

Urban growth and stratification: The role of locational externalities

Julián Moral-Carcedo^{*}

Universidad Autónoma de Madrid, Spain
Economic Theory Department & L.R.Klein Institute, USA

ARTICLE INFO

JEL classification:

R23

R15

Keywords:

Urban growth

Stratification

Spatial equilibrium

ABSTRACT

Urban areas that have undergone rapid and recent growth processes can shed light on the forces driving urban stratification and gentrification. This article analyzes how the process of urban growth itself, driven by migration, affects the stratification of the city. The availability of detailed data makes it possible to observe how high-income settlement patterns are driven by attraction (homophily) and repulsion (heterophobia). We consider neighborhood composition (i.e., the number of high-income residents in the vicinity of a given location) as a potential factor that can explain agents' higher willingness to pay in some locations. We propose a model of urban spatial equilibrium with locational externalities, related to preexisting neighborhood characteristics, that can help explain why new rich neighborhoods are located close to existing rich neighborhoods and far from poor neighborhoods. In this model a monopolistic owner will allocate available locations to those agents willing to pay the most, taking advantage of agents' willingness to pay more when locations are surrounded by high-income individuals. The location externality influences the dynamics of urban configuration through attraction-repulsion effects, as new rich neighborhoods are located near other rich neighborhoods and far from poor neighborhoods, thus reinforcing socio-spatial-stratification. This mechanism sheds light on why gentrification occurs only in some areas, while others with a priori better localization do not experiment changes.

1. Introduction

Although urban areas generally show varying degrees of spatial mixing (Bayer & McMillan, 2012) socio-spatial stratification is a common feature that can be observed in large cities. Neighborhoods with extreme incomes, where high- or low-income people live, are easily identifiable in the urban structure as being grouped in clusters. As previous works such as (Brueckner & Rosenthal, 2009) (Baum-Snow & Pavan, 2013) (Bayer & McMillan, 2012) have shown, clustered spatial configuration is the result of income inequality, commuting and mobility costs, the availability of local amenities, and housing characteristics. However, socio-spatial stratification also reflects the propensity of high-income people to live near other high-income people, as pointed out by (Guerrieri et al., 2013) (Bayer et al., 2007), among others. In addition to the fact that higher-income individuals may be located in areas with better characteristics (low pollution, better access to green areas, better access to amenities, landscape values, etc.), the choice of locations close to other people with higher incomes may be related to the formation of social networks (Kleit, 2008), where factors such as homophily, in which ties between similar individuals are likely

to be formed, play a role (Schelling, 1969). As (Kaufman, 1952) points out, these networks of contacts, and the individual's position in the network (Bothner et al., 2010), are important factors in the self-perception of status. Similarly, the extent and reach of this network of contacts also depends on social status, as (Cao & Smith, 2020) and (Smith et al., 2011) point out. Since physical proximity influences the frequency of contacts in networks (Stopczynski et al., 2018) (Hill & Dunbar, 2003) it is clear that individuals' choice of residential location is closely related to social status.

High-income individuals agglomerate in spatial clusters in high-income neighborhoods, but low-income individuals are also observed to be spatially grouped. While income allows individuals in the wealthier group to optimize the result of housing preferences, the availability of local amenities, or proximity to workplaces, among other factors, for low-income households housing affordability is the main reason to choose a location, regardless of other important aspects like commuting time. According to the "spatial mismatch hypothesis," low-income residents (Kain, 1968), often immigrants or members of minority groups, tend to live in neighborhoods that are far from the areas where jobs are located, which is detrimental to their labor market

^{*} Universidad Autónoma de Madrid, Fac. CC. Economicas, Cantoblanco Campus, Madrid 28049, Spain.

E-mail address: julian.moral@uam.es.

outcomes and lifetime earnings. (Turner, 2008). Also (McCormick, 1986) showed that minority workers in Great Britain have significantly longer commutes, while (Blázquez et al., 2010) found similar results for immigrants living in the metropolitan area of Madrid.

Static monocentric spatial models à la Alonso-Mills-Muth can explain a socio-spatially stratified urban configuration in which high-income people are located near the city center or central business district (CBD), minimizing commuting costs, due to their higher willingness to pay (Brueckner, 1987). Considering heterogeneous agents (high and low income with different commuting costs/value of time) these models predict that housing/rental prices will decrease with distance from the CBD. The high opportunity cost of commuting time for more affluent people (Su, 2022) makes the central areas of cities more valuable for households with higher incomes, despite their preference for larger houses that are cheaper in places further away from the center.

Although theoretical models derived from the standard Alonso-Mills-Muth model can provide adequate explanation about the observed stratification in cities, these models are purely static while urban spaces are essentially dynamic. The urban area expands through sprawl processes, but existing neighborhoods can also change through increasing land use intensity (verticalization and urban densification). Urban spatial configuration may change over time as the urban area grows, as people move both within the same urban area and to other urban areas. The immediate question is how these changes affect the observed socio-spatial stratification in a particular urban area.

When intense migration inflows occur or if better transport infrastructures or other transport innovations are available, cities may undergo expansion processes (Duranton & Turner, 2012) (Edward L. Glaeser, 2004). In some cases, the expansion of cities is limited by geological factors (coastline, water bodies, mountains, ...) so that the intensity of residential land use should increase to accommodate newcomers, who must compete with established residents for the existing sites. The variation in the price of housing is the basic mechanism that rations the limited availability of housing to those who can afford it. The expansion of cities and their socio-spatial stratification is influenced by the demand for housing from individuals with a high willingness to pay. The demand for housing from these groups can lead to gentrification processes and push up housing prices as high-income residents move into previously disadvantaged neighborhoods, as pointed in (Kidokoro et al., 2023) (Zhu et al., 2022), usually after some kind of urban intervention (i.e., the demolition of buildings in ruins, the re-routing of railroad tracks, the undergrounding of stations, or the redevelopment of industrial areas, etc.). On the other hand, low-income new residents tend to occupy the spaces left by other residents who move to other parts of the urban area (Bayona-Carrasco & Gil-Alonso, 2011) (Gallego-Valadés et al., 2021). If new high-income residents replace the previous low/medium-income residents, areas undergoing gentrification processes may be revitalized. However, if new residents with low income replace other low-income residents, the neighborhood income characteristics, and spatial inequalities will persist. Although their inhabitants change, the problems they suffer will not.

