Tourism Management 86 (2021) 104319

Contents lists available at ScienceDirect

Tourism Management

journal homepage: http://www.elsevier.com/locate/tourman

Determinants of Airbnb prices in European cities: A spatial econometrics approach

Kristóf Gyódi *, Łukasz Nawaro

University of Warsaw, Faculty of Economic Sciences, Długa 44/50, 00-241, Warsaw, Poland

ARTICLE INFO

Keywords:
Airbnb
Tourism
Digitalization
Spatial econometrics

ABSTRACT

We examine the determinants of Airbnb prices in 10 major EU cities, focusing on the role of location. The results confirm that attributes related to size, quality, and location are all significant drivers of Airbnb rates. Novel indices based on TripAdvisor data are used to measure the attractiveness of neighbourhoods, and the results show a more robust impact on price than standard location variables based on selected points of interest. The analysis confirms that Airbnb prices are spatially dependent, requiring the implementation of spatial regression models. Following recent studies on spatial econometrics, we examine various spatial models, including specifications with multiple sources of spatial dependence. The results show significant differences between the coefficients estimated with OLS and the various spatial models, especially in the case of location-specific variables. As well as having managerial and policy implications, our study contributes to the hedonic price literature by providing a methodological guide on spatial regression models.

1. Introduction

The short-term home rental market has been growing dynamically across major tourist destinations in recent years. While peer-to-peer accommodation provision is not an entirely new concept, the emergence of online platforms and the smartphone revolution have drastically changed its scale (Guttentag, 2015). Airbnb is the most important platform for short-term home sharing, having expanded rapidly across the world. According to Airbnb, an average of two million travellers are staying in accommodations provided via the platform every night (Airbnb, 2019).

The managerial aspects of Airbnb accommodation provision have been examined in a number of studies (Kwok & Xie, 2018; Lutz & Newlands, 2018; Oskam et al., 2018). A popular research method in this field is hedonic price analysis. The hedonic price model, based on Lancaster's characteristics theory, assumes that the price of a product reflects the consumer evaluation of the product's various attributes (Lancaster, 1966; Tang et al., 2019). Therefore, the final price is the sum of the consumer's willingness to pay for each attribute (Rosen, 1974). Previous studies on Airbnb have shown that size, quality, and host reputation play a significant role in hedonic price regressions (Gibbs et al., 2018; Teubner & Dann, 2017; Wang & Nicolau, 2017).

The hospitality literature supports the assumption that location is a key factor in hotel prices (Kim et al., 2018; Zhang et al., 2011). Empirical studies have shown that tourists value various location attributes, such as distance from the city centre, access to public transportation, and proximity to attractions (Yang et al., 2018). However, previous hedonic price analyses of Airbnb listings included only simple location variables, without exploring the various components of the location price premium (Gibbs et al., 2018; Wang & Nicolau, 2017).

Moreover, most of these works did not consider that Airbnb listings from the same area share latent location-specific attributes. When the values of a variable at a given location depend on the values of observations at neighbouring locations, the observations are spatially dependent (LeSage & Pace, 2009). The Ordinary Least Squares (OLS) model ignores the spatial dependence of Airbnb prices, leading to biased results, especially in the case of location variables. While there is an established literature on spatial regression models, only a few studies have used them for the analysis of the short-term rental market (Lawani et al., 2019; Tang et al., 2019). These examples have demonstrated that there are significant differences in the magnitudes of estimated coefficients for location variables between OLS and spatial models. Moreover, spatial models provide a better fit for data with spatial heterogeneity. However, the existing Airbnb literature considered only a subset of spatial models, while the state-of-the-art in spatial econometrics suggests that a more complex model selection process is needed.

* Corresponding author.
E-mail addresses: kgyodi@wne.uw.edu.pl (K. Gyódi), l.nawaro@uw.edu.pl (Ł. Nawaro).

https://doi.org/10.1016/j.tourman.2021.104319
Received 18 March 2020; Received in revised form 18 January 2021; Accepted 9 March 2021
Available online 8 April 2021

An accurate assessment of Airbnb price determinants is essential not only for tourism management, but also for urban planning. Various studies have shown that Airbnb is increasingly professionalised, transforming housing supply into tourist accommodations. This process has contributed to the housing shortage in major cities, and to the gentrification of neighbourhoods. Therefore, gaining a better understanding of the role of location in the price premium has considerable policy relevance.

The main aim of this study is to assess the impact of Airbnb listing attributes on price. Besides testing the influence of size, guest ratings, and host characteristics, various location variables will be examined. We evaluate two main hypotheses in this study:

H1. Airbnb prices are spatially dependent.

The real estate market is characterised by strong spatial dependence, as homes in a given area share a set of common attributes that affect price (Tang et al., 2019). This effect has been explored in the case of the traditional hotel industry (Kim et al., 2018; Soler & Gemar, 2018). While the literature on Airbnb has identified the spatial dependence of prices (Tang et al., 2019; Zhang et al., 2017), the different sources of these spatial relationships have yet to be fully explored. In our analysis, we will test various spatial processes (such as the influence of the attributes of neighbouring listings), and apply appropriate spatial models.

H2. Location attributes significantly affect Airbnb prices.

The attractiveness of the listing area may be influenced not only by the linear distance from selected points, but by the overall accessibility of tourist venues and points of interest. Therefore, following recent literature on guest preferences (Yang et al., 2018), we implement novel indices to measure the attractiveness of neighbourhoods based on TripAdvisor data.

In particular, we formulate the following hypotheses:

H2a. There is a significant and negative relationship between the price and the distance to the city centre.

H2b. There is a significant and negative relationship between the price and the distance to the nearest metro station.

H2c. There is a significant and positive relationship between the price and the attractiveness of the listing's neighbourhood.

In our analysis, we examine the price determinants in 10 major EU cities: Amsterdam, Athens, Barcelona, Berlin, Budapest, Lisbon, London, Paris, Rome, and Vienna. These cities represent major destinations of urban tourism in various regions of Europe (Western Europe, Central Europe, and Southern Europe). The sample includes cities that differ in terms of their size (area and population), housing market conditions, and Airbnb regulatory frameworks. This large and varied sample allows for the formulation of more robust conclusions.

2. Literature review

The hedonic price model has been widely used in the analysis of hotel prices (Espinet et al., 2003; White & Mulligan, 2002; Zhang et al., 2011). Wang & Nicolau (2017) and Kim et al. (2018) provide an overview of such studies, and conclude that hotel prices are mainly determined by location, class, size, customer ratings, amenities, and other property characteristics. The tested location variables include the distance to the city centre (Bull, 1994), the central train station (Thrane, 2007), the major tourist district (Carvell & Herrin, 1990), the airport, and the main business district (Lee & Jang, 2011).

The earlier hedonic price analyses on Airbnb listings, which were inspired by the literature on hotel prices, tested the effects of size, rating scores, popularity, host characteristics, amenities, and distance to the city centre (Gibbs et al., 2018; Teubner & Dann, 2017; Wang & Nicolau, 2017). The results confirmed that listings with more space, better reviews, and greater proximity to the city centre are more expensive. The relationship with the number of reviews was more ambiguous: while it was expected that more reviews would lead to a price premium, the results were negative and significant, suggesting a problem of reverse causality. These studies showed that while the price drivers of Airbnb listings are similar to those of the hotel industry, reputation (reviews, host attributes) also plays a role. More location characteristics have been considered by Deboosere et al. (2019) and Perez-Sanchez et al. (2018). Deboosere et al. (2019) found that access to public transportation and neighbourhood socio-economic factors have significant effects on Airbnb accommodation prices in New York. Perez-Sanchez et al. (2018) showed that listings in sightseeing areas and near the coastline are more expensive in the Mediterranean Spanish cities.

However, the OLS estimation method used in these analyses is not appropriate, as observations that are close to each other in space share latent location-specific attributes (Tang et al., 2019). If the value similarity of observations coincides with location similarity (*spatial autocorrelation*), and it is not treated in the model specification, the OLS estimates will be biased and inconsistent (Anselin & Bera, 1998).

A popular method for addressing spatial autocorrelation in the hospitality literature is the geographically weighted regression (GWR) approach. GWR assumes that the relationship between variables differs between locations, and generates a regression coefficient for each observation (Kim et al., 2018; Zhang et al., 2011). Therefore, GWR is useful for analysing the local variations in the estimated coefficients. GWR has been implemented in hotel price studies to show the significant role of distance to transportation infrastructure (Kim et al., 2018; Soler & Gemar, 2018; Zhang et al., 2011) and to tourist attractions (Kim et al., 2018; Zhang et al., 2011). For the analysis of Airbnb prices in Nashville, Zhang et al. (2017) presented GWR regressions with a limited number of variables, including distance from a convention centre and from the highway. Although the method controls for spatial dependencies, the final model omits various factors, such as listing type.

While GWR highlights local differences between the relationships of variables within the studied area, global spatial models can be used to calculate the average effects for the entire region. Moreover, global spatial models enable us to examine various spatial effects potentially relevant for Airbnb prices, such as the impact of neighbouring listings. Therefore, when the aim is to analyse the determinants of Airbnb prices across multiple cities, global models are more relevant. Prior studies considered two models: the spatial autoregressive model (SAR) and the spatial error model (SEM). Tang et al. (2019) demonstrated that it is necessary to use spatial models in the presence of spatial autocorrelation, and to test the impact of two groups of variables in US cities: namely, listing characteristics and zip code-level location attributes. However, the results provide relatively few insights, as the authors did not pursue a general-to-specific variable selection, but instead estimated two separate models for the two groups of variables. Lawani et al. (2019) estimated a hedonic price model for Boston, and confirmed that distance to the city centre, the business district, and the closest convention centre significantly affect prices; while distance to the train station is not significant.

These studies provide insights into the importance of correct model specifications. First, spatial models provide a better fit for the data than OLS. Second, the magnitudes of the estimated coefficients from OLS are significantly different, especially in the case of location variables. Therefore, spatial models should be used when conducting statistical analysis in which hedonic price models are applied to Airbnb rooms.

Our study is also inspired by research on guest satisfaction with hotel location. Yang et al. (2018) proposed a gravity-type index to measure the attractiveness of hotel location, based on TripAdvisor data. The authors showed that guest satisfaction is significantly affected by distance to the closest metro station, to highly valued attractions, and to airports and universities.

While the literature on Airbnb prices is growing, crucial research gaps can be identified. The most important contributions of our study can be summarised in three points. First, in the previous literature, only simple location variables were tested, such as the distance from the city centre or the train station. On the other hand, using the distance from certain points as an explanatory variable assumes that the relationship is linear in space, which is usually not the case (Lu et al., 2014). Therefore, following the insights from the literature on the hotel industry (Yang et al., 2018), we will use TripAdvisor data to measure the price premium related to an attractive neighbourhood. We will estimate the impact of the distance from the city centre, the distance from the nearest metro station, and the attractiveness of the neighbourhood in the same analysis.

Second, in the presence of spatial autocorrelation, spatial models should be used to estimate the impact of Airbnb price determinants. Existing studies (Lawani et al., 2019; Tang et al., 2019) that recognised the necessity of treating spatial dependence considered only a subset of global spatial models with a single source of spatial process. However, the state-of-the-art in spatial econometrics recommends testing for various other models, including models with multiple sources of spatial dependence (Elhorst, 2010). In this study, we consider previously omitted models, and provide a brief methodological guide for their implementation. Our study presents the main steps of the modelling strategy, including the analysis of spatial autocorrelation, the model and variable selection, and the interpretation of the results.

Third, there are only few studies that have analysed cross-sectional Airbnb data for European cities (Gyódi, 2019; Teubner & Dann, 2017; Wang & Nicolau, 2017). This article constitutes the first attempt to compare price determinants in a large and varied sample of major cities across Europe, thereby revealing a number of differences in Airbnb pricing in different locations.

