2019.1592699

Delgado, F.J. and Mayor, M. (2011). Tax mimicking among local governments: Some evidence from Spanish municipalities. *Portuguese Economic Journal*, 10, 149–164. https://doi.org/10.1007/s10258-010-0067-3

Delgado, F.J., Lago-Peñas, S. and Mayor, M. (2015). On the determinants of local tax rates: new evidence from Spain. *Contemporary Economic Policy*, 33, 351–368. https://doi.org/10.1111/coep.12075

Dogru, T., Mody, M., Suess, C., Line, N. and Bonn, M. (2020). Airbnb 2.0: Is it a sharing economy platform or a lodging corporation? *Tourism Management*, https://doi.org/10.1016/j.tourman.2019.104049

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

Elhorst J.P. and Fréret S. (2009). Evidence of political yardstick competition in France using a two-regime spatial Durbin model with fixed effects. *Journal of Regional Science*, 49, 931-951. https://doi.org/10.1111/j.1467-9787.2009.00613.x
Ert. E. and Fleischer, A. (2019). The evolution of trust in Airbnb: A case of home rental. *Annals of Tourism Research*, 75, 279-287. https://doi.org/10.1016/j.annals.2019.01.004
Eugenio-Martín, J.L., Cazorla-Artiles, J.M. and González-Martel, C. (2019). On the determinants of Airbnb location and its spatial distribution. *Tourism Economics*, 25(8), 1224-1244. https://doi.org/10.1177/1354816618825415
Falk, M., Larpin, B. and Scaglione, M. (2019). The role of specific attributes in determining prices of Airbnb listings in rural and urban locations. *International Journal of Hospitality Management*, 83, 132-140. https://doi.org/10.1016/j.ijhm.2019.04.023
Farmaki, A. and Kaniadakis, A. (2020). Power dynamics in peer-to-peer accommodation: insights from Airbnb hosts. *International Journal of Hospitality Management*, https://doi.org/10.1016/j.ijhm.2020.102571
Franco, S.F. and Macdonald, J.L. (2018). The effects of cultural heritage on residential property values: Evidence from Lisbon, Portugal. *Regional Science and Urban Economics*, 70, 35-56. https://doi.org/10.1016/j.regsciurbeco.2018.02.001
Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., and Goodwill, A. (2018). Pricing in the sharing economy: A hedonic pricing model applied to Airbnb listings. *Journal of Travel & Tourism Marketing*, 35(1), 46-56. https://doi.org/10.1080/10548408.2017.1308292
Gil, J. and Sequera, J. (2020). The professionalization of Airbnb in Madrid: far from collaborative economy. *Current Issues in Tourism*. DOI:10.1080/13683500.2020.1757628.
Gunter, U. and Önder, I. (2018). Determinants of Airbnb demand in Vienna and their implications for the traditional accommodation industry. *Tourism Economics*, 24(3), 270-293. https://doi.org/10.1177/1354816617731196
Gunter, U., Önder, I. and Zekan, B. (2020). Modeling Airbnb demand to New York City while employing spatial panel data at the listing level. *Tourism Management*, https://doi.org/10.1016/j.tourman.2019.104000
Gyódi, K. (2019). Airbnb in European cities: business as usual or true sharing economy? *Journal of Cleaner Production*, 221, 536-551.
Halleck Vega, S. and Elhorst, J.P. (2015). The SLX model. *Journal of Regional Science*, 55 (3), 339–363. https://doi.org/10.1111/jors.12188
Halvorsen, R. and Palmquist, R. (1980). The interpretation of dummy variables in semilogarithmic equations. *American Economic Review*, 70, 474–475.
Hill, D. (2015). How much is your spare room worth? *IEEE Spectrum*, 52 (9). https://doi.org/10.1109/mspec.2015.7226609
Kakar, V., Voelz, J., Wu, J. and Franco, J. (2018). The Visible Host: Does race guide Airbnb rental rates in San Francisco? *Journal of Housing Economics*, 40, 25-40. https://doi.org/10.1016/j.jhe.2017.08.001
Karlsson, L. and Dolnicar, S. (2016). Someone's been sleeping in my bed. *Annals of Tourism Research*, 58, 159-162. https://doi.org/10.1016/j.annals.2016.02.006
Kwok, L. and Xie, L. (2019). Pricing strategies on Airbnb: are multi-unit hosts revenue pros? *International Journal of Hospitality Management*, 82, 252-259. https://doi.org/10.1016/j.ijhm.2018.09.013
Lawani, A., Reed, M.R., Mark, T. and Zheng, Y. (2019). Reviews and price on online platforms: evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics*, 75, 22-34. https://doi.org/10.1016/j.regsciurbeco.2018.11.003
Lazrak, F., Nijkamp, P., Rietveld, P. and Rouwendal, J. (2014). The market value of cultural heritage in urban areas: an application of spatial hedonic pricing. *Journal of Geographical Systems*, 16, 89-114. DOI 10.1007/s10109-013-0188-1
LeSage, J. and Pace, R.K. (2009). *Introduction to Spatial Econometrics*. CRC Press. https://doi.org/10.1201/9781420064254
Lewbel, A. (2012). Using heterocedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business Economics & Statistics*, 30(1), 67-80. https://doi.org/10.1080/07350015.2012.643126