This paper explores the characteristics of the expansion process of the urban area of Madrid (Spain) using highly granular data and how this process affects their socio-spatial stratification. As a novelty, in this study we analyze the complete dataset of residential buildings by construction date in the entire metropolitan area of Madrid, which includes not only the most populous municipality (the municipality of Madrid), but also the neighboring municipalities with which it forms an area characterized by intense mobility flows to and from the municipality of Madrid. Without the inclusion of these urban centers, it is impossible to understand the process of urban growth in the metropolitan area and the location patterns of individuals, which do not depend on infra-regional administrative boundaries, but on other factors such as housing affordability and job accessibility. The metropolitan area of Madrid is a particularly interesting case for study due to the city's intense growth in

recent years in terms of both its extension and population. The population living in urban areas in Spain increased the share of from 65 % to >87 % over the period 1950–2018, as by (Gutiérrez et al., 2020a, 2020b) (Sánchez-Moral et al., 2018) pointed out. Since the end of the 20th century, urban areas with a population of >500,000 inhabitants have experienced strong demographic momentum mainly due to flows from medium-sized cities and towns to more populated urban areas. In the case of the Madrid region, the population has increased from 1.6 million (1940) to 6.8 million (2021), with a notable expansion in the 1960s, 1970s, and 2000s. While internal migration flows (from the rural countryside to industrial cities) explain the huge variation in population in the 1960s, the baby boom played a key role in the increase in population and size of the city in the 1970s. In the 2000s, external migration inflows explain why the Madrid region, and the whole country, registered the highest population growth in the last 50 years. In addition to this inflow of low-skilled migrant workers, the city of Madrid also had a significant inflow of high-skilled workers. According to data of the Spanish National Statistics Institute (INE) for 2019, technicians and scientific professionals accounted for 18.5 % of all employed persons in Spain, but 25.5 % of total employment in Madrid. This concentration of high-skilled workers is due to a “headquarters” effect as Madrid is the administrative capital of Spain, but also, as (Sánchez-Moral et al., 2018) (Roca & Puga, 2017) have argued, skilled workers move to Madrid due to the positive effect such a move can have on their careers. In recent years, the progressive reduction in wealth and inheritance taxes in Madrid may have been a factor of attraction for wealthier individuals.

Under the spatial configuration predicted by the monocentric Alonso-Mills-Muth model, one of the most striking aspects of Madrid is its distribution of individuals according to income level and distance from the city center. Although the data show that it is unlikely to find high-income individuals at distances >20 km from the city center, both low- and high-income individuals can be found in areas close to the city center. Another interesting fact is that new high-income urban areas seem to have appeared following a particular north/north-west orientation. In this article, we propose a model of urban spatial equilibrium with locational externalities related to the characteristics of preexisting neighborhoods that can help to explain these facts. As in (Guerrieri et al., 2013), we consider neighborhood composition (i.e., the number of high-income workers in the vicinity of a given location) as a potential factor that can explain agents' higher willingness to pay in some locations. As a departure from the works of other authors, in the model presented in this paper we include a monopolistic landlord who will allocate available locations to those agents who will pay the most, taking advantage of agents' willingness to pay more when locations are surrounded by high-income individuals. This modification makes it possible to exclude the possibility of multiple urban configurations in equilibrium, as occurs in (Guerrieri et al., 2013), and to characterize the equilibrium through a new approach that exploits the recursion in the location externality.

As an additional contribution, the proposed model includes the possibility that the spatial stratification in equilibrium can be modified because of variations in the inflows of inhabitants. When high-income individuals are more likely to enter the city (e.g., due to the growth of sectors that demand high-skilled jobs), gentrification processes can reduce the urban area occupied by low-income individuals. The replacement of low-income residents with more affluent residents increases the value of the locational externality, even in previously affluent neighborhoods, reinforcing urban stratification. As will be analyzed below, this effect is observed in the case of Madrid, as the areas that have experienced the highest growth in per capita income are those where there has historically been a greater concentration of high-income neighbors.

In the following section, we examine the characteristics of urban growth in Madrid after the Spanish Civil War and the role of migration flows. We then describe the observed income stratification in the Madrid region and present the main conclusions of the proposed theoretical

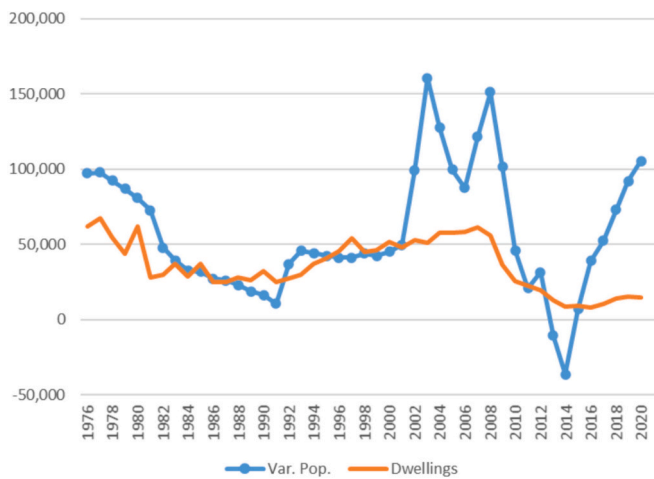


Fig. 1. Urban developments in Madrid region. Number of dwellings per year and variation in population (persons) 1975–2020.

Source: own elaboration with data from the Dirección general de Catastro: Ministerio de Hacienda. Spanish government. <http://www.sedecatastro.gob.es/>.

model of urban spatial equilibrium with locational externalities. Finally, we test the role of locational externalities in the social stratification observed in the Madrid region.

2. Evolution of Madrid metropolitan area: urban expansion and migration

2.1. General overview

Population inflows shape urban areas and drive the construction of new buildings and neighborhoods. The metropolitan area of Madrid has expanded in a non-uniform manner. Data¹ from the Dirección General del Catastro and population data indicate (Figs. 1–2) notable changes in the city's urban expansion, with two clear periods characterized by very high construction rates and significant increases in population in the periods 1975–1980 and 1990–2010. This was followed by a period of low construction activity due to the severe impact of the subprime crisis

¹ To analyze these developments in the urban area of Madrid we employ two key datasets. Firstly, we use administrative data on buildings from the Spanish registry of real estate properties and taxation (Dirección General del Catastro: Ministerio de Hacienda). This dataset collects georeferenced information about building surface area, the final year of building construction, and the number of dwellings per building (for 404,219 residential buildings). Data for the whole region of Madrid is considered. For the sake of clarity, however, the maps show only the municipality of Madrid and the surrounding municipalities where the largest share of population is concentrated. Secondly, we employ granular data on population and income from the INE. The INE dataset contains detailed information about the population and per capita income at the census tract level which allow us to assign a theoretical per capita income to each building as per capita income in the census tract where the building is located. A census tract comprises a geographical area belonging to a municipality in which approximately 2500 people reside. The Madrid region is divided into a total of 4240 census tracts. The census tract level has a variable area that depends both on the area of the municipality the section belongs to and the population that resides in it. The least populated municipalities have a small number of census tracts or only one tract. Due to statistical confidentiality, data on municipalities or census tracts with <100 inhabitants are not published. Highly populated municipalities have several census tracts whose extension depends on the degree to which the population is concentrated. Per capita income data is capped at €45,000. For income data, see https://www.ine.es/experimental/atlas/experimental_atlas.htm. Mobility data can be found at https://www.ine.es/experimental/movilidad/experimental_em.htm.

on the Spanish economy, but significant population growth due to migration flows. Another particularly interesting fact is the progressive reduction in the intensity of construction (number of dwellings per building surface area) in the Madrid region until 1990, which stabilized thereafter. This behavior points to a progressive process of urban sprawl, in which the urban area is expanded with new buildings that have a smaller number of dwellings (single-family houses, semi-detached houses, etc.) to the detriment of residential towers and high-rise buildings.