3. Dataset and methodology

3.1. Dataset

In order to collect Airbnb offers that would be presented to a real user, an automated experiment was conducted based on web-scraping. With the use of a web-automation framework (*Selenium WebDriver*), search queries were executed on the Airbnb platform that referred to accommodations in 10 major European cities for two people and two nights. The offers were collected four to six weeks in advance of the travel dates, and the collected prices refer to the full amount due for the accommodation, including the reservation fee and cleaning fee. For each city, two datasets were prepared, including offers for weekdays (Tuesday-Thursday) and weekends (Friday-Sunday). The analysis is based on the weekday samples, while the weekend data are used for robustness checks.

The search queries included various city districts; e.g., 80 *quartiers* in the case of Paris. Airbnb also returned listings outside the administrative borders of cities that were excluded from the analysis based on shapefiles containing the city areas. To improve the comparability of the results across the cities, the London sample is restricted to Inner London.

A further consideration is the size of the listing. Although all prices in our samples refer to an accommodation for two people, Airbnb presents users with considerably larger listings as well (e.g., a house for 10 people). Following the analysis of histograms with guest capacities, we decided to exclude listings that accommodate more than six people. Further details on the data collection, filtering, and samples are included in Appendix A. The datasets are publicly available at Zenodo repository.[1] The following sections provide a brief summary of the variables (listed in Table 1).

Table 2 shows that median prices are below the average values and the samples are characterised by high standard deviation (SD). These statistics suggest skewed price distributions with a long tail towards higher values. Therefore, the logarithm of price is used as the dependent variable.

Table 1
Listing attributes.

price	log of the final sum
bedrooms	number of bedrooms
person_capacity	maximum number of guests
room_private	dummy for private rooms
room_shared	dummy for shared rooms
cleanliness	guest reviews: scale to 10
guest_satisfaction	guest reviews: scale to 100
superhost	dummy for hosts with the superhost status
multi	dummy for listings offered by hosts with 2–4 listings
biz	dummy for listings offered by hosts with more than 4 listings
dist	distance to the city centre in kilometres
metro_dist	distance to the closest metro station in kilometres
attr_index	attraction index: scale to 100
rest_index	restaurant index: scale to 100

Table 2
Number of listings and descriptive statistics for prices (in EUR).

City	Number of listings	Median	Avg	SD
Amsterdam	1103	430.25	545.02	416.79
Athens	2653	127.72	155.87	366.5
Barcelona	1555	208.53	288.39	321.08
Berlin	1284	187.79	240.22	230.23
Budapest	2074	147.46	168.43	126.03
Lisbon	2857	223.03	236.35	108.44
London	4614	256.36	360.23	507.73
Paris	3130	318.53	398.79	396.3
Rome	4492	179.79	201.62	117.74
Vienna	1738	204.52	240.38	454.4

The descriptive statistics for ordinal explanatory variables (size and guest reviews) are presented in Table 3. The data reveals that certain characteristics of listings are common across the analysed cities, including the high values of guest ratings. The very high average values for cleanliness (above 9/10) and guest satisfaction (above 90/100) support the findings of previous studies on the rating system of Airbnb (Teubner & Dann, 2017). The average values of size variables (*person_capacity* and *bedrooms*) are also similar across the cities. On the other hand, Table 4 shows significant differences in the shares of the various listing types. The vast majority (above 70%) of listings are entire homes in Athens, Budapest, Paris and Vienna, while private rooms constitute more than 50% of listings in Amsterdam, Barcelona, Berlin and London.

To examine the impact of professionalisation on price, three dummy variables are created based on the number of listings offered by the host: *single*: one listing; *multi*: two to four listings and *biz*: more than four listings. The share of single-listings significantly differ in the examined cities: they form the majority in Amsterdam, Berlin and Paris (52–58%), while their share is less than 30% in Barcelona, Lisbon and Rome. The analysis of listing types and professionalisation variables suggest that regulatory differences have a significant impact on the structure of Airbnb networks. The more strictly regulated markets (e.g. Amsterdam, Berlin) are characterised by higher shares of single-listings and private rooms than the cities with a more relaxed approach (e.g. Budapest, Athens). However, the professionalisation of Airbnb is a general trend, as offers that belong to professional hosts constitute more than 30% of listings in 7 analysed cities.

The next variable related to professionalisation is the *superhost* status that is granted to hosts providing high-quality service. Across the examined cities, 14–43% of listings are by hosts with this quality signal.

Besides the standard control variables from the literature, a number of location variables are tested based on Euclidean distances. Following previous studies (Wang & Nicolau (2017); Teubner and Dann (2017); Deboosere et al. (2019)), the baseline measure (*dist*) is the distance of the listing to the city centre. The central points are evaluated for each city individually: i.e., they are located in the historic city centre (e.g.,

[1] Available at http://doi.org/10.5281/zenodo.4446043.

Table 3
Descriptive statistics of ordinal explanatory variables (various units).

	bedrooms (number)		person_capacity (2-6)		guest_satisfaction (0-100)		cleanliness (0-10)	
	Avg	SD	Avg	SD	Avg	SD	Avg	SD
Amsterdam	1.28	0.74	2.79	1.04	94.36	6.09	9.46	0.8
Athens	1.28	0.66	3.71	1.29	95.1	8.27	9.64	0.83
Barcelona	1.22	0.57	2.76	1.28	90.93	8.69	9.29	1.01
Berlin	1.08	0.57	2.8	1.22	94.3	6.91	9.48	0.83
Budapest	1.12	0.67	3.58	1.26	94.53	6.53	9.47	0.85
Lisbon	1.28	0.74	3.37	1.36	91.05	9.15	9.36	0.92
London	1.12	0.58	2.83	1.24	90.32	11.75	9.15	1.18
Paris	0.97	0.64	2.95	1.23	91.85	9	9.25	0.99
Rome	1.23	0.55	3.35	1.31	93.2	7.75	9.52	0.81
Vienna	1.1	0.59	3.33	1.27	93.8	7.09	9.47	0.85

Table 4
Descriptive statistics of dummy explanatory variables (share of listings (%)).

	room type			professionalisation			superhost	
	room shared	room private	home entire	single	multi	biz	yes	no
Amsterdam	0.54	50.68	48.78	57.66	30.83	11.51	29.28	70.72
Athens	0.19	7.58	92.24	34.83	27.37	37.81	43.08	56.92
Barcelona	0.51	76.21	23.28	27.27	37.68	35.05	18.07	81.93
Berlin	2.73	60.36	36.92	54.05	27.57	18.38	26.64	73.36
Budapest	0.34	10.41	89.25	34.81	29.85	35.34	36.89	63.11
Lisbon	1.23	31.29	67.48	17.12	23.98	58.91	21.35	78.65
London	0.5	56.89	42.61	30.78	26.94	42.28	14.69	85.31
Paris	1.5	24.22	74.28	51.69	22.49	25.81	13.71	86.29
Rome	0.16	38.54	61.31	28.41	38.67	32.93	33.64	66.36
Vienna	0.4	21.98	77.62	38.2	28.25	33.54	29.34	70.66

Plaça de Catalunya in Barcelona, Stephansplatz in Vienna) or near busy transportation hubs (e.g., Alexanderplatz in Berlin). The list of locations is included in the Appendix (Table A1). The second tested variable (*metro_dist*) measures the distance from the listing to the closest metro station. The coordinates of the metro stations were scraped from corresponding Wikipedia pages that contain the list of metro stations for each city.

While these variables should explain part of the location price premium, they have a serious shortcoming. When we look at the price of housing, we see that clusters of high-price and low-price neighbourhoods can be found across various city districts, and that these clusters are not necessarily in concentric circles (Brueckner et al., 1999). Therefore, we implement another approach to measuring the price premium related to attractive locations.

Tourists value the proximity to sightseeing venues and points of interest (Yang et al., 2018). When people are planning a leisure trip, they rely heavily on online sources (Pan & Fesenmaier, 2006). Among the most important websites for tourists is TripAdvisor, with around 460 million monthly visitors (TripAdvisor, 2020). The platform aggregates information on all kinds of attractions (e.g., sights, museums, and parks), and allows visitors to rate them. Therefore, TripAdvisor provides insight into not just the location and the number of relevant visitor attractions, but their popularity. While the authenticity of user-generated content on travel platforms is often questioned (Ayeh et al., 2013), empirical studies have found that TripAdvisor contains reliable reviews (Chua & Banerjee, 2013; Díaz & Espino-Rodríguez, 2018).

Following the work of Yang et al. (2018), we measure the accessibility of Airbnb listings with respect to destinations relevant to tourists based on TripAdvisor data. The locations and the numbers of reviews were collected in all cities for two categories: Attractions (Things To Do) and Restaurants. The venues without any reviews and outside the city borders were excluded from the dataset.

The *attr_index* for listing j, based on K points of interest is calculated as:

$$attr\_index_j = \sum_{k=1}^{K} \frac{R_k}{d_{jk}} \tag{1}$$

where R is the number of reviews for attraction k, and d_{jk} is the distance between the listing and point k. The calculated value is divided by the maximum value in a given city and multiplied by 100; therefore, the index has a range of 0–100 in all cities. In addition to *attr_index*, which is based on the venues from the Attractions category, *rest_index* is considered for restaurants.

Relative to the accessibility index proposed by Yang et al. (2018), we make two adjustments. First, we use d_{jk}, instead of d_{jk}^2 in our index. Without this modification, the values of the indices would be close to zero. Second, Yang et al. (2018) used a threshold of 10 reviews to exclude irrelevant attractions. In our analysis, we consider those attractions and restaurants that have more reviews than the first quartile.

Table 5 presents the statistics for the four measures. The data supports the assumption that Airbnb listings are highly concentrated in the vicinity of central points, with average distances of only 2–3 km in most cities. The data also suggests that access to public transportation may be an important attribute, as average distances to the closest metro station are less than 1 km in most cities.

Fig. 1 shows heatmaps for Airbnb prices in the four cities with the largest number of listings in our sample. The heatmaps contain squares with sides of 500 m: the colour of the square indicates the mean value of the listing price. In all cities, listings in central areas form more expensive clusters, while suburban areas tend to be characterised by lower values. This finding supports the hypothesis that location is a crucial price determinant.

Similarly, Fig. 2 presents heatmaps for mean values of *attr_index*. The maps show the concentration of attractions in the city centre: in general, the listings in central areas have a higher accessibility of attractions; while further from the centre, the index values decline. The maps with *rest_index* display similar characteristics (Figure A.1).

