Geospatial

Martin Reinhard

April 2025

1 Introduction

Finding suitable accommodation is one of the most pressing challenges faced by university students in Italy today. During the past decade, major Italian cities such as Milan, Bologna, Turin, and Rome have witnessed a significant increase in rental prices, often disproportionate to the financial means of students and their families. The rising cost of living, coupled with a shortage of affordable housing near the university, has exacerbated the situation. Milan, for example, has seen average rental prices rise steadily year after year, pushing many students to seek housing in increasingly peripheral areas with limited access to key urban services and academic facilities [1]. For university students, the ideal accommodation is one that offers a balance between affordability, quality, and accessibility. The proximity to university buildings and efficient public transport are consistently cited as top priorities, since these factors not only reduce commute times but also improve integration into student life, increasing academic performance [2]. However, in dense urban settings, centrally located and well-connected apartments often come with a high price tag. As a result, many students turn to shared apartments as a more affordable and practical solution [3], allowing them to access better locations and amenities while splitting costs with flatmates.

In this context, spatial analysis becomes a valuable approach to investigate the key factors that influence apartment prices in urban environments. Real estate markets have long been guided by the mantra "location, location, location", emphasizing the pivotal role of geographic context in determining property value. For student accommodation, various spatial and structural characteristics, such as proximity to university faculties, accessibility to public transportation, and overall neighborhood quality, play a crucial role in determining affordability. The increasing availability of open data sets, including the GTFS (General Transit Feed Specification) for public transit and detailed housing listings, allows a more precise investigation of how these factors interact to influence rental price patterns.

This paper focuses on the city of Turin, a significant academic hub in northern Italy, home to the University of Turin and the Polytechnic of Turin. Compared to cities like Milan, Turin has a more stable and affordable housing market, yet accessibility and location remain crucial concerns for its large student population. The city's urban transport system that includes buses, trams, and a single metro line is the main means of transportation for the vast majority of students, shaping how they move through the city. Furthermore, university faculties are scattered across different districts, from the city center to more peripheral areas, adding another layer of spatial complexity to the housing search.

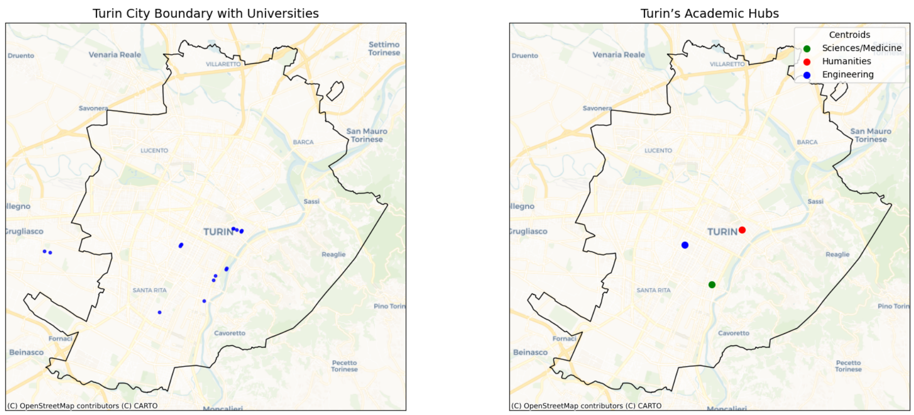
The main objective of this study is to investigate how spatial factors such as proximity to public transport routes and connection to university faculties influence apartment prices in Turin. By integrating GTFS transit data, university facilities positions, and housing listings with relative attributes (price, area, number of rooms, etc.), we aim to identify patterns in accessibility and cost. In doing so, this research focuses on analyzing how connectivity of an apartment to the university and its characteristics influences its price, as well as examining the possible spatial spillover effects and their magnitude.

By answering these questions, the study hopes to contribute to the ongoing debate on the right to education, urban equity, student housing policy, and the role of spatial accessibility in shaping apartment rent dynamics in Italian university cities.