Migration flows to and from other areas of Spain and the rest of the world have had an important effect on population changes in the Madrid region. According to the survey on residential variations conducted by the INE, the percentage of individuals who change their residence per year (i.e., the sum of ingoing and outgoing flows) in the Madrid region is usually <10 % of the total resident population. Therefore, a significant percentage of individuals do not change their residence or if they do, they remain in the same municipal area. Table 1 shows the origin and destination of individuals who have changed their place of residence at the municipal level for all of Spain in the period 1988–2020.

The data in Table 1 reveal significant migratory inflows to Spain from other countries, with >10 million people entering the country over the period 1988–2020 that is partially offset by an outflow of close to 5 million people (both returns of migrants and migration flows of Spaniards abroad). The movements of people in the Madrid region show a positive net balance that is especially high in regional municipalities other than Madrid. This net entry of people can be explained by flows from the rest of the world and from people from the Madrid municipality. In the case of the municipality of Madrid, the net balance with the rest of Spain is positive (885,000 entries and 863,000 departures).

Between 1988 and 2020, around 1.2 million residents of Madrid moved to other municipalities in the region, while there was a flow of 700,000 people in the opposite direction. Similarly, almost one million residents of the Madrid region moved to other municipalities in the region, excluding the municipality of Madrid. These changes in residential patterns in the Madrid metropolitan area (which includes both the Community of Madrid and the neighboring municipalities of the same region) are linked to the process of urban expansion away from the city center. During the period considered, the number of dwellings built in the urban area outside Madrid exceeded those built in the municipality of Madrid to accommodate the growing number of inhabitants in this area, many of whom came from the same municipality of Madrid, especially in the period after 2000. This process may have been driven not only by demographic pressure, given that the baby boom cohort (1965–1976) in Spain began to emancipate themselves in those years and are trying to find affordable housing in the outskirts and municipalities bordering Madrid, but also by the decision to change residence in search of larger houses on the city outskirts. In the case of the municipality of Madrid, it is observed that the variations in residence offer a negative net balance over the period 1988–1999 and in the toughest years of the subprime crisis and subsequent sovereign debt crisis in Europe (2010–2015). In the recovery of positive balances, the flows from the rest of the world (mostly economic immigration) are of fundamental importance.

2.2. Characteristics of the expansion of the Madrid urban area

Population growth in urban areas tends to increase rents and house prices when the available housing stock is fixed, as pointed by (Morteza Moallemi, 2020) (Libertad Gonzalez, 2013) (Mussa et al., 2017) (Saiz, 2007), which stimulates new housing construction. Population growth and migration reduces the risk that new property developments will not attract buyers, thus facilitating the construction of new houses that drives urban expansion and the renewal of consolidated urban areas (Chiu On Ki, 2010). The way an urban area evolves will depend on both the space available for expansion, that can be limited by geological barriers, and the institutional framework (land and property laws,

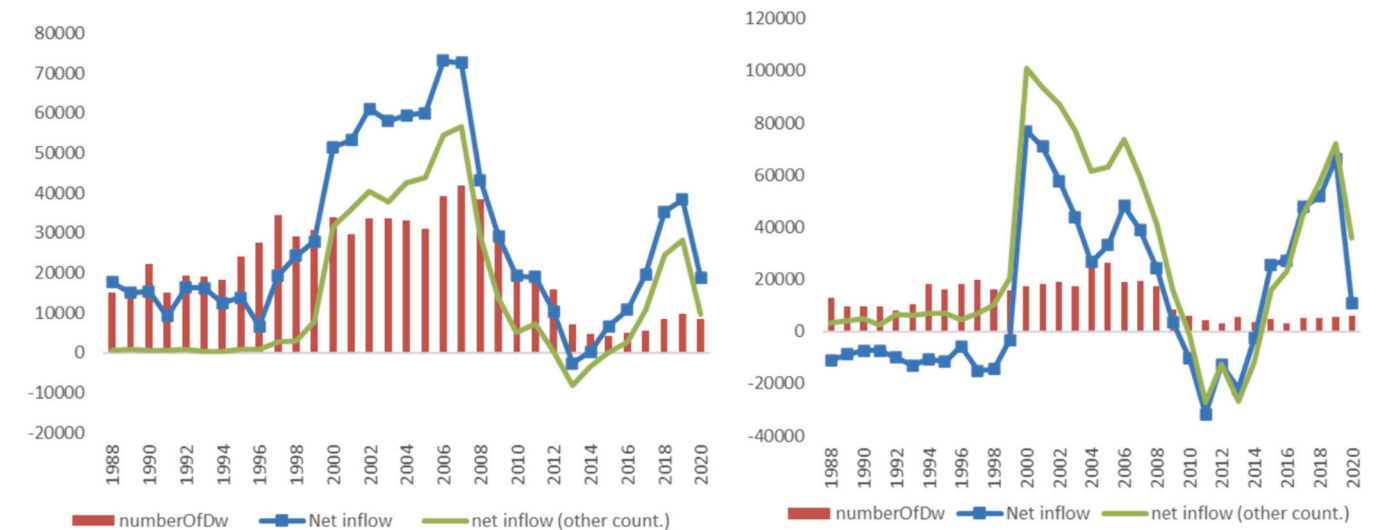


Fig. 2. Left. Total net inflow of people by year, net inflow due to external migration, and number of dwellings in buildings by year of construction. Right: Madrid municipality and (left) the rest of the Madrid region.
Source: own elaboration using INE data and data from the Dirección General del Catastro: Ministerio de Hacienda. Government of Spain. <http://www.sedecatastro.gob.es/>.

Table 1
Number of individuals that changed their place of residence (municipality) in Spain by origin and destination. Cumulative data for the period 1988–2020.