In the supplementary videos available online, the 3D heatmaps

Table 5
Descriptive statistics for location variables (various units).

	dist (km)		metro_dist (km)		attr_index (0-100)		rest_index (0-100)	
	Avg	SD	Avg	SD	Avg	SD	Avg	SD
Amsterdam	2.84	2.12	1.09	0.84	14.35	10.43	23.8	16.48
Athens	1.78	0.95	0.48	0.28	5.85	4.81	16.61	11.56
Barcelona	2.12	1.35	0.43	0.28	15.83	9.14	19.32	10.15
Berlin	5.26	3.68	0.84	1.24	16.74	10.67	30.23	16.44
Budapest	1.9	1.91	0.56	0.9	12.69	6.75	31.83	17.56
Lisbon	1.96	1.75	0.7	0.9	7.36	5.1	25.24	15.72
London	5.33	2.72	0.99	1.26	20.61	11.99	11.28	6.98
Paris	3.01	1.47	0.23	0.12	18.12	7.73	47.9	16.69
Rome	3.04	1.64	0.84	0.63	10.32	6.43	25.06	13.59
Vienna	3.14	1.95	0.54	0.52	8.81	6.24	4.24	3.7

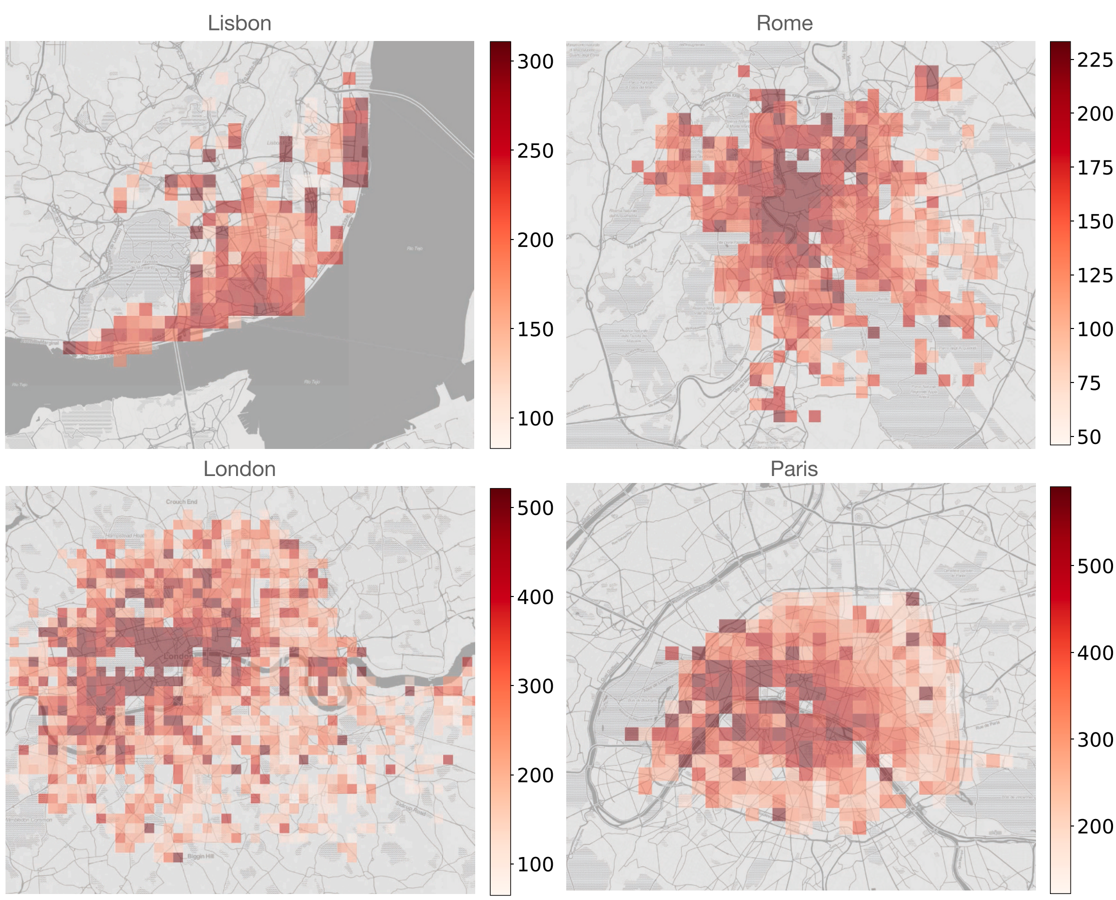


Fig. 1. Mean Airbnb prices (in EUR).

combine price and *attr_index*. The visualisations show that prices are higher in areas dense in tourist attractions, suggesting a positive relationship between the two variables.

Supplementary video related to this article can be found at https://doi.org/10.1016/j.tourman.2021.104319.

3.2. Spatial models

If the observations of the explained variable are affected by the neighbouring observations, we need to include a spatial lag in our model. The spatial lag of the dependent variable (also noted as WY) represents the linear combination of *y* constructed from the observations that neighbour observation *i*: $\sum_{j=1}^{n} W_{ij} y_j$ (LeSage & Pace, 2009, p. 8). *W* is the *spatial weights matrix* that describes the structure of spatial dependence between observations (Halleck Vega & Elhorst, 2015).

The spatial dependence may originate not only from the interaction effects among the dependent variables, but from the interaction effects among the explanatory variables and among the error terms (Halleck Vega & Elhorst, 2015). The Manski model (Manski, 1993), also referred to as the general nesting spatial model (GNS), assumes all possible sources of spatial dependence (Halleck Vega & Elhorst, 2015, p. 6):

$$Y = \rho WY + \alpha \iota_n + X\beta + WX\theta + u, u = \lambda Wu + \varepsilon \tag{2}$$

where:

- Y is the dependent variable with N observations: a N × 1 vector
- W is the N × N spatial weights matrix
- $\alpha\ \iota_n$ is the constant term: a N × 1 vector

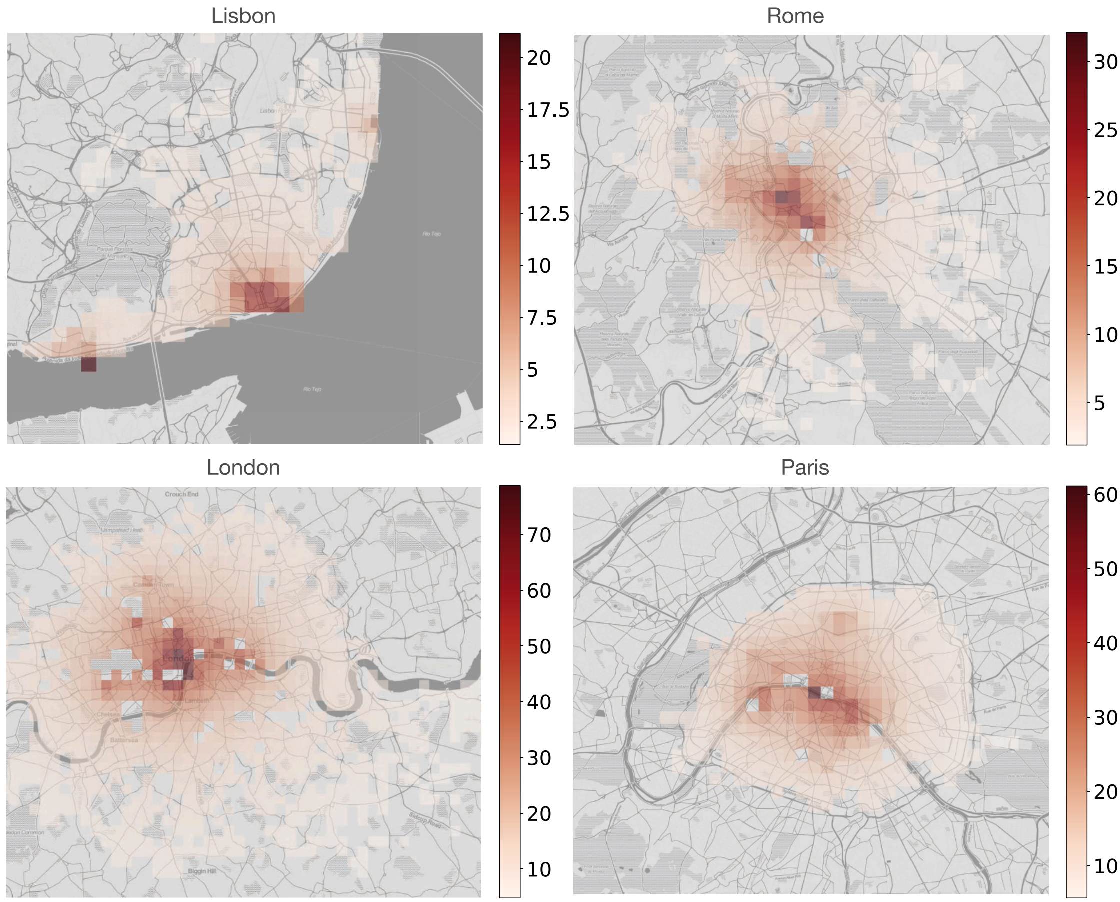


Fig. 2. Attraction index in selected cities.

- X is a N × K matrix for K explanatory variables
- β is a K × 1 vector that measures the influence of the explanatory variables
- θ is a K × 1 vector that measures the influence of the lagged explanatory variables
- ε is a N × 1 vector of disturbance terms
- ρ, λ are scalar parameters that measure the strength of spatial dependence

In general, the GNS model is not preferable due to the difficulties involved in disentangling the various spatial processes (Elhorst, 2010). However, we can formulate various spatial models from GNS by imposing restrictions on the parameters ρ, λ and θ. Based on observations from the functioning of the Airbnb market, various hypotheses can be formulated that help us in our preliminary selection of candidate models.

We can assume that Airbnb hosts set their prices based on the rates of similar listings in their area. As an example, Airbnb provides hosts a data science tool that suggests rates for their listings. Among various control variables, such as the type of the listing, the algorithm takes into account the prices of listings in the area (Hill, 2015). In such a case, we should consider the spatial autoregressive model (SAR) that includes WY:

$$Y = \rho WY + \alpha \iota_n + X\beta + \varepsilon \tag{3}$$

We can also hypothesise that the price of a listing is affected by the characteristics of the neighbouring listings. This can be related to competition; e.g., a host surrounded by upscale, high-rated listings may decide to decrease the price. The spatial lag of X model (SLX) assumes non-zero impact of WX:

$$Y = \alpha \iota_n + X\beta + WX\theta + \varepsilon \tag{4}$$

Finally, we can also assume that both effects take place. In this scenario, a multi-source spatial model should be considered: i.e., the spatial Durbin model (SDM) that includes both WY and WX:

$$Y = \rho WY + \alpha \iota_n + X\beta + WX\theta + \varepsilon \tag{5}$$

Empirical studies put a greater focus on models with a single spatial process, and thus mainly implement the SEM (WU) and SAR (WY) models (Halleck Vega & Elhorst, 2015). However, there are various benefits of assuming fewer restrictions in the model, especially when the exact data-generating process is unknown. LeSage & Pace (2009) pointed out that the SDM provides unbiased estimates for β, even if the true data-generating process is a spatial lag or a spatial error model (Elhorst, 2010).

In contrast to linear regression models, the interpretation of the β coefficients is less straightforward, as the derivative of the dependent variable with respect to the explanatory variable is not equal to β if the model includes WY or WX (Golgher & Voss, 2016; LeSage & Pace, 2009). On the other hand, the inclusion of spatially lagged variables enables the calculation of different metrics on the impact of a variable on the dependent variable. First, the *direct effect* provides the impact of a change from an explanatory variable to the dependent variable in the same location. Second, the spatial spillover, or *indirect effect*, measures the change from an explanatory variable to the dependent variable in other locations (Halleck Vega & Elhorst, 2015). Elhorst (2010) presented the properties of the direct, the indirect, and the total (direct + indirect) effects. Additionally, Golgher and Voss (2016) provided the general formulas for calculations. As our analysis focuses on the determinants of Airbnb prices (e.g., the impact of an improvement in the listing's quality rating on the price of the listing), the direct effects will be calculated and presented in the analysis.

During our analysis, we experimented with various spatial weight matrices. As the Airbnb listings in the analysed cities differ in terms of area and density, we decided to calculate row-standardised W with the 10 closest neighbours. As a robustness check, we also calculated the results for 5, 25, and 50 neighbours; and for distance-based W with distances of 500 and 1000 m.

The matrix of correlation (Table A3) reveals that there is a relatively strong relationship between various pairs of location variables, between the two guest review measures (*cleanliness*, and *guest satisfaction*) and between *person_capacity* and *bedrooms*. To avoid collinearity in the regression analysis, the variance inflation factor (VIF) test is calculated (Table A5). The results suggest that multicollinearity may be an issue for regressions that include both accessibility indices. Hence, the two variables will be tested separately in the analysis.