Li, H. and Srinivasan, K. (2019). Competitive Dynamics in the Sharing Economy: an Analysis in the Context of Airbnb and Hotels. *Marketing Science*, 38(3), 365-391. https://doi.org/10.2139/ssrn.3197969

Li, M.M. and Brown, H.J. (1980). Micro-Neighbourhood Externalities and Hedonic Housing Prices. *Land Economics*, 56(2), 125-141. https://doi.org/10.2307/3145857

López, F.A., Mínguez, R. and Mur, J. (2019). ML versus IV estimates of spatial SUR models: evidence from the case of Airbnb in Madrid urban area. *The Annals of Regional Science*, https://doi.org/10.1007/s00168-019-00914-1

Moreno-Izquierdo, L., Rubia-Serrano, A., Perles-Ribes, J.F., Ramón-Rodríguez, A.B. and Such-Devesa, M.J. (2020). Determining factors in the choice of prices of tourist rental accommodation. New evidence using the quantile regression approach. *Tourism Management Perspectives*, https://doi.org/10.1016/j.tmp.2019.100632.

Önder, I., Weismayer, C. and Gunter, U. (2019). Spatial price dependencies between the traditional accommodation sector and the sharing economy. *Tourism Economics*, 25(8), 1150-1166. https://doi.org/10.1177/1354816618805860

Oskam, J., Van der Rest, J.P. and Telkamp, B. (2019). What's mine is yours - but at way price? Dynamic pricing behavior as an indicator of Airbnb host profesionalization. *Journal of Revenue Pricing Management*, 17(5), 311-318. https://doi.org/10.1057/s41272-018-00157-3

Padovano, F. and Petrarca, I. (2014). Are the responsibility and yardstick competition hypotheses mutually consistent? *European Journal of Political Economy*, 34, 459-477. https://doi.org/10.1016/j.ejpoleco.2013.09.006

Parente, P.M.D.C. and Santos Silva, J.M.C. (2016). Quantile regression with clustered data. *Journal of Econometric Methods*, 5, 1-15. https://doi.org/10.1515/jem-2014-0011

Rietveld, P. and P. Wintershoven. (1998). Border Effects and Spatial Autocorrelation in the Supply of Network Infrastructure. *Papers in Regional Science*, 77, 265–276. https://doi.org/10.1111/j.1435-5597.1998.tb00717.x

Rosen, S. (1974). Hedonic prices and implicit markets: products differentiation in pure competition. *Journal of Political Economy*, 82, 34-55. https://doi.org/10.1086/260169

Sainaghi, R. (2020). The current state of academic research into peer-to-peer accommodation platforms. *International Journal of Hospitality Management*, https://doi.org/10.1016/j.ijhm.2020.102555

Sainaghi, R., Abrate, G. and Mauri, A. (2021). Price and RevPAR determinants of Airbnb listings: convergent and divergent evidence. *International Journal of Hospitality Management*, https://doi.org/10.1016/j.ijhm.2020.102709.

