Determinants of Airbnb prices in Bologna: A spatial econometrics approach

Damiano Orlandi

1 Introduction

Nowadays tourism differs significantly from the past. Globalization, technological innovation, and individualism reshaped its characteristics and have transformed our way of intend travelling. A central driver of this change is the spread of short-term rentals: while the idea of renting private accommodation for leisure or business is not new, the scale and form it has recently taken, enabled by digital platforms and mobile apps, are unprecedented [Gut13] At the heart of this transformation lies Airbnb, now the world's leading home sharing platform, whose rapid expansion has made it a crucial actor in debates on overtourism.

This study focuses on Bologna, a medium-sized Italian city, which, like many other attractive destinations in the country, is currently facing severe housing challenges. Historically tied to its university, Bologna must accommodate thousands of students each year. Yet in recent years, a shortage of affordable housing has led student groups to call for new policies to regulate rising rents. Among their main demands is the containment of short-term rental services, which are seen as reducing the availability of long-term housing for students.

In this context, it becomes essential to investigate the factors shaping Airbnb prices in Bologna. Prior research highlights location as a key determinant, with evidence showing that tourists particularly value proximity to the city center and major attractions [Yan+17]. However, traditional OLS models often fail to capture the spatial dependence among such variables, leading to biased estimates. Spatial econometric models, by accounting for unobserved location specific effects across neighboring areas, offer a more robust approach to understanding the dynamics of short-term rental pricing([Wan+17], [Gyó+21]).

The first research question addresses the role of spatial proximity to the city center of Bologna. Specifically, it asks how the distance from the central areas (Piazza Maggiore) of the city influences the price per night of Airbnb offer. Since location is widely recognized as a key determinant of price and demand in the hospitality sector, it becomes crucial to examine whether listings situated closer to the heart of Bologna, enjoy a measurable advantage over those located in more peripheral zones.

The second research question focuses on the role of professional hosts in shaping the market. It explores whether hosts managing multiple rooms, often regarded as more business-oriented operators, could determine higher prices of rooms.

2 Dataset

The dataset used in this analysis was downloaded from the official Airbnb portal https://insideairbnb.com/get-the-data/ for the time window 12/09/2024 – 20/12/2024. In particular, from the four retrieved sources: calendar.csv.gz, listing.csv.gz, reviews.csv.gz, neighbourhoods.geojson, a new clean dataset was constructed by selecting and transforming relevant variables. Table 1 summarizes its structure.

Table 1: Dataset

lat	long	room	price	reviews	host_class	distance
44.491630	11.333980	Private room	3.73	2.63	multi	0.76
44.488170	11.341240	Private room	3.91	0.00	multi	0.65
44.478880	11.356250	Entire home/apt	4.78	1.09	single	1.97
44.487870	11.352030	Entire home/apt	4.94	1.94	multi	1.98

It is important to note that the variable `price` represents the logarithm of the median value of each room in the 4 months period (in Eur), while `number_of_reviews` corresponds to the logarithm+1 of the number of reviews. These transformations were applied because the official Airbnb documentation does not clearly describe how such values are obtained, and in the case of reviews, data were incomplete for the four-month study period. Furthermore, the use of log-transformed values mitigates skewed distributions and provided more robust insights during the preliminary analysis (Table 2 shows non-log data to prove skewed distributions). Other variables were considered for the reviews (like review rankings), but no clear data collection and analysis were provided by Airbnb.

The variable `distance` measures the linear distance (in kilometers) between each Airbnb listing and Bologna's city center, identified as *Piazza Maggiore*. To calculate this distance, the dataset of listings and the rerence point were projected into a metric coordinate system (EPSG:32632 - WGS 84 / UTM zone 32N), ensuring accurate

Table 2: Descriptive statistics of raw data

Variable	Min	Q1	Median	Q3	Mean	Std
Price	10.0	70.0	99.0	150.0	211.51	620.24
Reviews	0.0	0.0	1.0	6.0	3.87	5.52

measurements in meters. Euclidean distances were then computed and converted into kilometers.