As the value of location variables is similar among neighbours (e.g., the distance to the closest metro station will be similar among neighbouring listings), X and WX will be highly correlated. For this reason, location variables will be included only among the X variables, and omitted from WX.

4. Results

4.1. Model selection

Our analysis is based on Python programming language and the *PySAL* package (Rey & Anselin, 2007). The scripts prepared for the spatial regressions and robustness checks are published along with the datasets at Zenodo.

First, Moran's I is calculated to test whether Airbnb rates are spatially correlated (Table 6). For all cities, the null hypothesis is rejected, thus Airbnb rates are spatially dependent.

Next, the three spatial models are estimated, along with the baseline OLS. OLS and the SLX models are estimated with ordinary least squares, while SAR and SDM are calculated with the maximum likelihood (ML) method. Table B1 presents the Akaike Information Criterion (AIC) scores and log-likelihood (LL) values for the four considered models (with specifications including *attr_index*). The SDM model (WX, WY) achieves the best results, with the highest log-likelihood values in all cities, and the lowest AIC scores in seven cities. The second-best performance is provided by the SAR model (WY), while SLX (WX) and OLS perform significantly worse. Therefore, we will focus on the results of the SDM model.

In order to verify that the WX, WY model is appropriate, the significance and the magnitude of the lagged variables is examined. The ρ coefficients for WY (Tables B2 and B3) are significant for all cities, with values between 0.18 (Vienna) and 0.54 (Rome). Therefore, WY seems to have a robust influence on the price setting, and should be included in the model specification. The θ coefficients for the spatial lags of the independent variables (WX) are significant for the size (*lag_room_private*, *lag_bedrooms*), and professionalisation (*lag_biz*) attributes in a number of cities. This finding supports our assumption that a part of the price variation is explained by the characteristics of neighbouring observations; e.g., the price of the offer is higher when the nearby offers are by professional hosts, and is lower when the neighbouring Airbnb homes include more bedrooms. For this reason, models including WX should be considered.

Table 6
Spatial autocorrelation tests.

	Moran's I	p-value
Amsterdam	0.21	0.00
Athens	0.34	0.00
Barcelona	0.2	0.00
Berlin	0.19	0.00
Budapest	0.16	0.00
Lisbon	0.26	0.00
London	0.35	0.00
Paris	0.28	0.00
Rome	0.41	0.00
Vienna	0.15	0.00

To conclude, the SDM model is well motivated theoretically (hosts take into account the prices of similar listings in their neighbourhood), and achieved the best goodness of fit.

4.2. Regression results

Figs. 3–5 summarise the results for size, quality, and location attributes. The graphs present the results for the baseline OLS and the three spatial models: the colour of the circle reveals the estimation method, while the transparency shows whether the p-value is below five percent (if the circle is transparent, the regression coefficient is not significant).

As regression coefficients cannot be interpreted as partial derivatives in the case of models with WY, the direct effects are calculated, and are presented in the figures. The estimated coefficients of the WX, WY (SDM) regressions are presented in Tables B2 and B3.

Two different model specifications were estimated that differ only in the TripAdvisor-based accessibility index used: one includes *attr_index*, while the other includes *rest_index*. The results are presented for the regressions with *attr_index*. However, the impact of *rest_index* is shown separately at the end of the section as well (Fig. 6).

All size variables are found to be significant across all models and cities, which supports the assumption that size attributes are major drivers of price. Relative to entire homes, shared rooms and private rooms are significantly cheaper. In the case of Barcelona, where the coefficients are the greatest, shared rooms are, on average, around 66% cheaper (*exp (-1.07)-1*), and private rooms are around 46% cheaper, than entire homes. For each additional person the listing can accommodate, the prices increase between 5.9% (Budapest, Rome) and almost 20% (Amsterdam). Similarly, the number of rooms pushes up prices: for each additional bedroom, the prices increase by 6.6–25.6%. The results vary moderately between the different models, which suggests that the impact of additional space on price is relatively stable across the various city areas.

Fig. 4 summarises the variables controlling for listing quality and host attributes. Out of the two variables for guest reviews, the cleanliness rating seems to be more important, with significant coefficients in most cities. An improvement of one point in a scale of 10 leads to a 2.1–5.2% increase in price. As an example, an improvement from 9 to 10 in cleanliness means that the full price of a two night stay increases by around 6 EUR in Lisbon, 9 EUR in London, 17 EUR in Paris and 4 EUR in Rome. The other important quality signal is the superhost status that is significant and positive in five cities (see Tables B2 and B3). By acquiring this quality signal, hosts can raise prices by 4.9% in London (18 EUR), 7.8% in Paris (31 EUR) and 2.6% in Rome (5 EUR), while the effect is not significant in Lisbon.

The direct effects for the dummy variables that control for the number of listings provide detailed insights. In most cities, listings that belong to hosts with multiple offers are more expensive, with significant effects of *multi* (Berlin), *biz* (Barcelona, Vienna) or both (Athens, Paris, Rome). On the other hand, the dummies are insignificant in Amsterdam and Budapest; and the impact is negative in Lisbon, and London. A negative effect may suggest that the offers in these cities are focused on a lower segment of the market. The results also signal that the most professional hosts charge the highest prices: the price premiums for *biz* are significantly higher than for *multi* in a number of cities, especially in Athens and Paris.

The variation between the coefficients is relatively low for the quality variables, but increases for the variables measuring professionalisation, especially in Paris and Barcelona. Additionally, the spatial lag of *biz* is significant and positive in various cities (Tables B2

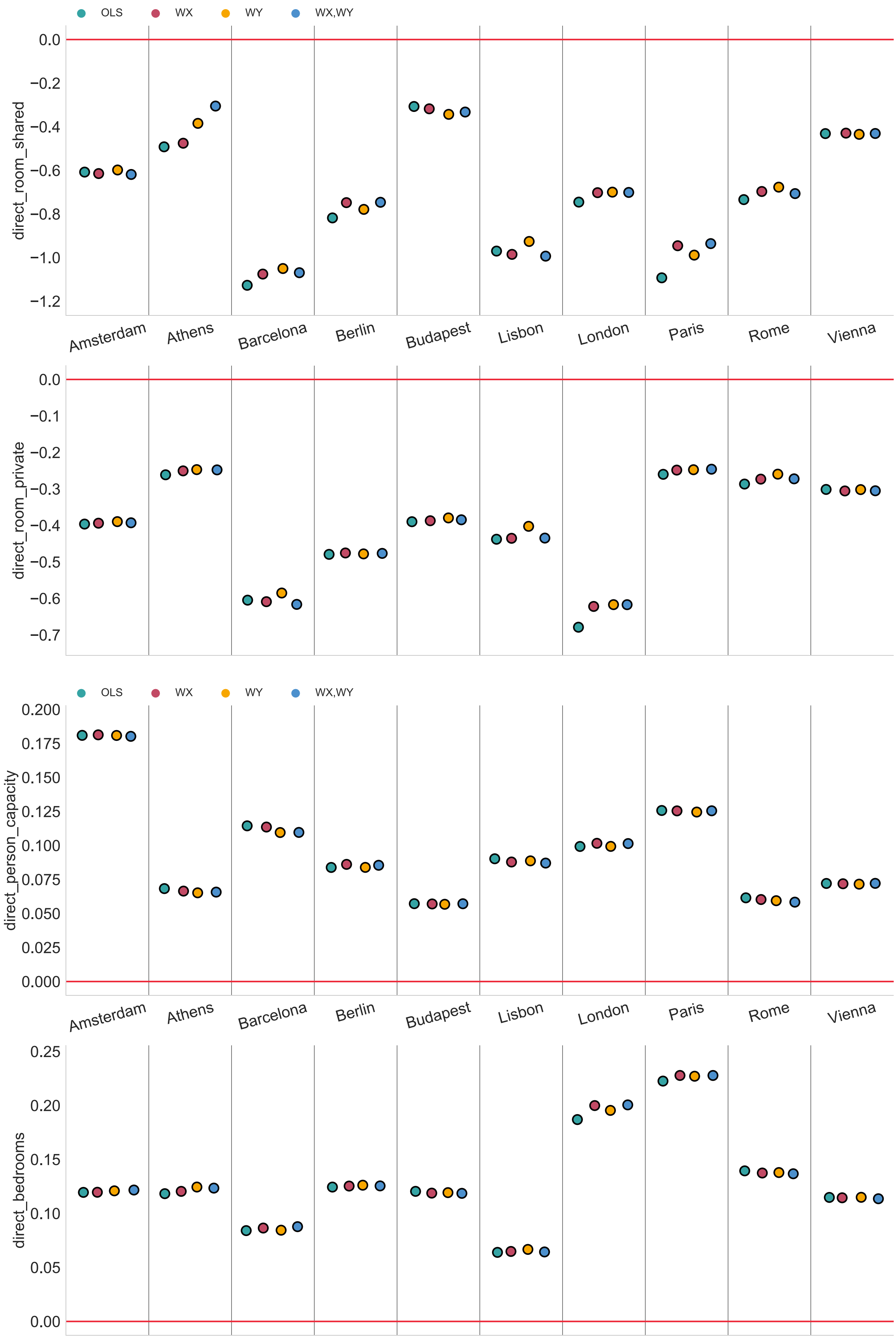


Fig. 3. Regression results 1: size attributes.

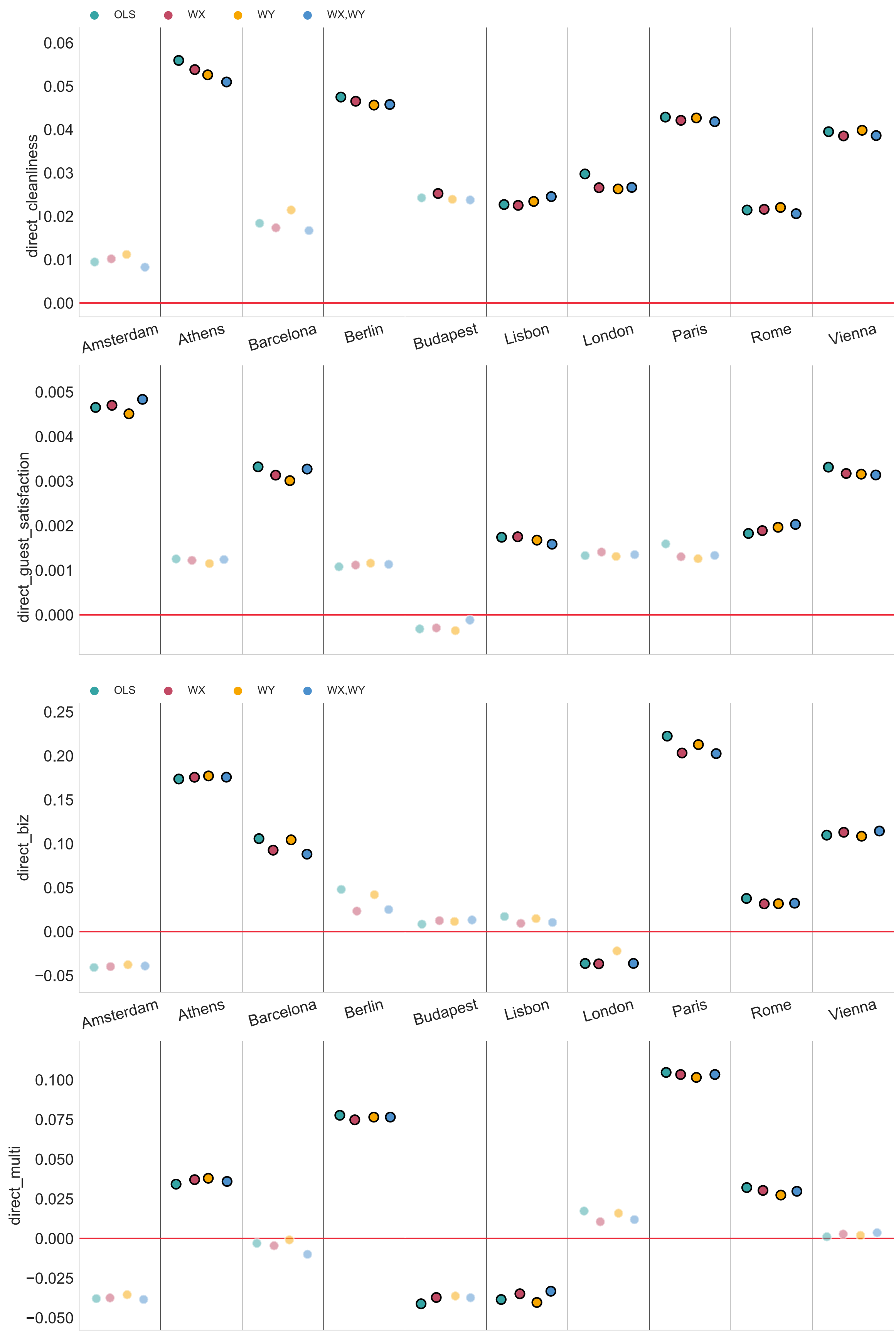


Fig. 4. Regression results 2: Quality and host attributes.