Salas-Olmedo, M.H., García-Palomares, J.C. and Gutiérrez, J. (2018). Tourists' digital footprint in cities: comparing big data sources. *Tourism Management*, 66, 13-25. https://doi.org/10.1016/j.tourman.2017.11.001

Shleifer, A. (1985). A theory of yardstick competition. *RAND Journal of Economics*, 16(3), 319-327. https://doi.org/10.2307/2555560

Smith, T.E. (2009). Estimating Bias in Spatial Models with Strongly Connected Weight Matrices. *Geographical Analysis*, 41, 307-332. https://doi.org/10.1111/j.1538-4632.2009.00758.x

Tang, L., Kim, J. and Wang, X. (2019). Estimating spatial effects on peer-to-peer accommodation prices: towards an innovative hedonic model approach. *International Journal of Hospitality Management*, 81, 43-53. https://doi.org/10.1016/j.ijhm.2019.03.012

Varma, A., Jukic, N., Pestek, A., Shultz, C.J. and Nestorov, S. (2016). Airbnb: exciting innovation or passing fad? *Tourism Management Perspectives*, 20, 228-237. https://doi.org/10.1016/j.tmp.2016.09.002

Wang, D., and Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management*, 62, 120-131. https://doi.org/10.1016/j.ijhm.2016.12.007

Wegmann, J. and Jiao, J. (2017). Taming Airbnb: toward guiding principles for local regulation of urban vacation rentals based on empirical results from five US cities. *Land Use Policy*, 69, 494-501. https://doi.org/10.1016/j.landusepol.2017.09.025

Xie, K. and Mao, E. (2017). The impacts of quality and quantity attributes of Airbnb hosts on listing performance. *International Journal of Contemporary Hospitality Management*, 29(9), 2240-2260. https://doi.org/10.1108/ijchm-07-2016-0345

Yang, L., Chau, K.W., Szeto, W.Y., Cui, X. and Wang, X. (2020a). Accessibility to transit, by transit, and property prices: spatially varying relationships. *Transportation Research Part D*, https://doi.org/10.1016/j.trd.2020.102387.

Yang, L., Chu, X. Gou, Z., Yang, H., Lu, Y. and Huang, W. (2020b). Accessibility and proximity effects of bus rapid transit on housing prices: heterogeneity across price quantiles and space. *Journal of Transport Geography*, https://doi.org/10.1016/j.jtrangeo.2020.102850.

APPENDIX

SPATIAL PRICE MIMICKING ON AIRBNB: MULTI-HOST VS SINGLE-HOST

APPENDIX A.- Sightseeing spots considered in the analysis

Museums

ID	Description	District	Neighbourhood
1	*Congreso de los Diputados*	*CENTRO*	*CORTES*
2	Spanish National Library	*SALAMANCA*	*RECOLETOS*
3	*Museo Nacional de Artes Decorativas*	*RETIRO*	*LOS JERONIMOS*
4	Natural Science National Museum	*CHAMARTIN*	*EL VISO*
5	El Prado Museum	*RETIRO*	*LOS JERONIMOS*
6	*Casón del Buen Retiro*	*RETIRO*	*LOS JERONIMOS*
7	Madrid's Train Museum	*ARGANZUELA*	*DELICIAS*
8	Thyssen-Bornemisza National Museum	*CENTRO*	*CORTES*
9	*Museo de la Real Academia de Bellas Artes de San Fernando*	*CENTRO*	*SOL*
10	*Plaza Monumental de Toros de las Ventas*	*SALAMANCA*	*GUINDALERA*
11	Velázquez's House	*MONCLOA-ARAVACA*	*CIUDAD UNIVERSITARIA*
12	*Galería de Cristal del Palacio de Cibeles*	*RETIRO*	*LOS JERONIMOS*
13	Botanical Garden	*RETIRO*	*LOS JERONIMOS*

Table A1.- List of museums

Monuments

ID	Description	District	Neighbourhood
14	*Puerta del Sol*	*CENTRO*	*SOL*
15	Almudena Cathedral	*CENTRO*	*PALACIO*
16	El Pardo Royal Palace	*FUENCARRAL-EL PARDO*	*EL PARDO*
17	*Plaza Mayor*	*CENTRO*	*SOL*
18	Debod Temple	*MONCLOA-ARAVACA*	*CASA DE CAMPO*
19	Madrid Royal Palace	*CENTRO*	*PALACIO*

Table A2.- List of monuments

*Note: the names in italics preserve the original name in Spanish since translation would be meaningless.