2 Description of the data

The city of Turin has a complex academic geography, with university faculties spread across various neighborhoods and grouped into three main sectors according to their academic orientation. This spatial arrangement reflects both the historical development and the patterns of academic specialization. The humanistic faculties are located primarily in and around the historic city center and adjacent northern neighborhoods, highlighting their integration with cultural and institutional landmarks. The engineering departments are concentrated west of the city center, forming a distinct academic hub. Meanwhile, the faculties related to Medicine, Mathematics and Natural Sciences are situated further south, near the areas of Lingotto and San Salvario.

Although this intra-urban dispersion of university faculties supports academic specialization and enhances accessibility within the city, two faculties, Agricultural Sciences and Veterinary Medicine, are located well beyond the city limits of Turin. Despite being part of the University of Turin, these peripheral courses are considered of limited general interest and are located in areas with scarce housing opportunities and weak connections to the urban transit network. For these reasons and to preserve the spatial consistency of the analysis, they were excluded from the study. In addition, to better capture the spatial structure of the university system within the city, KMeans clustering was applied to the geographic locations of the remaining faculties. The three clusters obtained, which predictably aligned with the university's main academic orientations: Humanities, Engineering and Sciences/Medicine. The centroids of



(a) All University Faculties in Turin

(b) Academic Hub of Turin

each cluster represent the spatial focal point that will be used in the analysis to measure the proximity and accessibility of the apartments. This approach enables a more nuanced assessment of how student housing preferences may vary depending on the specific academic context.

When searching for accommodation, students today have access to a variety of platforms and resources. Many still turn to Facebook groups, university forums, or word of mouth to find shared housing opportunities. However, there are also several specialized online platforms dedicated to rental listings in Italy. In this study, apartment data was scraped from Idealista.com, each apartment includes key characteristics such as total area in square meters, number of rooms/roommates, number of bathrooms, and listed rental price. To enable spatial analysis, apartment locations were geocoded using Nominatim API (OpenStreetMap), although an ArcGIS-based geocoding solution was also implemented as a potential alternative and was used to refine some missing locations.

The dataset was carefully cleaned to ensure accuracy and consistency. This included refining the web scraping procedure to filter out incomplete or malformed entries at the source, minimizing the inclusion of irrelevant or duplicate listings. Further preprocessing steps involved removing any residual duplicates and excluding records with missing or ambiguous geographic information. Only listings with clearly identifiable and geocodable addresses were retained. For spatial operations, all apartment coordinates were projected to the EPSG:3857 coordinate reference system, a standard for web mapping that allows for consistent spatial analysis and integration with map tile services. This projection facilitates accurate distance calculations and alignment with base maps used throughout the analysis.

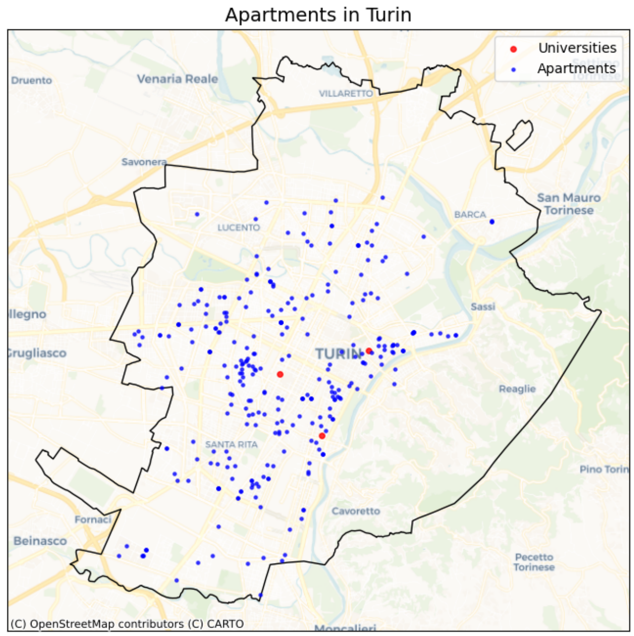


Figure 2: Students apartments

As shown in the map above2, the distribution of apartments is relatively dense and widespread throughout the urban fabric of Turin. Notably, there is a visible concentration of listings in neighborhoods such as San Salvario, Crocetta, and Santa Rita—areas (around the city Centre) that are both well-connected and in reasonable proximity to several university buildings. This spatial overlap between residential supply and university infrastructure supports the relevance of analyzing accessibility as a factor in housing pricing and distribution, especially from the perspective of students searching for affordable and well-located shared accommodations.