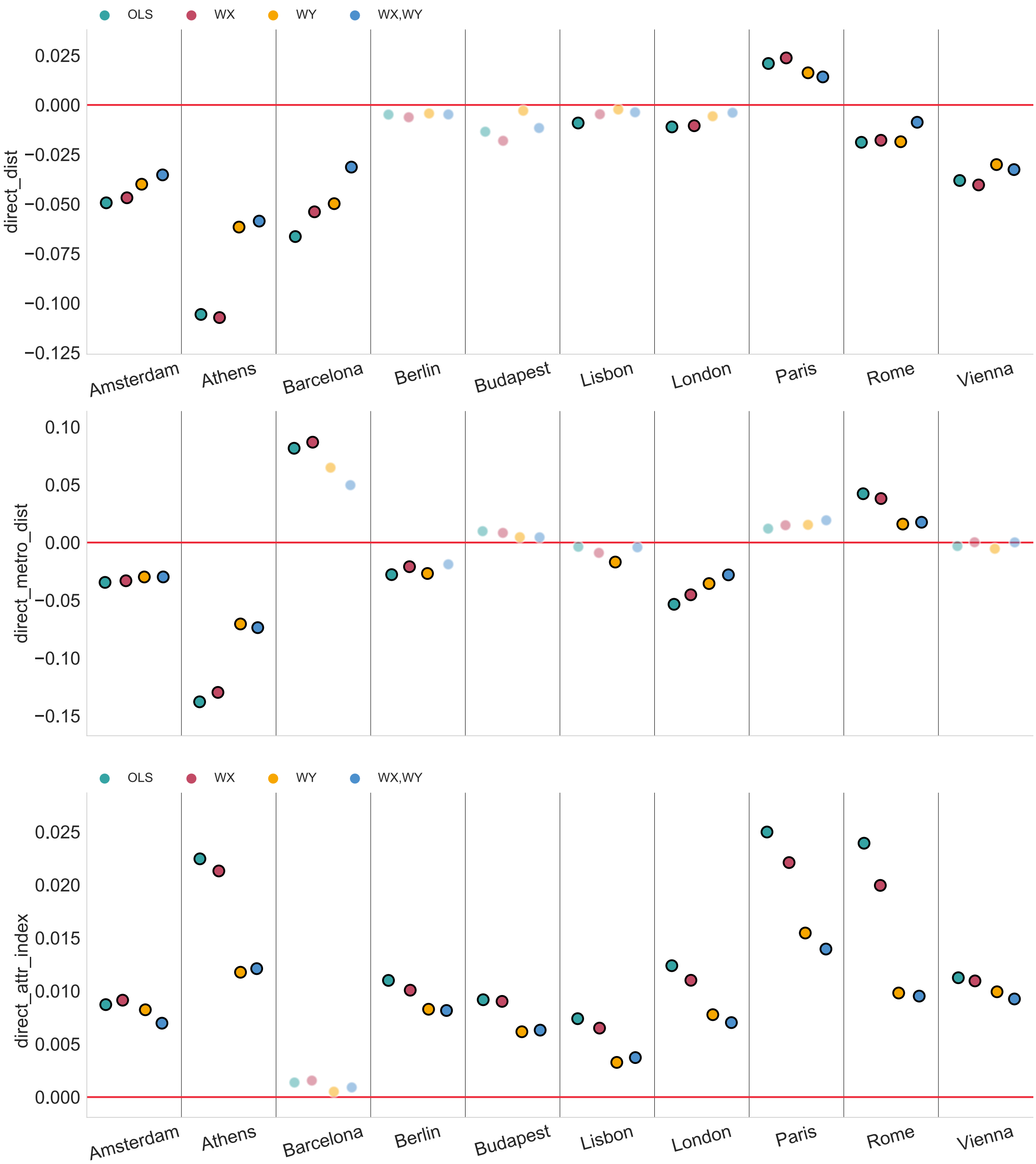


Fig. 5. Regression results 3: Location attributes.

and B3). These findings suggests that professional hosts tend to invest in selected, more expensive neighbourhoods.

Figs. 5 and 6 present the results for the location variables. The method of estimation has a much greater impact on the results and OLS usually yields higher coefficients than the spatial regressions. This is expected, as the treatment of spatial autocorrelation enables us to distinguish the influence of location from the other examined variables, while OLS provides biased results. Additionally, the various spatial models also report coefficients of different magnitudes. The models including a spatial lag of the dependent variable report lower effects than OLS or the WX model. Therefore, models that do not control for the influence of neighbouring observations overestimate the coefficients for the location variables.

The results support the hypothesis that location is a significant factor in determining the price. When we compare the influence of the various variables, we see that *attr_index* or *rest_index* are more robust determinants of price than distance to the city centre or distance to the nearest metro station. *Attr_index* is significant in all cities except Barcelona: relative to the most accessible Airbnb listing (that is, the closest to the most popular attractions), a one-unit improvement in the index (scale to 100) provides a 0.4–1.4% price premium. *Rest_index* is also significant in nine cities, with an influence similar to that of *attr_index*. All things equal, an accommodation with 10 points higher *attr_index* costs more by 9 EUR in Lisbon, 25 EUR in London, 56 EUR in Paris and 19 EUR in Rome.

In comparison, the distance to the city centre is a significant variable in six cities, although in Paris the impact is positive. In the case of metro stations, the coefficients are significant in four cities only, with a positive effect in Rome.

To conclude, the results suggest that measuring the impact of

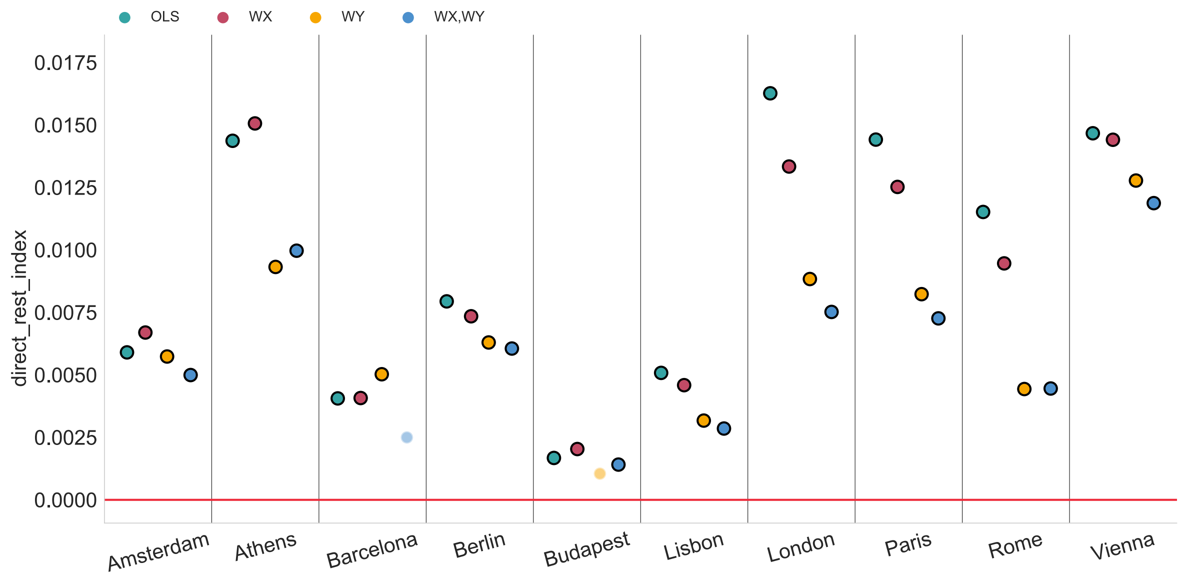


Fig. 6. Regression results 4: Location attributes.

location with distances from single points may be misleading. The choice of the points can be challenging, as in the case of Paris, where regressions suggested that prices are increasing further from the city centre. The central point was assigned to the City Hall (Hôtel de Ville), while various high-price clusters can be found in the western areas of the city (Fig. 1). On the other hand, the TripAdvisor-based indices contain much richer information on attractive tourist locations, and thus provide a more reliable measure of the attractiveness of neighbourhoods. The very large differences between the baseline OLS and the SDM regressions (WX, WY) are also worth noting.

4.3. Robustness checks

To further verify our results, various robustness checks were carried out. Fig. B1 shows the coefficients of selected variables for the weekend and weekday samples. The differences in statistical significance are minor: e.g., there are some changes in the case of *biz* that becomes significant in Berlin and Budapest, and signals a positive impact in London.

As well as verifying our results on the weekend sample, different spatial weight matrices were tested for the SDM regressions. In our study, weights based on 10 closest neighbours are used as the baseline. Choosing a specification that corresponds to the underlying data-generating process is a major challenge in spatial econometrics (Anselin, 2002). For this reason, we test numerous weight types based on k-nearest neighbours, as well as on distance. The results show that the value of ρ differs greatly with alternative specifications: increasing the number of closest neighbours pushes up the strength of the spatial lag, with the smallest coefficient found for five neighbours, and the largest coefficient found for 50 neighbours. On the other hand, the results for distance-based W matrices are inconsistent across cities, often producing lower or higher ρ than the W matrices based on neighbours. Therefore, our baseline W seems to be an appropriate choice, as it results in credible values of ρ: significant and consistent, but not extreme compared to other weight matrices. The results are presented in the online supplementary materials.

5. Conclusions

We have provided a detailed analysis on Airbnb price determinants in 10 major EU cities. The most important contribution of the article is the deeper exploration of the relationship between location and prices, and the accurate estimation of price premiums using spatial models.

The results support the formulated hypotheses

- H1: Airbnb prices are spatially correlated.
- H2: Location attributes significantly affect Airbnb prices.

The analysis also supports the hypotheses on specific location variables: i.e., prices decrease further from the city centre and metro stations, and increase in the vicinity of popular tourist attractions and restaurants. However, among these variables, the TripAdvisor-based accessibility indices are the more robust price determinants.

The results concerning listing type, size and overall guest satisfaction support the findings of previous studies (Gibbs et al., 2018; Perez-Sanchez et al., 2018; Teubner & Dann, 2017; Wang & Nicolau, 2017). Similarly to Lawani et al. (2019), our results also show that the cleanliness rating has a major impact on price. While we have found evidence on the significant and positive impact of the superhost status in half of the analysed cities, the literature reports mixed results: Wang & Nicolau (2017), Gibbs et al. (2018) and Deboosere et al. (2019) found a similar positive effect, while Teubner and Dann (2017) did not report significant coefficients. Our study also provides new insights on the role of professionalisation. The results of prior studies are ambiguous: according to Wang & Nicolau (2017), the price increases with the number of listings owned by the host, while Deboosere et al. (2019) found that rates are lower in the case of hosts with more than 10 listings. In comparison, our analysis suggests that hosts with multiple listings charge higher rates than hosts with one offer, especially hosts with more than four listings.