Origin	Destination				Total outflow
	Madrid city	Rest of Madrid region	Other countries	Rest of Spain	
Madrid city	–	1,193,124	686,654	863,863	2,743,641
Rest of Madrid region	701,495	974,382	357,168	578,363	2,611,408
Other countries	1,613,167	840,678	–	8,223,850	10,677,695
Rest of Spain	885,170	535,902	3,694,285	17,591,690	22,707,047
Total inflow	3,199,832	3,544,086	4,738,107	27,257,766	
Net inflow	456,191	932,678	- 5,939,588	4,550,719	

Note: data for municipalities with <10,000 inhabitants are not included, so the figures presented in the table may be underestimating true flows.
Source: own elaboration using INE data.

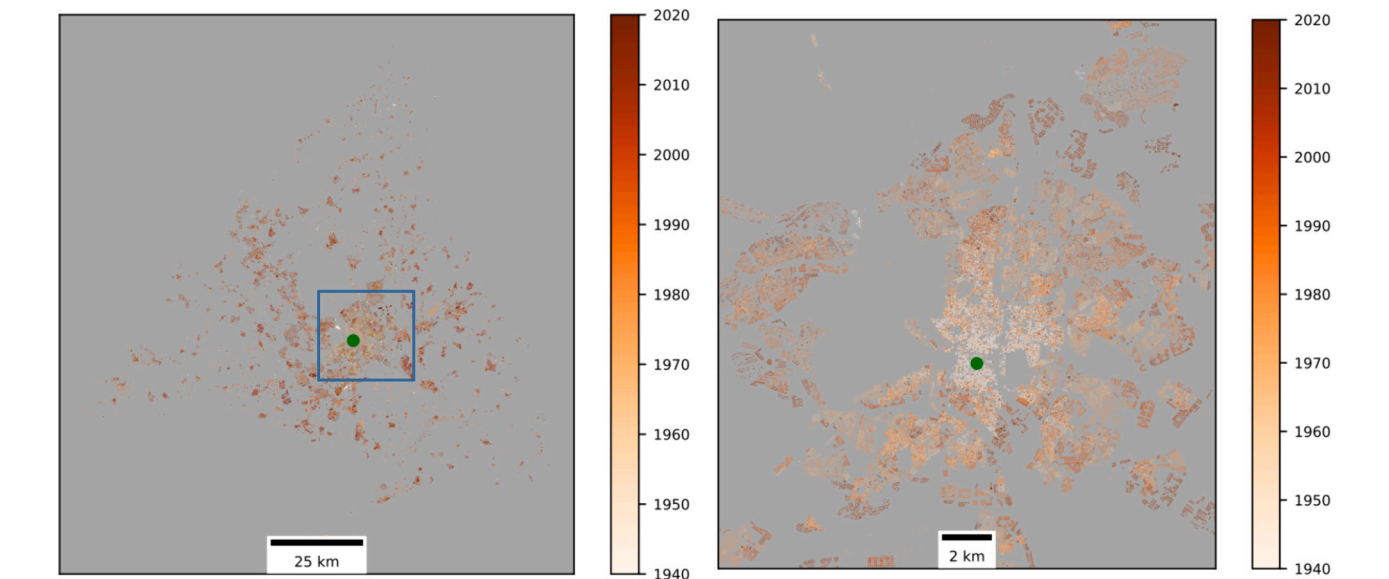


Fig. 3. Residential buildings in the region (left) and urban area (right) of Madrid by year of construction and city center (Green dot: Puerta del Sol).
Note: the rectangle in the left map shows the area plotted in the right map in detail.
Source: own elaboration using data from the Dirección General del Catastro: Ministerio de Hacienda. Government of Spain. <http://www.sedecatastro.gob.es/>. The Geopandas library for Python (Jordahl et al., 2020) was used to plot the data.

financial markets, environmental protection regulation, cultural heritage laws, ...). The current size of the urban area and the existing transportation network are also important factors in determining the potential for expansion. Commuting costs and accessibility also affect the attractiveness of available housing and house prices (Jayantha Wadu Mesthrige, 2022) (Ahlfeldt, 2013) (Berkas & Gaetani, 2023).

An overview of the evolution of the metropolitan Madrid urban area (City of Madrid and their surrounding area) is shown in Fig. 3, where a detailed map of buildings by year of construction (1940–2020) in the Madrid urban area is provided. An immediate conclusion is that as we move away from the urban center (green point) the date of construction is more recent, but it is also clear that the behavior is not uniform. Fig. 4 uses the same information, but in this case this graph shows the histogram of buildings per construction year according to their distance to the city center (Puerta del Sol). Before 1970, new buildings were mainly concentrated in areas near the center of the city, while in the following decades urban growth became increasingly dispersed with a clear tendency to be located on the outskirts due to improvements in the road network in the Madrid region.²

The resulting urban configuration shows a high concentration of residential buildings with an average age of <50 years, located >10 km from the city center, while older buildings are mainly located near the CBD. Urban expansion in the Madrid region is a relatively recent phenomenon (Díaz-Palacios-Sisternes et al., 2014) and is highly concentrated in municipal entities surrounding the Madrid municipality (the municipality has an extension of only 604 km², that is, a ball of radius of around 14 km). While the distribution of buildings according to their distance to the city center is coherent with the spatial monocentric model (i.e., a decreasing number of buildings with distance from the city center), the differences in location according to the income of dwellings built after 1970 (before that date the buildings were mainly located at distances of <10 km from the center of the city) are of more interest, as shown in Fig. 5. The data, in line with the outcomes of the monocentric spatial equilibrium model, clearly show a negative gradient for the number of dwellings with distance, where new dwellings are progressively displaced to more distant locations. New dwellings near the city center seem to point to partial neighborhood renewal and a “filling in the gaps” dynamics in the urban space, but the most important part of the new buildings is concentrated in locations at distances >10 km.³ Both in the 1970–1990 and post 1990 period,⁴ wealthier households showed a clear preference for locations near the city center, and it is unlikely that new buildings at distances >25 km were occupied by high-income households. The concentration of high-income households at shorter distances poses the immediate question as to what the main patterns of income stratification in the urban area are.

3. Income stratification and urban expansion

Big urban areas are centers of attraction, both because of the greater supply and diversity of goods and services and because of their opportunities for workers and firms. Some authors (Dauth et al., 2019) (Yankow, 2006) (Roca & Puga, 2017) (Lehmer & Möller, 2010) point to an stylized positive relationship between wages and city size.

² While the high capacity road network covered 218 km in 1985, it increased to 684 km in 2000. In 2019, the network was further extended to 976 km, with a significant increase of 225 km in the period 2000–2005. Simultaneously, the stock of vehicles increased exponentially with 566 vehicles per km of road (total road network) in 1987, reaching 1527 vehicles per km in 2019 (Data from the Madrid Region Statistics Institute).

³ As a reference, note that the distance between Madrid and the main towns of the surrounding provinces is 52 km (Guadalajara), 67 km (Toledo), 68 km (Segovia), and 81 km (Avila).