To further contextualize these patterns, a socioeconomic index sourced from the Torino Geoportale [4] was incorporated into the analysis. This index synthesizes various dimensions such as demographic composition, average income levels, the density of commercial activities and access to essential services, thus providing a comprehensive indicator of neighborhood quality. The map visualization reveals that neighborhoods with a higher density of apartment listings, particularly Centro, Crocetta and parts of San Salvario (three of the central areas of Turin), also score high on the socioeconomic index. This suggests that these areas are not only strategically located in terms of urban connectivity and academic proximity, but also offer desirable living conditions, making them especially attractive to students. Integrating this socioeconomic layer into the model could enable a more comprehensive understanding of how perceived neighborhood quality interacts with market dynamics to influence the distribution and pricing of student-oriented housing.

To represent and analyze the structure of Turin's public transportation system, this study leverages GTFS (General Transit Feed Specification) data [5], which offer a standardized framework containing comprehensive details on routes, stops and schedules. The Turin transport network is managed by GTT (Gruppo Torinese Trasporti) and includes an extensive range of bus services, multiple tram lines, and a single-line metro connecting the western suburb to the southern district. In this research, the focus is limited to urban transit routes, deliberately excluding regional and suburban services to better capture the relevant options available to students living within the city. As illustrated in the map 3, the public transport network in Turin is characterized by a dense and highly accessible system, primarily structured around an extensive web of urban bus lines that provide complete coverage throughout the city. This network reaches virtually all neighborhoods within the municipal boundaries, ensuring reliable access to key destinations, including university and residential areas. In addition, the spatial distribution of bus stops along with the frequency of services, as can be observed from the GTFS timetable data, highlights the efficiency and coverage of Turin's public transportation system.

Figure 3: Urban transport network and bus stops

To incorporate a representative spatial attribute into the analysis, a travel time-based accessibility indicator was calculated to quantify how reachable each apartment is from the three main university clusters. This was achieved using r5py, a Python library that interfaces with the R5 routing engine, capable of multimodal transport analysis. The core of the method involves generating a travel time matrix that calculates the duration needed to travel from each apartment listing to the centroid of each university specialization cluster.

Travel times were computed using public transport and walking, simulating departures during typical morning hours to reflect realistic commuting conditions. The three resulting attributes represent the minimum time required to reach each corresponding faculty centroids from a given apartment. Rather than relying solely on straight-line distances, this approach provides a more practical measure of accessibility, capturing the temporal cost of reaching relevant university areas and enhancing the model's ability to assess house suitability for students.

3 Data analysis

To have a broader and more in-depth view of the spatial dynamics behind the price of apartments in Turin, the local Moran statistic I was used. This method is a localized version of the global Moran's I test and is particularly effective in detecting spatial clusters and outliers by evaluating each observation in relation to its neighbors. Local Moran's I measures whether an apartment price is significantly similar or different from those of nearby listings, identifying spatial patterns that global measures may overlook. This granular approach enables the identification of specific areas where prices are clustered or divergent, offering a valuable perspective on the heterogeneity of the urban housing market.

The resulting cluster map4 highlights the spatial structure of apartments based on price. Different high-high clusters are clearly visible in central neighborhoods, areas known for better connectivity, amenities, and also proximity to universities, indicating groups of apartments with significantly above average prices surrounded by similarly expensive listings. In contrast, low-low clusters appear in more peripheral areas, suggesting concentrations of low-priced apartments likely tied to less central locations or weaker service provision. Some scattered high-low and low-high areas are also visible, marking localized price anomalies where individual apartments differ sharply from their surroundings; these may signal gentrifying areas, atypical properties, or isolated high-end or low-end accommodations.