Finally, following the methodology proposed in previous literature [Gyó+21], host professionality classes were added to the dataset by defining three thresholds based on the number of listings managed by each host:

- **single**: 1 listing
- **multi**: 2–4 listings
- **business**: more than 4 listings

Table 3: Descriptive statistics of explanatory variables (share of listings, %).

Room type				Professionalisation		
Shared	**Private**	**Entire home/apartment**	**Hotel**	**Single**	**Multi**	**Biz**
0.13	21.64	78.11	0.13	37.28	30.08	32.65

3 Data Analysis

Prior to model estimation, it was neede to assess whether Airbnb prices display spatial structure. Using a row-standardised spatial weights matrix W based on geographic proximity between listings, a global Moran's I on log(price) was performed. The proximity criterion k-nearest neighbours was set equal 9 (the best value not showing disconnected components) and the same W is used consistently across tests and models. The statistic equals $I = 0.17$, which indicates a positive spatial autocorrelation and that nearby listings tend to have more similar prices than distant ones. Thus, the null hypotesis of spatial randomness was rejected and the spatial econometric models were applied.

As a benchmark, an OLS model was estimated:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is log(price) and $\mathbf{X}$ contains distance to the centre, room-type indicators, host-class dummies, and number of reviews. Despite it is statistically significant (F = 77.11, p 0.001) it does not explain the variance ($R^2 = 0.099$). Conclusion confirmed by the Moran's I statistic calculated on the its residuals (0.149) as quality control.

Then, a Lagrange multiplier test has been conducted in order to verify the presence of spatial dependence. The simple LM–lag is significant, but its robust counterpart is not ($p = 0.719$), whereas both LM–error and robust LM–error are strongly significant. This pattern indicates that spatial dependence primarily resides in the error (unobserved, spatially structured factors) not in an endogenous spatial lag of the dependent variable. Therefore, given the rejection of homoskedasticity, a robust SEM estimator

(GM_Error_Het) seems appropriate to verify that residual spatial autocorrelation is eliminated after estimation.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \qquad \mathbf{u} = \lambda W\mathbf{u} + \boldsymbol{\xi}, \qquad \boldsymbol{\xi} \sim \mathcal{N}\big(\mathbf{0}, \sigma^2\mathbf{I}\big)\,.$$

Table 4 shows robust dependencies between variables and that the spatial structure has been captured by . Thanks to the last Moran's I test conducted on the filtered residual of SEM, is possible to say that there is no autocorrelation (Moran I : 0.0054 p: 0.39).

Table 4: SEM (GM_Error_Het) results — Dependent variable: log(price)

Variable	**Coefficient**	**Std. Error**	**z-Statistic**	**p-Value**
Intercept	4.94905	0.03580	138.22391	0.00000
Room: Hotel (vs Entire home)	-0.00648	0.31876	-0.02034	0.98377
Room: Private (vs Entire)	-0.38736	0.02632	-14.71706	0.00000
Room: Shared (vs Entire)	-1.27823	0.15655	-8.16514	0.00000
Host class: Business (vs Single)	0.17373	0.03075	5.64912	0.00000
Host class: Multi (vs Single)	0.09274	0.02389	3.88143	0.00010
Distance	-0.08265	0.01263	-6.54478	0.00000
Reviews	-0.08540	0.01099	-7.77273	0.00000
Spatial error (λ)	0.36181	0.02610	13.86444	0.00000

4 Conclusion

Thanks to the conducted research and as [Gyó+21] discovered, there is a clear, and free of neighbouring spatial biases, dependence among distance from the city centre and Airbnb accomodation prices. The results confirm the working hypothesis: prices decrease as listings are located further away from central areas, while proximity to the city centre is rewarded with higher nightly rates. The estimated impact per kilometer underscores how location has become a decisive factor in the steep increase of prices observed in recent years.