⁴ The second expansion phase of the urban area of Madrid began around 1990 (see Fig. 1).

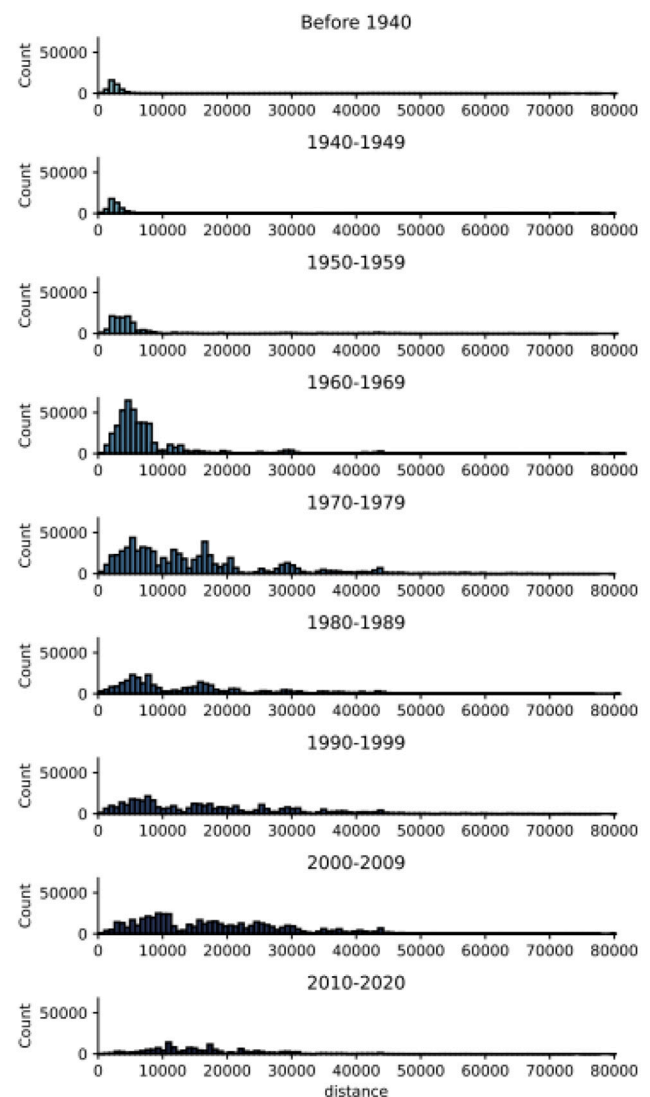


Fig. 4. Distribution plot of new buildings by year of construction (10-year periods) according to their distance (meters) to the city center (Puerta del Sol). Weighted by number of dwellings.

Source: own elaboration using data from the Dirección General del Catastro: Ministerio de Hacienda. Government of Spain. <http://www.sedecatastro.gob.es/>.

Agglomeration economies have a positive effect on wages, which widens with more years of work experience in big cities (Roca & Puga, 2017). The growth of high-skill, high-wage jobs also increase the demand for low-wage workers in low-skill jobs (Autor & Dorn, 2013) (Lindley & Machin, 2014), and both effects can foster income inequality between individuals living in the same city (Baum-Snow & Pavan, 2013). The concentration of high-income/low-income households in different spatial locations and the process of job polarization, marked by high job growth for both high- and low-wage occupations, can exacerbate spatial disparities in income (Lindley & Machin, 2014) (Berkas & Gaetani, 2023) and socio-spatial stratification of the urban area.

From a spatial point of view, the Madrid region (Fig. 6) shows a clear socio-spatial stratification in terms of income: households with higher incomes are located in the north and west of the Madrid metropolitan area, while households with the lowest incomes are concentrated in the south-east and in the areas furthest away from the metropolitan area. Under the spatial configuration predicted by the monocentric Alonso-Mills-Muth model, one of the most striking aspects in the configuration of the city of Madrid is how individuals are located (Fig. 7).

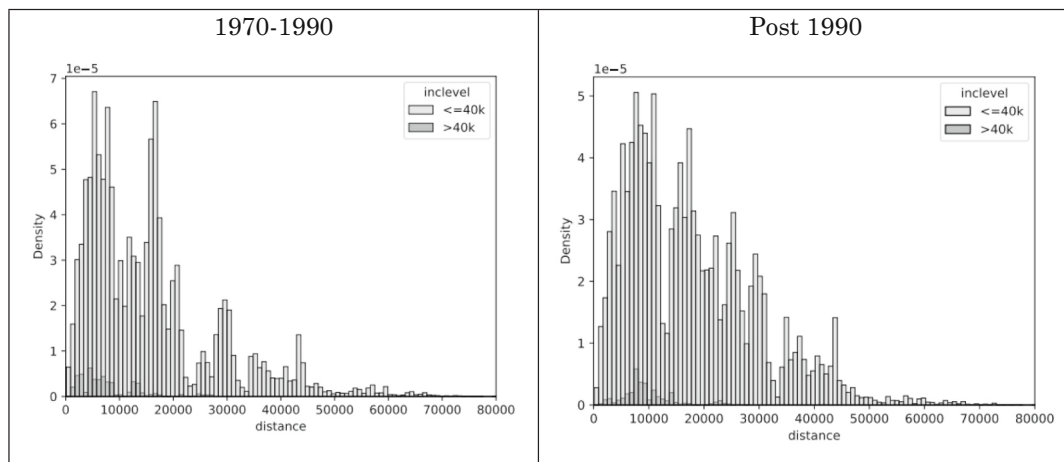


Fig. 5. Distribution plot of new buildings by year of construction according to their (Euclidean) distance in meters to the city center and neighborhood per capita income in 2019.

Note: weighted by number of dwellings. Own elaboration using data from the Dirección General del Catastro: Ministerio de Hacienda. Government of Spain. <http://www.sedecatastro.gob.es/>. Income level is per capita gross income in 2019.

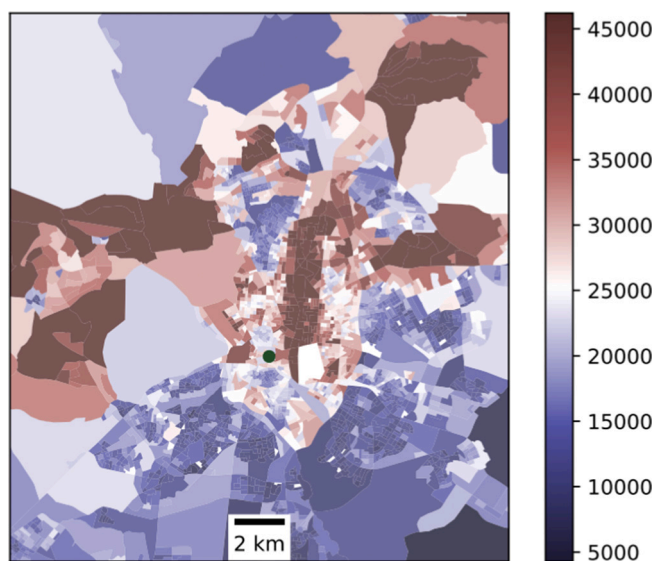


Fig. 6. Per capita gross income at census tract level. 2019. The green point indicates the center of the city (Puerta del Sol). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Source: own elaboration using INE data.