We also contribute to the literature with a novel analysis on location attributes. Similarly to Teubner and Dann (2017), Lawani et al. (2019) and Zhang et al. (2017), the results support that Airbnb prices decrease with distance from the city centre. Regarding access to public transportation, we report a negative relationship between price and distance to the nearest metro station. However, similarly to the analysis of Deboosere et al. (2019) on New York, we found a positive relationship in Rome.

5.1. Considerations for methodology

The results show that measures based on the distance from certain points (e.g., city centre) are not optimal for measuring the price premium for location. However, the TripAdvisor indices, based on up-to-date data on tourist preferences, provided detailed information on the attractiveness of neighbourhoods. A further advantage of this methodology is related to scalability: while the standard variables require a case-by-case examination of points of interest, web-scraped data on popular tourist destinations enable the implementation of a unified approach.

The study supports the use of spatial regression models with multiple sources of spatial dependence. In the case of basic control variables, such as size attributes, the various models reported similar results. However, the choice of the model had a significant influence on the location variables and attributes with spatial dependence (e.g., number of listings offered by the host). Therefore, researchers conducting hedonic price studies in the context of the hospitality sector should carefully consider their choice of model.

5.2. Managerial and policy implications

The proposed methodology enables a more accurate measurement of price premiums related to various listing attributes. Therefore, our results are highly relevant for both the hospitality sector, as well as for urban planners.

Investors and prospective hosts searching for flats for short-term rental should consider a number of location factors. Instead of focusing on the distance from the city centre, it is more important to evaluate the overall accessibility of popular attractions and restaurants. As the results suggest, even relatively small differences in location can lead to robust price premiums. Therefore, investments in neighbourhoods outside the city centre, but relatively close to various tourist interests can be highly profitable in the long run.

Existing hosts seeking to increase their revenue should focus on improving quality indicators. The results show that the superhost status is associated with a relatively high price premium. The criteria to become a superhost include high guest rating (above 4.8), low cancellation rate and high response rate (Airbnb, 2021). Besides reaching such level of overall guest satisfaction, hosts should also focus on the cleanliness rating. While raising these quality signals may require additional costs (e.g. hiring cleaning service), hosts can also adjust the listing descriptions to better highlight the location characteristics. As an example, a short summary of nearby attractions, restaurants and bars may increase the listing performance. Similarly, hosts should underline convenient access to public transportation and quick commute times to main points of interests.

In case of new listings, potential rates should be assessed based on the prices of similar listings from the neighbourhood. Additionally, hosts can adjust the price in case of a greater variety of popular attractions in the vicinity of the listing.

The analysis also has important insights for policy-makers. The results support the claim that there are strong economic incentives for transforming the housing supply into short-term rentals in attractive neighbourhoods. Moreover, the analysis suggests that profitable locations do not need to be in the city centers. This finding is in line with the literature on Airbnb's contribution to gentrification (Wachsmuth & Weisler, 2018), concluding that Airbnb is increasingly affecting peripheral neighbourhoods. The study also presents the important role of professionalisation: hosts with multiple offers provide the majority of listings in most cities. The results also suggest that professional hosts invest in specific city areas and provide accommodation at higher price points.

Therefore, our findings support the assumption that the uncontrolled growth of Airbnb supply may lead to stronger negative processes, including the gentrification of neighbourhoods. In order to contain these harmful effects, policy-makers may consider regulations that enable occasional home-sharing, while limit professional accommodation provision.

Funding

This research was supported by National Science Centre, Poland: Project number 2017/27/N/HS4/00951.

Declaration of competing interest

Declarations of interest: none.

Acknowledgements

We wish to express our gratitude for the helpful comments received at the NARSC 2019 Conference (Spatial Econometric Association Session), especially to Prof. J. Paul Elhorst, Prof. James LeSage, and Dr Hua Kiefer. We would also like to thank Prof. Dr hab. Katarzyna Śledziewska, Prof. Dr hab. Katarzyna Kopczewska, Dr hab. Piotr Wójcik, Dr Wojciech Hardy, Märtha Andersen, the anonymous reviewers and our colleagues at DELab for their help and comments.

Appendix A. Datasets

Table A.1 Airbnb data collection: search queries (2019).

City	Weekend dates	Weekday dates	Collection of offers	Collection of host info	Number of districts
Amsterdam	23–25.8	20–22.8	13–14.7	25–26.7	97
Athens	23–25.8	20–22.8	14–15.7	27–30.7	76
Barcelona	19–21.7	16–18.7	27–28.5	6–8.7	76
Berlin	26–28.7	23–25.7	3–4.6	13–14.7	94
Budapest	26–28.7	23–25.7	13–16.06	16–18.7	184
Lisbon	26–28.7	23–25.7	7–12.06	9–13.07	141
London	9–11.8	6–8.8	28.6–12.07	21–24.07	534
Paris	19–21.7	16–18.7	29–30.5	8–12.7	80
Rome	30.8–1.9	27–29.8	25–27.7	8–18.8	69
Vienna	26–28.7	23–25.7	4–6.6	14–15.7	107

Table A.2 Data filtering: Weekday and weekend samples.

City	Full Sample	Within city	For max 6 persons	Reviews min 1	With host info	Final
Amsterdam	3096	2531	2497	2146	2118	2080
Athens	7932	7513	6946	5357	5324	5280
Barcelona	4376	3962	3759	3064	2870	2833
Berlin	3461	3307	3160	2693	2516	2484
Budapest	5798	5024	4731	4134	4051	4022
Lisbon	9110	7629	7174	5907	5826	5763
London	27,207	16,560	15,714	11,987	10,241	9993
Paris	10,651	10,130	9716	7425	6781	6689
Rome	13,614	11,352	10,721	9228	9108	9027
Vienna	4638	4541	4263	3690	3567	3537

Table A.3 Central points: variable dist.

City	Central location
Amsterdam	Dam Square
Athens	Syntagma Square
Barcelona	Plaça de Catalunya
Berlin	Alexanderplatz
Budapest	Erzsebet Square
Lisbon	Carmo Convent
London	Trafalgar Square
Paris	Hôtel de Ville
Rome	Roma Termini
Vienna	Stephansplatz

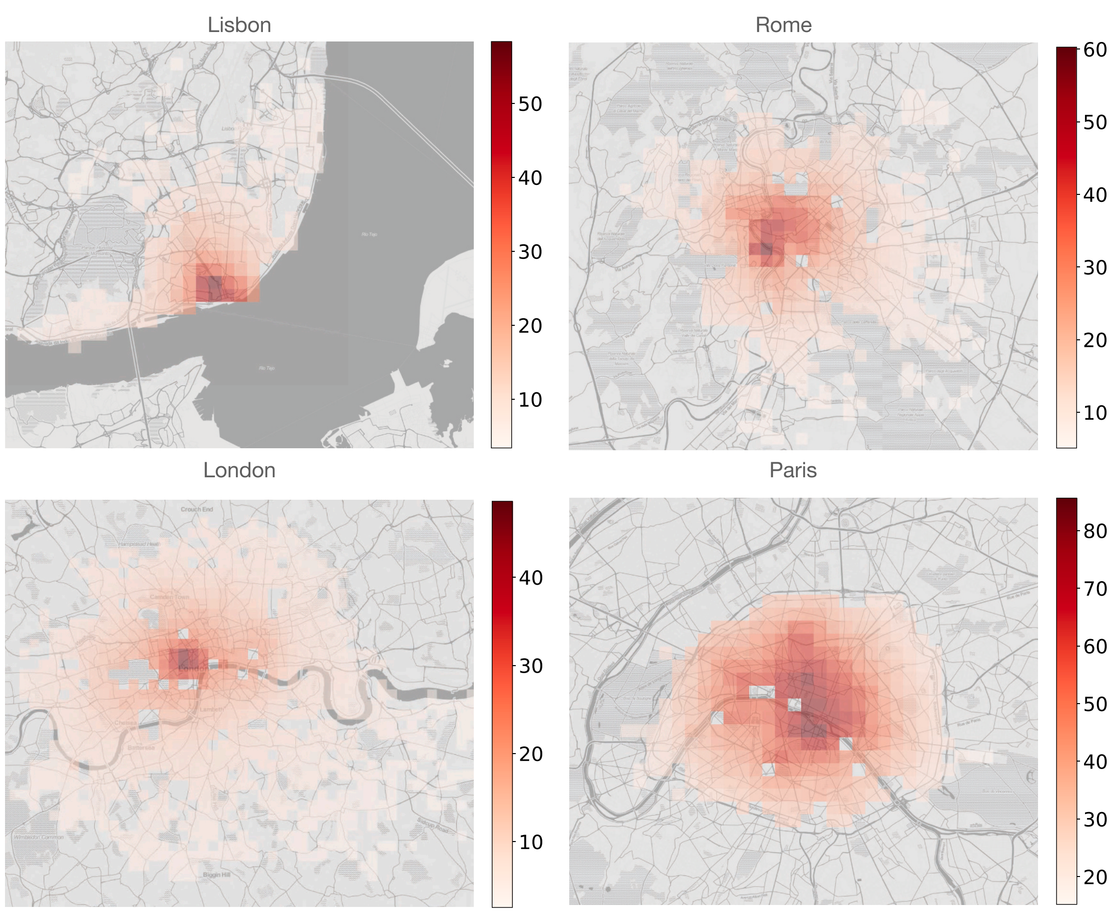


Fig. A.1. Restaurant index in selected cities.

Table A.4 Correlation matrix.

	dist	metro_dist	attr_index	rest_index	person_capacity	cleanliness	guest_satisfaction	bedrooms
dist	1.0***	0.55***	−0.25***	−0.47***	−0.13***	−0.03***	−0.01	−0.01*
metro_dist	0.55***	1.0***	−0.22***	−0.28***	−0.03***	0.01**	0.03***	0.04***
attr_index	−0.25***	−0.22***	1.0***	0.49***	−0.05***	−0.07***	−0.08***	−0.06***
rest_index	−0.47***	−0.28***	0.49***	1.0***	0.05***	0.0	−0.0	−0.07***
person_capacity	−0.13***	−0.03***	−0.05***	0.05***	1.0***	0.0	0.01	0.57***
cleanliness	−0.03***	0.01**	−0.07***	0.0	0.0	1.0***	0.72***	0.03***
guest_satisfaction	−0.01	0.03***	−0.08***	−0.0	0.01	0.72***	1.0***	0.04***
bedrooms	−0.01*	0.04***	−0.06***	−0.07***	0.57***	0.03***	0.04***	1.0***

***: p-val <0.01 **:p-val <0.05 *:p-val <0.1.

Table A.5 VIF test.

	room_shared	room_private	person_ capacity	superhost	multi	biz	cleanliness	guest_satisfaction	bedrooms	dist	metro_dist	attr_index	rest_index
Amsterdam	1.02	1.43	1.9	1.2	1.16	1.15	1.82	1.91	1.9	2.5	1.35	4.08	4.89
Athens	1.01	1.1	1.91	1.11	1.33	1.43	2.32	2.34	1.8	1.99	1.14	2.44	3.44
Barcelona	1.06	2.48	3.35	1.12	1.5	1.83	2.12	2.23	2.34	3.53	1.26	2.06	3.87
Berlin	1.12	1.33	1.64	1.11	1.15	1.42	1.56	1.62	1.25	2.38	1.71	4.36	4.81
Budapest	1.02	1.23	1.59	1.13	1.36	1.42	2.05	2.09	1.45	6.7	4.09	2.07	2.71
Lisbon	1.05	1.75	2.33	1.13	1.89	1.98	2.05	2.09	1.61	2.53	1.36	2.44	3.59
London	1.02	1.56	2.24	1.09	1.41	1.7	2.43	2.54	1.6	4.74	2.14	6.92	5.49
Paris	1.04	1.33	1.75	1.08	1.22	1.37	1.83	1.88	1.54	2.74	1.11	3.58	6.42
Rome	1.01	1.75	2.26	1.16	1.58	1.74	2.05	2.14	1.61	1.65	1.32	2.92	3.24
Vienna	1.02	1.26	1.7	1.11	1.27	1.37	1.69	1.76	1.43	2.21	1.54	4.88	4.65

Appendix B. Analysis

Table B.1 Ranking models based on log-likelihood values and AIC scores.