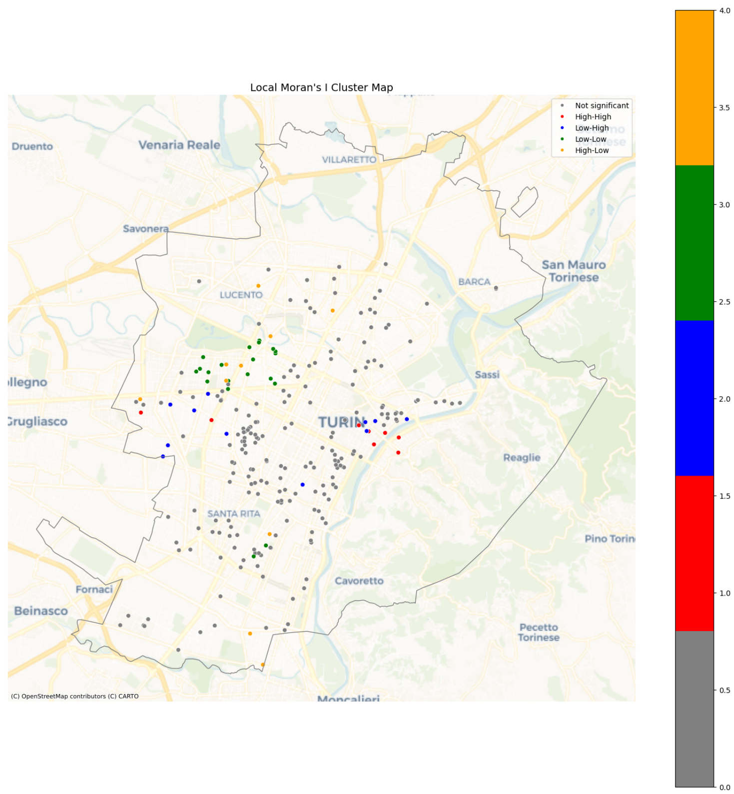


Figure 4: Local Moran I Clusters of Apartments for Price

In defining spatial relationships, the choice of spatial weight matrix must be defined. This matrix is a square, symmetric N×N array in which each element w_{ij} quantifies the spatial link between observations i and j. In the present analysis, the most straightforward binary form was adopted, where $w_{ij} = 1$ if a spatial relationship exists and $w_{ij} = 0$ otherwise. However, more complex weighting schemes, such as inverse distance or kernel functions, can also be employed. An initial attempt to define spatial relationships through a fixed geographic cuto-ff distance was ultimately discarded, as it resulted in disconnected subgroups and isolated listings, an undesirable outcome that impairs the ability to detect and interpret consistent spatial patterns. To mitigate this, a K-Nearest Neighbors (KNN) method was adopted, which connects each apartment to its **k** closest neighbors based on geographical proximity. This ensures a uniform and connected spatial network in the study area, enhancing comparability and robustness in identifying spatial autocorrelation. The Local Moran's I cluster map (Figure4) was produced using **k=20**, but this value is just an arbitrary parameter. The tuning process for the weight matrix parameter will be carried out in the subsequent modeling phase to better align with the assumptions and requirements of the spatial regression analysis.

This analysis applies spatial econometric techniques to account for spatial dependence in housing prices. According to the model selection strategy proposed by Elhorst [6], the analysis of spatial data should begin with the estimation of a conventional Ordinary Least Squares (OLS) model, followed by statistical tests on the residuals, specifically Lagrange Multiplier (LM) tests, to assess whether spatial effects are present and, if so, whether they are best captured by a Spatial Autoregressive Model (SAR), a Spatial Error Model (SEM) or a more general model such as the Spatial Durbin Model (SDM).