Although the data show that it is unlikely to find high-income individuals at distances >20 km, both low-income and high-income individuals can be found in areas close to the city center.⁵ Under any of the three distance measures considered, we find, coherently with the standard monocentric model, that individuals with higher incomes tend to be located closer to the CBD/city center. However, we also find a high

concentration of low-income individuals at relatively short distances. Although the consideration of alternative measures of distance reduces the theoretical commuting distance of higher-income individuals, it does not shed light on the anomaly regarding the location of low-income neighborhoods.

The apparent anomaly in the location of individuals in the Madrid region, with low-income agents near the city center, may be partially explained by two factors. Firstly, the “age” of these low-income urban areas. Given that the buildings are generally <60 years old (most were built between 1960 and 1980) and would therefore still be in an acceptable state of habitability, many are small and sufficiently old to deter high-income individuals (Brueckner & Rosenthal, 2009). Second, the large flow of immigrants from disadvantaged economic areas may have replaced former residents. These two factors could have served to slow down a possible process of building renewal and gentrification in disadvantaged neighborhoods in good locations near the city center. The original characteristics of neighborhoods will last for long periods of time, and urban laws can also limit the speed of the adjustment by protecting buildings and limiting neighborhood renewal. The localization of existing urban transport facilities and transport infrastructures (i. e., railways, high capacity roads, airports) may also reduce the attractiveness of some city locations, as the cost of living near these infrastructures outweighs the benefits of being close to the city center. These low-income neighborhoods were originally built in the post-Civil War period (1940s) in the city outskirts at that time and are located near transport infrastructures such as the M-30 beltway and the Abonigal freight terminal, which act as a de facto border for low-income neighborhoods in the south and east areas near the city center (Fig. 8).

More interesting is the peculiar configuration in the spatial distribution of the new neighborhoods occupied by high-income people. The stratification resulting from the urban evolution and population inflows is shown in Fig. 6 above, but the dynamics are missing. A simple and convenient way of including dynamics can be achieved by simultaneously considering data on income by neighborhoods (census tract) and the date of construction (mode of construction dates of buildings in a census tract) of the buildings. In this case, only the neighborhoods with extreme per capita income values (high per capita income $>€40$ k euros; low per capita income $<€10$ k) are considered. This analysis allows us to observe a clear tendency towards the formation of clusters (Fig. 9) such that recent neighborhoods occupied by people with extreme incomes are close to other older neighborhoods occupied by people with extreme income levels. Given that the process of urban growth in Madrid is relatively recent, the oldest neighborhoods (both high- and low-income

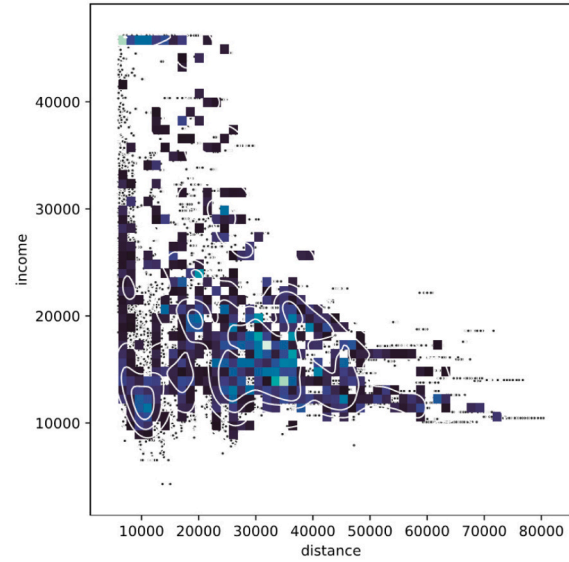
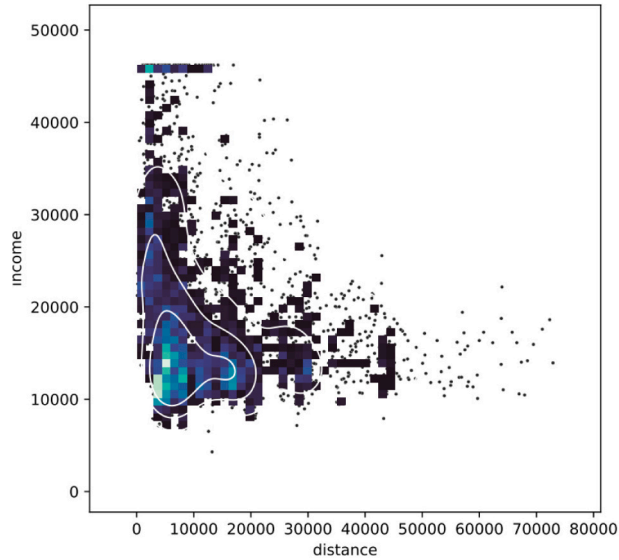
⁵ The location of the buildings is taken from the registry of real estate properties and taxation data (Dirección General del Catastro), and each building has been assigned the per capita income data (2019) at census tract level. Based on the INE mobility study using cell phone positioning data, the CBD is defined as an area with a population of $>10,000$ inhabitants in which the balance between inflows and outflows is $>50\%$ of the resident population and net entries of at least 5000 people. Sparsely populated areas are excluded to prevent them from being classified as CBD despite having insignificant movements of people. See Appendix B.

a) Distance (mts.) to inner center
(Puerta del Sol)

$$b) d_i^a = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$$

c) Mean distance (mts.) to CBD
(21 “attraction areas” Annex B)

$$d_i^b = \frac{1}{21} \sum_{j=1}^{21} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$



d) Minimum distance (mts.) to CBD.

$$d_i^c = \min \left\{ d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \right\}$$

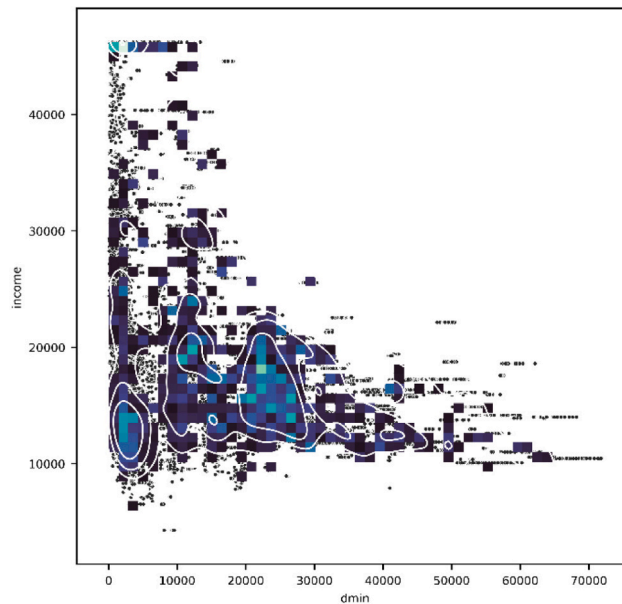


Fig. 7. Spatial distribution of per capita income (2019) and distance to CBD.