	OLS		WX		WY		WX, WY	
	AIC	LL	AIC	LL	AIC	LL	AIC	LL
Amsterdam	498.36	−236.18	511.65	−233.83	491.65	−231.82	495.64	−224.82
Athens	1802.75	−888.37	1791.9	−873.95	1603.04	−787.52	1595.67	−774.84
Barcelona	979.93	−476.97	960.04	−458.02	933.7	−452.85	890.32	−422.16
Berlin	847.38	−410.69	840.71	−398.36	835.75	−403.87	831.65	−392.83
Budapest	1420.82	−697.41	1417.82	−686.91	1361.56	−666.78	1364.69	−659.35
Lisbon	498.96	−236.48	453.28	−204.64	339.4	−155.7	260.97	−107.48
London	4035.36	−2004.68	3921.14	−1938.57	3713.02	−1842.51	3661.89	−1807.95
Paris	2270.54	−1122.27	2200.27	−1078.13	2075.86	−1023.93	2052.73	−1003.37
Rome	1739.75	−856.87	1653.88	−804.94	1145.27	−558.63	1078.58	−516.29
Vienna	990.62	−482.31	996.51	−476.25	981.25	−476.63	985.46	−469.73

Table B.2
Results of SDM regression 1: with attr_index, weekdays.

variables	Amsterdam	Athens	Barcelona	Berlin	Budapest
CONSTANT	4.1534*** (0.5679)	1.6586*** (0.2778)	2.6897*** (0.3566)	3.4084*** (0.5287)	2.6673*** (0.4224)
room_private	−0.3918*** (0.0217)	−0.2468*** (0.0251)	−0.6196*** (0.0304)	−0.4774*** (0.0218)	−0.3856*** (0.0271)
room_shared	−0.6083*** (0.1228)	−0.3614** (0.1453)	−1.0420*** (0.1168)	−0.7428*** (0.0669)	−0.3520*** (0.1274)
bedrooms	0.1225*** (0.0167)	0.1246*** (0.0127)	0.0905*** (0.0215)	0.1261*** (0.0180)	0.1184*** (0.0130)
person_capacity	0.1805*** (0.0118)	0.0643*** (0.0067)	0.1078*** (0.0115)	0.0857*** (0.0097)	0.0566*** (0.0073)
guest_satisfaction	0.0049** (0.0020)	0.0011 (0.0012)	0.0030** (0.0014)	0.0011 (0.0017)	−0.0001 (0.0016)
cleanliness	0.0082 (0.0151)	0.0498*** (0.0114)	0.0186 (0.0116)	0.0454*** (0.0137)	0.0228 * (0.0123)
superhost	0.0087 (0.0216)	0.0322** (0.0133)	0.0592*** (0.0224)	0.0004 (0.0219)	−0.0144 (0.0162)
multi	−0.0381 * (0.0209)	0.0376** (0.0161)	−0.0134 (0.0204)	0.0744*** (0.0219)	−0.0346 * (0.0185)
biz	−0.0386 (0.0313)	0.1773*** (0.0155)	0.0822*** (0.0236)	0.0218 (0.0293)	0.0148 (0.0182)
dist	−0.0351*** (0.0074)	−0.0572*** (0.0090)	−0.0309*** (0.0089)	−0.0048 (0.0039)	−0.0115 (0.0094)
metro_dist	−0.0296** (0.0130)	−0.0719*** (0.0243)	0.0489 (0.0333)	−0.0188 * (0.0100)	0.0045 (0.0159)
attr_index	0.0069*** (0.0014)	0.0118*** (0.0018)	0.0009 (0.0012)	0.0081*** (0.0013)	0.0062*** (0.0015)
lag_room_private	0.0606 (0.0634)	0.0943 (0.0667)	0.3182*** (0.0847)	0.1387** (0.0649)	0.1572** (0.0693)
lag_room_shared	−0.2862 (0.3797)	1.2125** (0.5072)	−0.2404 (0.3062)	−0.0651 (0.1298)	0.7289** (0.3491)
lag_bedrooms	−0.0646 (0.0516)	−0.0786** (0.0385)	−0.1007 (0.0663)	−0.0617 (0.0588)	−0.0322 (0.0396)
lag_person_capacity	−0.0550 (0.0377)	−0.0026 (0.0202)	0.0019 (0.0362)	−0.0285 (0.0289)	−0.0004 (0.0210)

(continued on next page)

Table B.2 (*continued*)

variables	Amsterdam	Athens	Barcelona	Berlin	Budapest
lag_guest_satisfaction	−0.0033 (0.0059)	0.0019 (0.0033)	0.0060 (0.0040)	−0.0001 (0.0054)	0.0008 (0.0050)
lag_cleanliness	−0.0009 (0.0455)	−0.0005 (0.0331)	−0.0510 (0.0343)	0.0157 (0.0439)	0.0224 (0.0393)
lag_superhost	0.0678 (0.0682)	−0.0347 (0.0370)	0.0444 (0.0583)	0.0549 (0.0653)	−0.0670 (0.0455)
lag_multi	−0.0060 (0.0637)	−0.0481 (0.0511)	0.0856 (0.0636)	0.1115 (0.0716)	−0.0778 (0.0542)
lag_biz	−0.0087 (0.0808)	−0.1097*** (0.0420)	0.1077 * (0.0590)	0.1907*** (0.0673)	−0.0521 (0.0489)
rho	0.2510*** (0.0581)	0.4580*** (0.0305)	0.3886*** (0.0434)	0.1935*** (0.0568)	0.3201*** (0.0407)
AIC	495.6408	1595.6739	890.3161	831.6514	1364.6903
Loglik	−224.8204	−774.8370	−422.1581	−392.8257	−659.3451

***: p-val <0.01 **:p-val <0.05 *:p-val <0.1.

Table B.3
Results of SDM regression 2: with with attr_index, weekdays.

variables	Lisbon	London	Paris	Rome	Vienna
CONSTANT	2.4863*** (0.2162)	2.8566*** (0.1878)	2.0977*** (0.2337)	2.0489*** (0.1953)	3.1185*** (0.3778)
room_private	−0.4362*** (0.0143)	−0.6083*** (0.0142)	−0.2425*** (0.0165)	−0.2688*** (0.0112)	−0.3063*** (0.0207)
room_shared	−0.9887*** (0.0457)	−0.6867*** (0.0750)	−0.9165*** (0.0555)	−0.7036*** (0.1031)	−0.4284*** (0.1217)
bedrooms	0.0654*** (0.0081)	0.1997*** (0.0117)	0.2287*** (0.0115)	0.1333*** (0.0093)	0.1136*** (0.0154)
person_capacity	0.0846*** (0.0052)	0.1026*** (0.0064)	0.1249*** (0.0064)	0.0577*** (0.0046)	0.0720*** (0.0078)
guest_satisfaction	0.0016** (0.0007)	0.0012 * (0.0007)	0.0011 (0.0009)	0.0022*** (0.0008)	0.0030** (0.0014)
cleanliness	0.0244*** (0.0073)	0.0272*** (0.0069)	0.0418*** (0.0081)	0.0201*** (0.0071)	0.0388*** (0.0116)
superhost	0.0190 (0.0122)	0.0489*** (0.0157)	0.0733*** (0.0180)	0.0254*** (0.0092)	−0.0146 (0.0179)
multi	−0.0347** (0.0151)	0.0084 (0.0141)	0.1010*** (0.0158)	0.0247** (0.0103)	0.0045 (0.0191)
biz	0.0055 (0.0133)	−0.0386*** (0.0148)	0.1967*** (0.0164)	0.0260** (0.0113)	0.1149*** (0.0197)
dist	−0.0036 (0.0039)	−0.0038 (0.0042)	0.0139*** (0.0051)	−0.0084*** (0.0032)	−0.0325*** (0.0065)
metro_dist	−0.0041 (0.0064)	−0.0274*** (0.0064)	0.0190 (0.0502)	0.0169** (0.0071)	0.0001 (0.0187)
attr_index	0.0036*** (0.0012)	0.0069*** (0.0009)	0.0137*** (0.0012)	0.0092*** (0.0010)	0.0092*** (0.0016)
lag_room_private	0.2233*** (0.0344)	0.0684 * (0.0355)	0.0094 (0.0444)	0.0954*** (0.0274)	0.1336** (0.0617)
lag_room_shared	0.3424*** (0.1225)	−0.0205 (0.2298)	−0.1204 (0.1111)	0.3403 (0.2215)	−0.0778 (0.3242)
lag_bedrooms	−0.0491** (0.0215)	−0.0654** (0.0282)	−0.1089*** (0.0336)	−0.0233 (0.0274)	−0.0200 (0.0466)
lag_person_capacity	0.0122 (0.0157)	−0.0678*** (0.0159)	−0.0319 * (0.0190)	−0.0226 * (0.0136)	−0.0023 (0.0223)
lag_guest_satisfaction	−0.0016 (0.0023)	0.0019 (0.0020)	0.0049 * (0.0028)	−0.0039 * (0.0022)	0.0051 (0.0044)
lag_cleanliness	−0.0071 (0.0223)	−0.0222 (0.0200)	−0.0164 (0.0254)	−0.0044 (0.0205)	−0.0168 (0.0352)
lag_superhost	0.0058 (0.0317)	−0.0328 (0.0397)	0.0170 (0.0506)	−0.0131 (0.0248)	0.0490 (0.0476)
lag_multi	0.0431 (0.0425)	0.0719 * (0.0397)	0.0189 (0.0462)	0.0583 * (0.0307)	−0.0573 (0.0557)
lag_biz	0.0965** (0.0398)	0.0702** (0.0332)	0.0677 * (0.0398)	0.0761** (0.0308)	−0.0486 (0.0462)
rho	0.4400*** (0.0299)	0.4211*** (0.0241)	0.3815*** (0.0302)	0.5425*** (0.0205)	0.1797*** (0.0496)
AIC	260.9676	3661.8908	2052.7305	1078.5817	985.4567
Loglik	−107.4838	−1807.9454	−1003.3652	−516.2909	−469.7284