Table 1: OLS: Variable Coefficients and Significance

Table 2

Variable	**Coefficient**	**Std. Error**	**p-value**
CONSTANT	469.243	55.567	0.000
Area (m^2)	0.464	0.239	0.053
Rooms	-13.353	6.344	0.036
Bathrooms	32.316	16.154	0.046
Ind. socioeconomico	-0.050	0.604	0.935
Travel time to centroid 0	-0.177	0.357	0.620
Travel time to centroid 1	-0.476	0.302	0.116
Travel time to centroid 2	-0.548	0.468	0.243

R-squared : 0.0654 / Akaike Info Criterion (AIC) : 3733.649

The table 2 presents the OLS results. The model achieves an R^2 of 0.065, indicating a very scarce ability to explain variations in housing prices. Among the predictors, area, rooms and bathrooms are statistically significant, with coefficients in line with economic intuition: larger properties and those with more bathrooms tend to be priced higher, while the number of rooms (so the number of flatmates) has a negative sign, indicating that shared living arrangements may reduce individual willingness to pay for a property.

Table 3: Diagnostic Tests

Test	**Value**	**p-value**
Jarque-Bera (Normality)	2904.659	0.0000
Breusch-Pagan (Heteroskedasticity)	133.148	0.0000
Koenker-Bassett (Heteroskedasticity)	16.020	0.0249

Other covariates, including travel times to academic hubs and the socioeconomic index - attributes depending on the position - were not statistically significant at conventional levels. The residual diagnostics indicate violations of key OLS assumptions. Specifically, the Jarque-Bera test strongly rejects the normality of residuals (p ¡ 0.001), and both Breusch-Pagan and Koenker-Bassett tests suggest the presence of heteroskedasticity.

To assess whether spatial dependence is present in the residuals of the OLS model, Lagrange Multiplier (LM) tests for spatial lag and spatial error dependence were applied using a series of spatial weight matrices defined by increasing values of **k** in a k-nearest neighbors framework. Varying the value of **k** allows for an examination of how the spatial configuration influences the presence and nature of residual autocorrelation. For each configuration, standard LM-lag and LM-error tests, as well as their robust counterparts, were computed. This approach avoids the pitfall of relying on an arbitrary spatial structure.

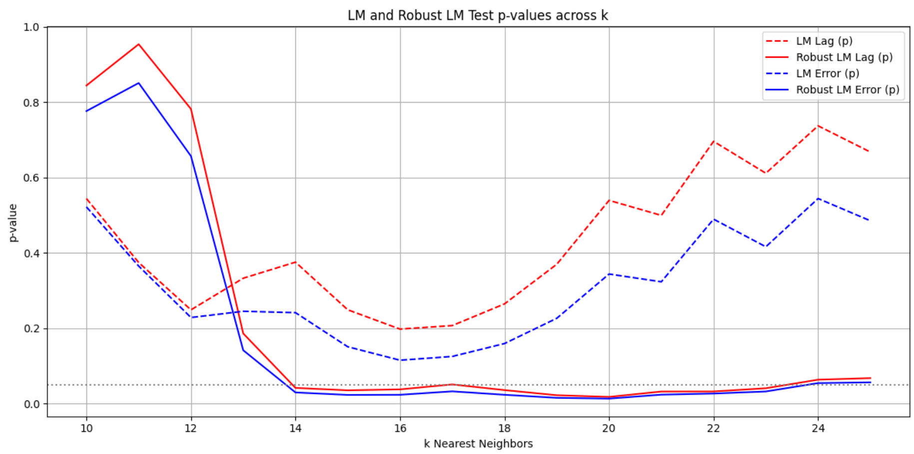


Figure 5: LM and Robust LM Test p-values across k

The results, as shown in the plot 5, demonstrate a clear trend: for $\mathbf{k} > 14$, only the robust LM-lag and robust LM-error tests yield statistically significant values, while the standard LM statistics fail to reach significance. This pattern indicates that traditional diagnostics may overlook spatial autocorrelation, particularly when heteroskedasticity is present in the residuals. This finding aligns with earlier evidence from the Koenker-Bassett and Breusch-Pagan tests, which strongly suggest the presence of heteroskedasticity in the OLS residuals. Therefore, the significance of only the robust LM tests was anticipated, reinforcing the necessity of a spatial econometric model. Given the dual presence of spatial dependence and heteroskedasticity, a more comprehensive specification, such as the spatial Durbin model (SDM), which accounts for both lagged dependent variables and spatially correlated errors, is recommended as the most suitable approach.