APPENDIX B.- Information criteria for choosing distance threshold

Threshold value	log L	AIC	LM test p-value
100-metre	-2691.74	5409.5	<0.01
150-metre	-2676.03	5378.1	<0.01
200-metre	-2660.96	5347.9	<0.01
250-metre	-2653.97	5333.9	0.002
300-metre	-2651.71	5329.4	0.08
350-metre	**-2650.98**	**5328.0**	**0.66**
400-metre	-2654.4	5334.8	0.8
450-metre	-2653.13	5332.3	0.71
500-metre	-2656.53	5339.1	0.16

Table A3.- Log likelihood, AIC and LM test for different weight matrix distance thresholds

APPENDIX C.- Robustness checks (I): extended regression with all the interaction terms

Dependent variable: *Ln P*	OLS	Quantile regression				
Explanatory variables	Coeff.	Coeff. (q=0.1)	Coeff. (q=0.25)	Coeff. (q=0.5)	Coeff. (q=0.75)	Coeff. (q=0.9)
entire	0.314***	0.327***	0.408***	0.561***	0.590***	0.709***
	(0.091)	(0.087)	(0.078)	(0.097)	(0.116)	(0.142)
bedrooms	0.039	0.003	0.021	0.024	0.033	0.010
	(0.030)	(0.045)	(0.036)	(0.034)	(0.032)	(0.017)
bedrooms x entire	-0.015	0.035	0.016	-0.010	-0.023	-0.004
	(0.034)	(0.049)	(0.039)	(0.046)	(0.036)	(0.025)
capacity	0.045	-0.010	-0.010	0.099**	0.217***	0.332***
	(0.032)	(0.014)	(0.060)	(0.050)	(0.048)	(0.062)
capacity x entire	0.068**	0.104***	0.104*	0.010	-0.088*	-0.184***
	(0.034)	(0.017)	(0.060)	(0.054)	(0.050)	(0.063)
Min.LOS	-0.028***	-0.017***	-0.015	-0.014***	-0.024***	-0.023***
	(0.008)	(0.005)	(0.016)	(0.005)	(0.004)	(0.006)
min.LOS x entire	0.038***	0.025***	0.024	0.020**	0.034***	0.046**
	(0.010)	(0.007)	(0.016)	(0.009)	(0.012)	(0.021)
reviews	-0.002***	7.9e-05	1.4e-05	-0.001***	-0.002***	-0.002**
	(0.000)	(4.8e-04)	(3.0e-04)	(0.000)	(0.000)	(0.001)
reviews x entire	8.4e-05	-8.8e-05	-4.8e-04	-1.3e-04	-3.6e-04	-6.4e-04
	(4.9e-04)	(5.2e-04)	(3.7e-04)	(3.7e-04)	(3.7e-04)	(9.6e-04)
num.list.centre	-0.001	8.6e-05	-4.2e-04	-0.001	-0.000	-0.001
	(0.001)	(9.2e-04)	(0.001)	(0.001)	(0.001)	(0.001)
num.list.centre x entire	-0.001	0.001	-0.000	0.001	-0.001	-0.001
	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)
mhost	-0.259**	-0.054	-0.087	-0.164*	-0.249***	-0.303***
	(0.122)	(0.149)	(0.107)	(0.084)	(0.094)	(0.115)
mhost x entire	0.460***	0.264*	0.300**	0.324***	0.415***	0.484***
	(0.134)	(0.153)	(0.125)	(0.098)	(0.111)	(0.140)
mus.500	0.011	0.007	0.007	-0.007	0.013	-0.008
	(0.020)	(0.031)	(0.040)	(0.026)	(0.018)	(0.016)
mus.500 x entire	0.025	0.027	0.035	0.058**	0.037*	0.047**
	(0.022)	(0.033)	(0.041)	(0.027)	(0.020)	(0.019)
mon.500	0.045	-0.004	0.040*	0.049**	0.039**	0.032

	(0.031)	(0.047)	(0.024)	(0.019)	(0.020)	(0.033)
mon.500 x entire	-0.024	0.033	0.000	0.007	-0.015	-0.028
	(0.034)	(0.050)	(0.025)	(0.024)	(0.026)	(0.039)
Atocha.500	-0.105	-0.017	-0.011	0.027	-0.121**	-0.141
	(0.064)	(0.089)	(0.063)	(0.061)	(0.058)	(0.123)
Atocha.500 x entire	0.078	0.038	-0.033	-0.060	0.078	0.075
	(0.075)	(0.095)	(0.077)	(0.074)	(0.070)	(0.139)
Num.sub.less750	0.003	-3.0e-04	0.005*	0.007***	0.004	0.006
	(0.002)	(0.005)	(0.003)	(0.002)	(0.003)	(0.006)
Num.sub.less750 x entire	-0.001	0.003	-0.004	-0.005**	-0.001	-0.004
	(0.003)	(0.005)	(0.003)	(0.002)	(0.003)	(0.006)
Constant	3.716***	3.265***	3.394***	3.453***	3.607***	3.701***
	(0.085)	(0.081)	(0.068)	(0.091)	(0.108)	(0.124)
Parente-Santos Silva test for		23.265	24.417	25.770	25.135	18.691
intra-cluster correlation [p-value]		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Observations	4,308	4,308	4,308	4,308	4,308	4,308
R-squared	0.574	0.543	0.554	0.565	0.548	0.504