***: p-val <0.01 **:p-val <0.05 *:p-val <0.1.

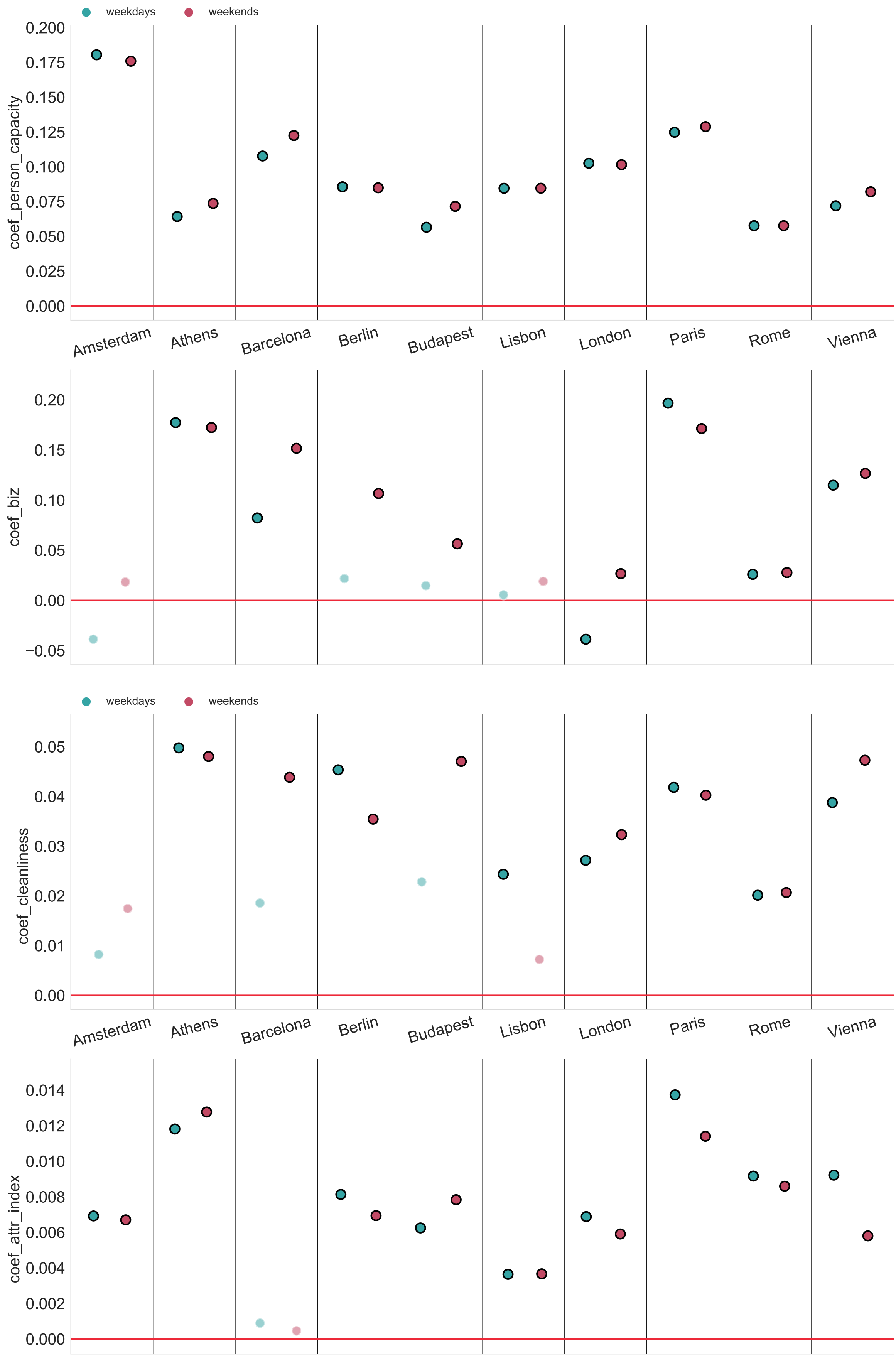


Fig. B.1. Weekend and weekday samples: coefficients of selected variables.

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tourman.2021.104319.

References

Airbnb. (2019). *Fast facts*. Available online at: https://news.airbnb.com/fast-facts/. (Accessed 30 October 2019).

Airbnb. (2021). *Superhost: Recognizing the best in hospitality*. Available online at: https://www.airbnb.com/superhost. (Accessed 4 January 2021).

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics, 27*, 247–267.

Anselin, L., & Bera, A. K. (1998). *Spatial dependence in linear regression models with an introduction to spatial econometrics*. Handbook of Applied Economic Statistics.

Ayeh, J. K., Au, N., & Law, R. (2013). Do we believe in TripAdvisor?" Examining credibility perceptions and online travelers' attitude toward using user-generated content. *Journal of Travel Research, 52*, 437–452. https://doi.org/10.1177/0047287512475217

Brueckner, J. K., Thisse, J. F., & Zenou, Y. (1999). Why is central Paris rich and downtown detroit poor? An amenity-based theory. *European Economic Review*. https://doi.org/10.1016/S0014-2921(98)00019-1

Bull, A. O. (1994). Pricing a motel's location. *International Journal of Contemporary Hospitality Management*. https://doi.org/10.1108/09596119410070422

Carvell, S. A., & Herrin, W. E. (1990). Pricing in the hospitality industry: An implicit markets approach. *Hospitality Review, 8*(2).

Chua, A. Y., & Banerjee, S. (2013). Reliability of reviews on the internet: The case of Tripadvisor. *Lecture Notes in Engineering and Computer Science, 1*, 453–457.

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

Díaz, M. R., & Espino-Rodríguez, T. F. (2018). Determining the reliability and validity of online reputation databases for lodging: Booking.com, TripAdvisor, and HolidayCheck. *Journal of Vacation Marketing, 24*, 261–274. https://doi.org/10.1177/1356766717706103

Elhorst, J. P. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*. https://doi.org/10.1080/17421770903541772

Espinet, J. M., Saez, M., Coenders, G., & Fluvia, M. (2003). Effect on prices of the attributes of holiday hotels: A hedonic prices approach. *Tourism Economics*. https://doi.org/10.5367/000000003101298330

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018). Pricing in the sharing economy: A hedonic pricing model applied to airbnb listings. *Journal of Travel & Tourism Marketing, 35*, 46–56. https://doi.org/10.1080/10548408.2017.1308292

Golgher, A. B., & Voss, P. R. (2016). *How to interpret the coefficients of spatial models: Spillovers, direct and indirect effects*. Spatial Demography. https://doi.org/10.1007/s40980-015-0016-y

Guttentag, D. (2015). Airbnb: Disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism, 18*, 1192–1217. https://doi.org/10.1080/13683500.2013.827159, 10.1080/13683500.2013.827159.

Gyódi, K. (2019). Airbnb in European cities: Business as usual or true sharing economy? *Journal of Cleaner Production, 221*, 536–551. https://doi.org/10.1016/j.jclepro.2019.02.221

Halleck Vega, S., & Elhorst, J. P. (2015). The slx model. *Journal of Regional Science, 55*, 339–363.

Hill, D. (2015). *Uber and the zero marginal cost revolution*. Available online at: http://www.huffingtonpost.com/jeremy-rifkin/uber-german-court_b_5758422.html. (Accessed 30 November 2017).

Kim, J., Jang, S., Kang, S., & Kim, S. H. J. (2018). Why are hotel room prices different? Exploring spatially varying relationships between room price and hotel attributes. *Journal of Business Research*, 1–12. https://doi.org/10.1016/j.jbusres.2018.09.006

Kwok, L., & Xie, K. L. (2018). Pricing strategies on Airbnb: Are multi-unit hosts revenue pros? *International Journal of Hospitality Management*, 1–8. https://doi.org/10.1016/j.ijhm.2018.09.013

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy, 74*(2), 10.1086/259131.

Lawani, A., Reed, M. R., Mark, T., & 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

Lee, S. K., & Jang, S. C. (2011). Room rates of U.S. airport hotels: Examining the dual effects of proximities. *Journal of Travel Research*. https://doi.org/10.1177/0047287510362778

LeSage, J., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton: Taylor & Francis.

Lu, B., Charlton, M., Harris, P., & Fotheringham, A. S. (2014). Geographically weighted regression with a non-euclidean distance metric: A case study using hedonic house price data. *International Journal of Geographical Information Science*. https://doi.org/10.1080/13658816.2013.865739

Lutz, C., & Newlands, G. (2018). Consumer segmentation within the sharing economy: The case of Airbnb. *Journal of Business Research, 88*, 187–196. https://doi.org/10.1016/j.jbusres.2018.03.019

Manski, C. F. (1993). Identification of endogenous social effects the reflection problem. *The Review of Economic Studies, 60*(3), 531–542, 10.2307/2298123.

Oskam, J., van der Rest, J. P., & Telkamp, B. (2018). What's mine is yours-but at what price? Dynamic pricing behavior as an indicator of Airbnb host professionalization. *Journal of Revenue and Pricing Management, 17*, 311–328. https://doi.org/10.1057/s41272-018-00157-3

Pan, B., & Fesenmaier, D. R. (2006). Online information search. Vacation planning process. *Annals of Tourism Research*. https://doi.org/10.1016/j.annals.2006.03.006

Perez-Sanchez, V. R., Serrano-Estrada, L., Marti, P., & Mora-Garcia, R. T. (2018). The what, where, and why of airbnb price determinants. *Sustainability, 10*. https://doi.org/10.3390/su10124596

Rey, Sergio J., & Anselin, Luc (2007). PySAL: A python library of spatial analytical methods. *The Review of Regional Studies, 37*(1), 5–27.

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

Soler, I. P., & Gemar, G. (2018). Hedonic price models with geographically weighted regression: An application to hospitality. *Journal of Destination Marketing and Management, 9*, 126–137. https://doi.org/10.1016/j.jdmm.2017.12.001

Tang, L. R., Kim, J., & 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

Teubner, T., & Dann, D. (2017). Price determinants on airbnb: How reputation pays off in the sharing economy. *Journal of Self-Governance and Management Economics, 5*, 53. https://doi.org/10.22381/jsme5420173

Thrane, C. (2007). Examining the determinants of room rates for hotels in capital cities: The Oslo experience. *Journal of Revenue and Pricing Management*. https://doi.org/10.1057/palgrave.rpm.5160055

TripAdvisor. (2020). *About tripadvisor*. Available online at: https://tripadvisor.mediaroom.com/us-about-us. (Accessed 19 January 2020).

Wachsmuth, D., & Weisler, A. (2018). Airbnb and the rent gap: Gentrification through the sharing economy. *Environment and Planning A*. https://doi.org/10.1177/0308518X18778038

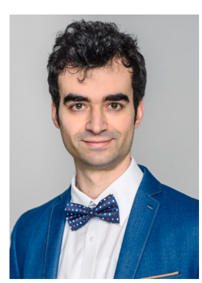
Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management, 62*, 120–131. https://doi.org/10.1016/j.ijhm.2016.12.007

White, P. J., & Mulligan, G. F. (2002). Hedonic estimates of lodging rates in the four corners region. *The Professional Geographer*. https://doi.org/10.1111/0033-0124.00348

Yang, Y., Mao, Z., & Tang, J. (2018). Understanding guest satisfaction with urban hotel location. *Journal of Travel Research, 57*, 243–259. https://doi.org/10.1177/0047287517691153

Zhang, Z., Chen, R. J., Han, L. D., & Yang, L. (2017). Key factors affecting the price of airbnb listings: A geographically weighted approach. *Sustainability, 9*, 1–13. https://doi.org/10.3390/su9091635

Zhang, H., Zhang, J., Lu, S., Cheng, S., & Zhang, J. (2011). Modeling hotel room price with geographically weighted regression. *International Journal of Hospitality Management*. https://doi.org/10.1016/j.ijhm.2011.03.010



Kristóf is a PhD candidate at the Faculty of Economic Sciences at the University of Warsaw and researcher at DELab UW. His research areas include the empirical analysis of Airbnb, online platforms, the sharing economy and the use of data-science methods in economics. He is a member of the Next Generation Internet research team aiming at developing a cutting-edge data-driven methodology for identifying early signals of new trends and technologies.

Łukasz is a PhD candidate in the Doctoral School of Social Sciences at the University of Warsaw, where he carries out research on digital transformation of transport: particularly micro-mobility and its impact on cities. He has completed Master's degree in Data Science with a thesis about benchmarking hyperparameter optimization algorithms in gradient boosting models and a Bachelor's in International Economics, both at the University of Warsaw's Faculty of Economic Sciences. At DELab UW, as a data scientist, he is responsible for programming, web applications, data analysis and processing.