Source: Own elaboration using data from the Dirección General del Catastro and INE. Lighter colors indicate a higher concentration. The first one is the simple Euclidean distance between the centroid of each building and the center of the city (Puerta del Sol). The second is the arithmetic mean of the Euclidean distances between each building and each of the centroids of the attraction areas (a total of 21 areas). The latter is simply the minimum distance between the location of the building and each of the 21 attraction areas.

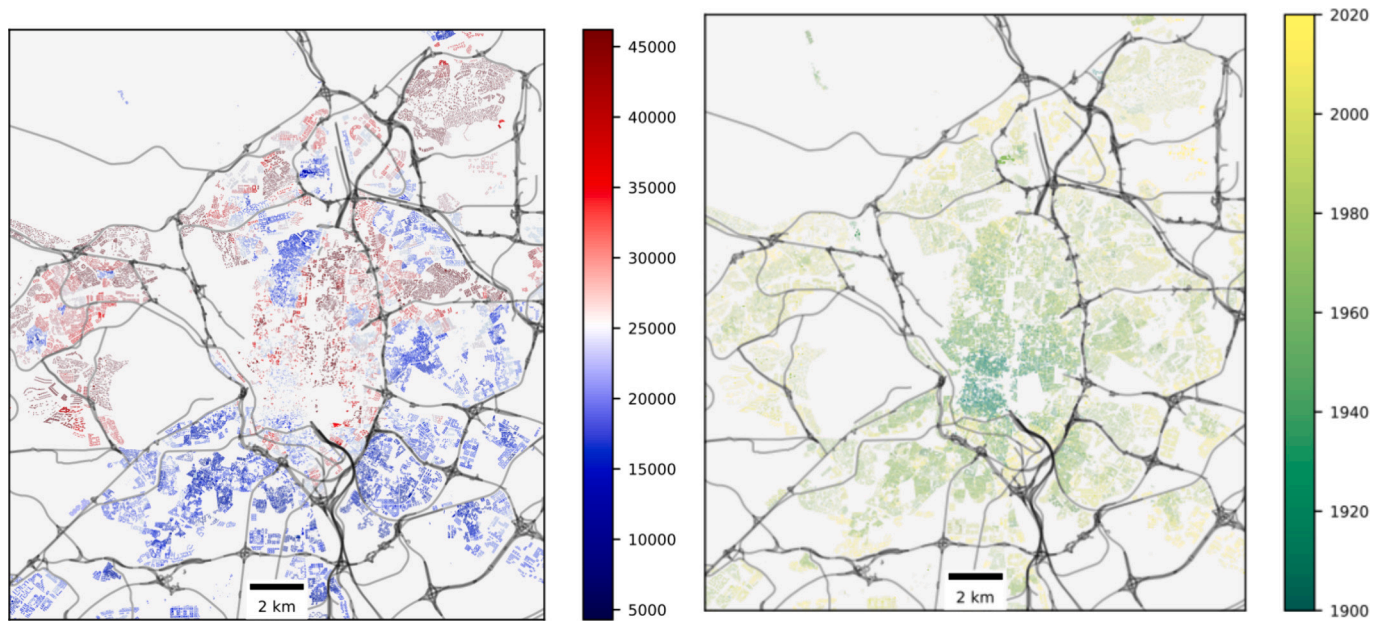


Fig. 8. Left. Per capita income (2019) and transport network. Right: Building construction year and transport network (grey lines: rail and highway networks). Source: own elaboration using Dirección General del Catastro and INE data. Rail networks exclude the metro network and the subterranean rail network. Only routes with a capacity greater than two-lane highways are shown.

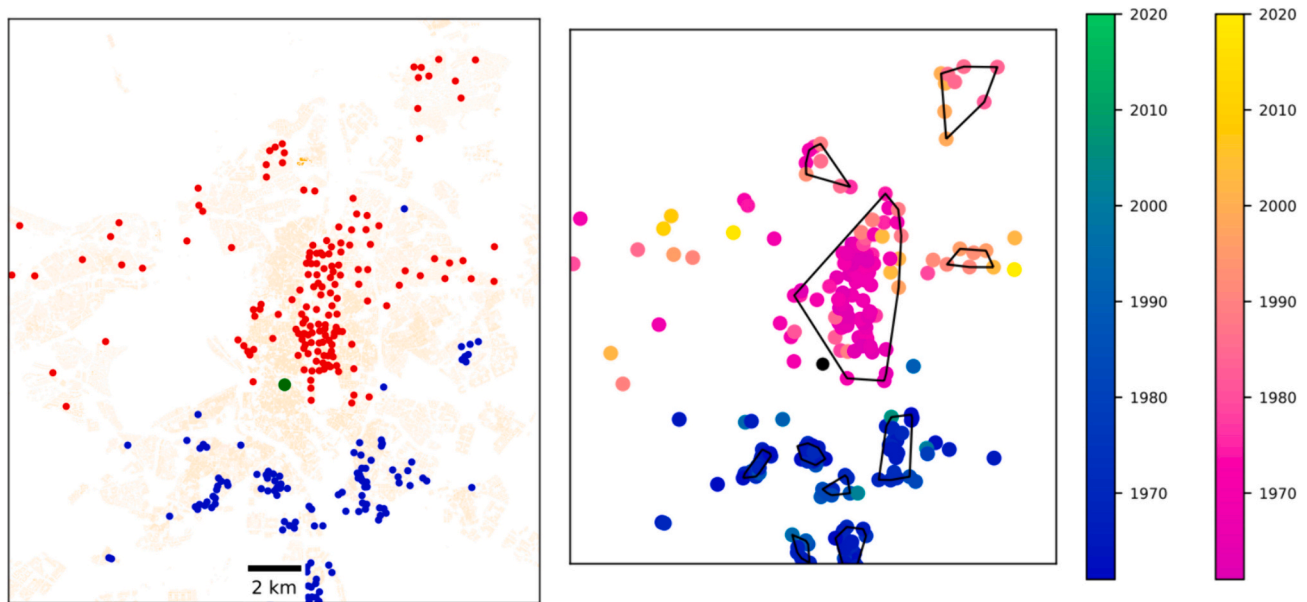


Fig. 9. Left graph- per capita gross income (2019) at census tract level (red points: per capita income \geq €40 k, blue points: per capita income $<$ €10 k). Right graph: construction date and per capita gross income (2019) at census tract level. The black point indicates the city center (Puerta del Sol). Pink-yellow points if per capita income \geq €40 k euros. Blue points if per capita income $<$ €10 k. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Source: own elaboration. Polygons represent spatial clusters detected with the HDBScan algorithm with a minimum cluster size of 5 units.

ones) can be found at a shorter distance from the city center but separated by natural (rivers and forest) and artificial barriers (airport, railways, beltways, and high-capacity routes).