Table A4.- Extended OLS and quantile regression parameter estimates with all the interaction terms
Standard errors adjusted for 2,636 clusters in parentheses. *** p<0.01, ** p<0.05, * p<0.1

APPENDIX C.- Robustness checks (II): endogeneity check for the number of reviews

Dependent variable: *Ln P* Explanatory variables	IV (2SLS) Coeff.
entire	0.613*** (0.018)
bedrooms	0.024*** (0.007)
capacity	0.111*** (0.005)
min.LOS	-0.012*** (0.003)
reviews	-0.001*** (0.000)
num.list.centre	-0.002*** (0.000)
mhost	0.102*** (0.020)
mus.500	0.026*** (0.008)
mon.500	0.012 (0.010)
Atocha.500	-0.043 (0.035)
num.sub.less750	0.002** (0.001)
Constant	3.528*** (0.022)
Anderson LM statistics for underidentification [p-value]	3479.97 [0.00]
Cragg-Donald Wald F statistic for first-stage [p-value]	1801.70
Sargan statistic for overidentification	11.39 [0.249]
Durbin-Wu-Hausman test of exogeneity	1.53 [0.215]
Observations	4,308
R-squared	0.553

Table A5.- IV regression using Lewbel-based instruments for the number of reviews
Standard errors adjusted for 2,636 clusters in parentheses. *** p<0.01, ** p<0.05, * p<0.1

APPENDIX C.- Robustness checks (III): model specification

Dependent variable: Ln P Explanatory variables	(1)	(2)	(3)	(4)
entire	0.298***	0.299***	0.302***	0.296***
	(0.076)	(0.081)	(0.077)	(0.077)
bedrooms	0.028*	0.018	0.023**	0.027*
	(0.015)	(0.013)	(0.011)	(0.015)
capacity	0.044	0.057	0.046	0.042
	(0.031)	(0.035)	(0.032)	(0.031)
capacity x entire	0.069**	0.062*	0.068**	0.070**
	(0.032)	(0.036)	(0.032)	(0.032)
min.LOS	-0.025***	-0.037***	-0.025***	-0.024***
	(0.007)	(0.008)	(0.008)	(0.007)
min.LOS x entire	0.034***	0.044***	0.034***	0.033***
	(0.010)	(0.011)	(0.010)	(0.010)
reviews	-0.001***	-0.002***	-0.002***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
num.list.centre	-0.001	-0.001***		-0.001
	(0.001)	(0.000)		(0.001)
num.list.CM			-0.001	
			(0.001)	
mhost	-0.223**		-0.202**	-0.228**
	(0.089)		(0.096)	(0.092)
mhost x entire	0.409***		0.432***	0.418***
	(0.092)		(0.087)	(0.096)
mhost_1		-0.062		
		(0.051)		
mhost_1 x entire		0.135**		
		(0.056)		
d_justicia	0.069*			
	(0.040)			
d_embaja	-0.164***			
	(0.034)			
d_cortes	0.020			
	(0.040)			
d_uni	-0.072*			
	(0.037)			
d_palacio	-0.068**			
	(0.031)			
mus.500		0.029***	0.029***	
		(0.008)	(0.009)	
mon.500		0.021*	0.026*	
		(0.012)	(0.013)	
Atocha.500		-0.044	-0.051	
		(0.037)	(0.034)	
num.sub.less750		0.002**	0.003**	
		(0.001)	(0.001)	
dist.mus				-0.294***
				(0.041)
dist.mon				-0.110**
				(0.052)
dis.Atocha				0.086***
				(0.031)
min.dist.sub				0.249***
				(0.085)
Constant	3.549***	3.744***	3.732***	4.495***
	(0.046)	(0.077)	(0.074)	(0.223)
Observations	4,308	4,308	4,308	4,308
R-squared	0.581	0.560	0.574	0.580

Table A6.- OLS parameter estimates from different model specifications

Standard errors adjusted for 2,636 clusters in parentheses. *** p<0.01, ** p<0.05, * p<0.1