Another important conclusion concerns the role of host professionalisation. The results show that business and multi hosts tend to charge higher prices compared to single hosts, supporting the idea that for many operators Airbnb represents a professional activity rather than a casual way of renting out family property. This aligns with the second hypothesis and confirms the presence of another key factor of short-term rentals.

Hotel rooms were excluded from the analysis due to inconsistent data, both in terms of sample size and statistical significance of results (4). By contrast, private and shared rooms emerged as negative price determinants when compared to entire homes or apartments, suggesting that tourists place a significant premium on privacy and exclusivity.

A further noteworthy result is the effect of the number of reviews: listings with more reviews are generally associated with lower prices. This could indicate that hosts who have already secured a large number of bookings are more inclined to adjust prices downward to maintain occupancy, while those with fewer reviews keep prices higher in order to compensate for lower and less predictable income streams.

Overall, these findings are consistent with those observed across 33 major European cities by [Wan+17], strengthening the external validity of the analysis. They also resonate with the wider debate on the gentrification effects linked to short-term rentals. In Bologna, as in other cities, the attractiveness of central areas for short-term tourists reduces the availability of affordable long-term rentals, thereby exacerbating tensions in the housing market. Living close to amenities and services is often more profitable for short-term rental purposes than for long-term leases, further fuelling the displacement of local residents.

5 Limitations

While the analysis provides meaningful insights into the determinants of Airbnb prices in Bologna, several limitations must be acknowledged. First, in order to simulate more faithfully the methodology adopted in the quoted research, additional data sources would have been necessary. In particular, the use of a TripAdvisor scraper could have enriched the dataset with more detailed indicators of tourist demand and attraction density, thereby offering a more precise measure of location-specific amenities.

Second, the set of explanatory variables employed in this study remains somewhat limited. Future work should consider integrating a broader range of factors that are likely to influence short-term rental prices. For instance, the inclusion of information on nearby amenities (restaurants, cultural attractions, commercial areas), the availability and quality of local transport connections, and a more nuanced measure of neighbourhood characteristics could substantially improve the explanatory power of the model.

Third, although the review variable was incorporated, it was restricted to a log-transformed count of reviews. A richer operationalisation of this factor could involve not only the number of reviews, but also their qualitative dimension. For example, a classification of reviews into positive and negative categories, or the use of sentiment analysis, would allow for a better assessment of how guest satisfaction translates into pricing strategies.

Finally, the temporal scope of the dataset is relatively narrow. Short-term fluctuations in demand or seasonal effects could not be fully captured, limiting the generalisability of the results. Extending the analysis to a longer time frame would help verify the robustness of the observed relationships and account for potential shocks affecting the housing and tourism markets.

References

[Gut13] D. Guttentag. “Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector”. In: *Current Issues in Tourism* 18.12 (2013), pp. 1192–1217. DOI: 10.1080/13683500.2013.827159. URL: https://doi.org/10.1080/13683500.2013.827159.

[Gyó+21] Kristóf Gyódi and Łukasz Nawaro. “Determinants of Airbnb prices in European cities: A spatial econometrics approach”. In: *Tourism Management* 86 (2021), p. 104319. ISSN: 0261-5177. DOI: https://doi.org/10.1016/j.tourman.2021.104319. URL: https://www.sciencedirect.com/science/article/pii/S0261517721000388.

[Wan+17] Dan Wang and Juan L. Nicolau. “Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com”. In: *International Journal of Hospitality Management* 62 (2017), pp. 120–131. ISSN: 0278-4319. DOI: https://doi.org/10.1016/j.ijhm.2016.12.007. URL: https://www.sciencedirect.com/science/article/pii/S0278431916305618.

[Yan+17] Y. Yang, Z. Mao, and J. Tang. “Understanding Guest Satisfaction with Urban Hotel Location”. In: *Journal of Travel Research* 57.2 (2017). Original work published 2018, pp. 243–259. DOI: 10.1177/0047287517691153. URL: https://doi.org/10.1177/0047287517691153.

Code

The code is available at https://github.com/Damn18/Geospatial_datascience.git.