Based on previous diagnostic tests that will not be presented in this report (altought present in the Python notebook), the parameter $\mathbf{k}$=15 for the spatial weight matrix was selected and used in the SDM analysis, as it yielded the best overall fit and stability of the parameter estimates. The SDM specification augments the standard spatial lag model by including spatially lagged covariates.

In the context of student apartment rents, this allows us both to capture direct effects of an apartment's own characteristics and to quantify how the same attributes in surrounding units exert spillover influences.

Compared to the baseline OLS model, the SDM exhibits an improved overall fit, with a higher pseudo R-squared (0.1574) and a lower Akaike Information Criterion (AIC = 3724.34), indicating a slightly better specification of spatial effects and price interactions.

Table 4: SDM: Variable Coefficients and Significance

Variable	**Coefficient**	**p-value**
CONSTANT	690.28375	0.00000***
Area (m^2)	0.37676	0.10326
Rooms	-15.55499	0.01008**
Bathrooms	37.73818	0.01381**
Ind. socioeconomico	-0.36203	0.65999
travel_time_to_centroid_0	-2.48212	0.32630
travel_time_to_centroid_1	5.77420	0.02593**
travel_time_to_centroid_2	-3.95461	0.04937**
W_Area (m^2)	2.20295	0.03432**
W_Rooms	-124.33903	0.00001***
W_Bathrooms	253.70495	0.00111***
W_Ind. socioeconomico	0.40019	0.78335
W_travel_time_to_centroid_0	3.35087	0.20102
W_travel_time_to_centroid_1	-7.09768	0.00728***
W_travel_time_to_centroid_2	2.47270	0.26947
W_Price	-0.64485	0.00283***

Pseudo R-squared : 0.1574 / Spatial Pseudo R-squared : 0.1063 / AIC : 3724.336

In terms of covariates, the number of rooms and bathrooms remains strongly significant, consistent with the OLS results, but their spatial spillovers amplify in magnitude and clarity. The direct effects confirm that more rooms reduce price (–15.55, $p < 0.05$), while more bathrooms significantly increase it (37.74, $p < 0.05$). However, indirect (spillover) effects are even more pronounced: nearby apartments with more bathrooms significantly push prices upward (indirect effect = 139.45), while an increase in neighboring room count lowers prices (–69.49). This might reflect the fact that areas dominated by large multi-room flats could be associated with less demand for individual student rentals, whereas high bathroom count signals better amenities, attracting student interest and influencing nearby rental markets.

Interestingly, the significance of travel times to faculties changes considerably. In the OLS model, no travel-time variables were statistically significant. However, in the SDM, accessibility to the Humanities (centroid1) and Engineering (centroid2) hubs shows significant effects: being farther from the Humanities hub increases apartment prices (coefficient = 5.77, $p < 0.05$), suggesting these areas might be less attractive to students and thus have higher prices due to different tenant profiles. In contrast, distance from Engineering is negatively associated with price (–3.95, $p < 0.05$), implying that proximity to this hub is capitalized into housing costs, perhaps reflecting higher demand from engineering students for nearby apartments. The indirect effects partially offset these trends, though the total impacts remain consistent in direction.

The spatial lag coefficients of explanatory variables (the "W_" terms) further reinforce the spatial dependence in the housing market. For instance, neighboring prices (W_Price) are negatively associated with the price of a given apartment (–0.64, $p < 0.01$), which could suggest some local price competition or saturation effects. The significance of spatial lags for rooms, bathrooms, and accessibility metrics further confirms that student housing markets in Turin exhibit strong local interactions, both in physical space and in attribute space.

Finally, the spatial lag of price (ρ=-0.6448, p=0.0028) is highly significant, confirming that once local amenities are maintained constant, higher neighboring rents actually reduce a unit's own rent, according to the competitive nature of the student market.

4 Conclusions

This study aimed to explore how spatial dynamics, particularly connectivity to university hubs, affect the pricing of student-oriented apartments in Turin. By integrating spatial econometric techniques with data on public transport and university locations, the analysis aimed to capture not only the direct influence of proximity and housing characteristics, but also the spillover effects from neighboring units and areas. The Spatial Durbin Model was employed to account for these complexities, offering a comprehensive framework to interpret both local and spatially lagged influences on apartment prices.

The findings confirm that traditional structural variables, such as the number of rooms and bathrooms, remain significant determinants of price. Notably, these effects extend spatially: the presence of more bathrooms or rooms in nearby listings positively influences the price of a given apartment, suggesting that students value these features not only for their own accommodation but perhaps also as proxies for the quality of the surrounding housing stock. Interestingly, proximity to different university hubs yields differentiated effects. Although accessibility to the Engineering centroid is associated with lower prices, the Humanities centroid shows a positive relationship between proximity and price. This contrast might reflect student preferences or underlying differences in neighborhood desirability. For instance, higher annual application numbers to Engineering programs may create demand pressure across a broader area, driving prices up even farther away, while the Humanities hub, located in a more central or sought-after district, commands higher rents in its vicinity.

The Science/Medicine centroid did not display significant effects, which could imply either that housing prices around this cluster are already equilibrated with respect to accessibility or that the spatial variance in connectivity to this faculty is limited. Furthermore, the insignificance of the neighborhood socioeconomic index in both direct and spillover effects suggests that students prioritize practical aspects such as apartment characteristics and commute times over broader contextual factors like income level or social prestige of a neighborhood. This aligns with a utilitarian view of student behavior in rental markets, where affordability and access are paramount, potentially overriding longer-term considerations typical of family or investment buyers.

Another notable outcome is the significant and negative spatial lag of the dependent variable, implying that higher prices in neighboring areas exert a downward pressure on a unit's price. This could signal competitive effects on the market, possibly due to an oversupply in some microareas or indicate that students compare rents across nearby listings to determine relative value. It underlines the importance of modeling both direct and indirect interactions when evaluating price mechanisms.

Despite the valuable insights offered, the analysis is subject to limitations. The use of cross-sectional data prevents capturing dynamic temporal trends in student preferences or market responses. The model also relies on approximated centroids for university clusters, which may not perfectly reflect actual student flows or perceptions of accessibility. Additionally, GTFS-based travel times, while robust, may not fully account for individual travel behaviors or multimodal integration, such as biking or walking. Expanding the dataset to include more granular indicators (e.g. distance to nightlife, libraries, or supermarkets) or panel data could enrich future analyses. In summary, the SDM reveals a complex but interpretable spatial structure in the student housing market in Turin. Proximity to certain academic hubs and the configuration of nearby listings play a significant role on the price. These findings reinforce the importance of incorporating spatial dependence and spillovers into housing models, particularly in urban contexts where mobility and localized preferences are critical.

References

[1] Mugnano, S., Giannotti Mura, C., Costarelli, I., Ramello, R. (2024). *Le nuove sfide dell'abitare studentesco e del diritto allo studio nella città di Milano*, L'università pubblica italiana per città e territori. Politiche, casi e pratiche (pp. 19-31). Università di Bologna. Available at: https://hdl.handle.net/10281/543102

[2] Iqbal, M., Morrissey, H. and Ball, P. (2020). *The influence of accommodation on the academic performance of university students*, International Journal of Current Research, 12(4), pp. 10900-10907. Available at: http://hdl.handle.net/2436/623199

[3] Giuliana Costa and Tommaso Frangioni (2024). *The Housing Market for Students in Milan: Pathways, Actors and Dynamics*, Dipartimento di Architettura e Studi Urbani Politecnico di Milano Delft, 28 August 2024. Available at: https://www.academia.edu/124727871/The_Housing_

Market_for_Students_in_Milan_Pathways_Actors_and_Dynamics#loswp-work-container

[4] Available at: http://geoportale.comune.torino.it/web/node/2211

[5] Available at: http://aperto.comune.torino.it/dataset/feed-gtfs-trasporti-gtt

[6] J. Paul Elhorst (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, University of Groningen Faculty of Economics and Business, Groningen, The Netherlands. Available at: https://doi.org/10.1007/978-3-642-40340-8