High-income neighborhoods seem to agglomerate in the north-west direction, while in the case of low-income areas no significant changes can be observed since the construction date of 1970. After that date, the Spanish land use legislation and social housing policy has clearly stood against urban segregation and promoted social cohesion and spatial social mix (Ponce, 2010) (Ponce, 2019). The fact that most buildings in low-income areas were constructed prior to 1970 could explain the

relative proximity of these clusters to the city center, given that those years were characterized by high migratory flows from rural areas and housing shortage problems. After the Spanish Civil War, the government attempted to intervene to mitigate the scarcity of housing (Sambricio, 1999) (Valenzuela-Rubio, 1974) through several housing plans such as the General Urban Plan of Madrid (Bidagor Plan) in 1941 or the Social Housing Plan of 1959. These plans were conceived based on the idea of developing a set of satellite nuclei occupied by “workers” around the municipality of Madrid in 1940 who had to work in nearby industrial areas. However, they did not prevent disorderly construction by the

migrants themselves or by private real estate developers in areas that were relatively close to the urban center in those years. As pointed out by (Sambricio, 1999), the Madrid General Urban Plan of 1941 conceived urban planning based on the concept of social hierarchy, with space for a “bourgeois” area and a set of “working class” nuclei. Extreme examples are the dwellings for members of the army built in 1958 by the army housing services (*Patronato de Casas Militares Jefes y Oficiales*) in the northern area of Madrid with a floor area of $>100 \text{ m}^2$ for officers and “social dwellings” built in the southeast area for workers with a floor area of $<60 \text{ m}^2$.

The high intensity of dwellings per area (with many dwellings with a floor area in the range of $38\text{--}75 \text{ m}^2$) and the rate of construction of new dwellings (Fig. 1) at that time point to intense migratory flows. As will be discussed in the next section, the low-income neighborhoods in the southern part of the city may be an explanation for the growth of the new high-income neighborhoods in the exact opposite direction. The construction of large administrative centers of the central government in the north direction from the city center (*Nuevos Ministerios*, 1941, *Complejo Cuzco*, 1971) and the existence and development of golf courses in the west and north (*Real Club Puerta de Hierro*, 1904, *Club de Campo Villa de Madrid*, 1931, *Real Club La Moraleja*, 1976) and other facilities usually associated with the wealthier social classes (*Hipódromo de la Zarzuela*, 1878) may also have contributed to the development of the wealthier neighborhoods in this north west direction due to their impact on residential values as found in other works such as in (Yates & Cowart, 2019).

4. Spatial equilibrium model with locational externalities

The urban growth pattern observed in Madrid (Fig. 4) shows an initial expansion close to the city center after the Spanish Civil War (1939) which continued in later years with new neighborhoods in both distant places and in areas near the city center. As a result of this process a clear income stratification by distance to the city center/CBD emerged. High-income neighborhoods appear to be agglomerating around the north-west axis, while the low-income neighborhoods have changed very little since the 1970s.

This section presents a simple spatial equilibrium model largely inspired in (Guerrieri et al., 2013) where city size and stratification are jointly analyzed. In this model, a monopolistic absentee landlord owns all the available land. The land value for non-residential use is equal to zero but can be intended for residential use at no cost. An absentee builder-landlord decides the rental value $R(x)$ of housing in location x .

Agents live and work in a linear urban area composed of a discrete set of locations X with arbitrary origin, $x = 0$, where the urban business district (CBD) is located. Each location $x \in X$ is a residential area of normalized size 1 at a distance x from the CBD. All jobs are in the CBD and agents' commutes are costly.

For the sake of simplicity, we will consider jobs of two types/wages: low L and high H wages. Competitive firms in the CBD employ labor as the only production factor producing a consumption good with price normalized to 1. Productivity is constant among workers of the same type, but workers with a higher wage H are more productive than low-wage workers L . Commuting costs differ by type of worker, $r_H > r_L$. As in (Su, 2022), we assume that the opportunity cost of commuting is higher for high-wage workers due to the higher value of their time. As a particular assumption, we assume that commuting costs increase with wage more than proportionally, $\frac{r_H}{w_H} > \frac{r_L}{w_L}$.

Agent utility is $U_i(x) = U_i(c, G(x))$, $i = H, L$, where c is the consumption good and $G(x)$ is a locational “reward” that depends on location x . We assume that, from the point of view of the agent, $G(x)$ is exogenously given, so $G(x)$ is not affected by the agent's housing location choice. For the purposes of this paper, as in (Guerrieri et al., 2013), the value of $G(x)$ is related to the characteristics of the neighborhood of x . Specifically, we assume that $G(x)$ depends on the number of H workers

that reside between the city limits and x . As shown below, this assumption will be very convenient to make the model easily manageable.

The agent's resource constraint will satisfy,

$$c + R(x) + r_i x \leq w_i, i = H, L, \quad (1)$$

For convenience sake, we assume a linear separable utility function specified as $U(c, G(x)) = c + \beta G(x)$, $\beta > 0$ conditional to location and externality utility given by.

$$U_i(x, w_i, r_i) = w_i - r_i x - R(x) + \beta G(x), i = H, L \quad (2)$$

The agent's utility will depend on commuting distance, wage level, locational externality, and rental prices. As the landlord enjoys monopolistic power, she can fix rental prices until no consumer surplus remains (over reservation utility \bar{U}_i), and in equilibria agents will be indifferent among distinct locations in the city. Therefore, in equilibrium the agent's utility will be the same among locations for each agent type, $U_i(x, w_i, r_i) = \bar{U}_i$ for all x . This mechanism leads to a similar result as the agent's perfect mobility assumption in (Zenou, 2009) that is behind the bid-rent relationship (Eq. (3)) where agents are indifferent between all locations x . Agent willingness to pay (bid-rent) for location x is given by,

$$R_i(x) = w_i - r_i x + \beta G(x) - \bar{U}_i \quad (3)$$

where reservation utility \bar{U}_i will be set to zero for the sake of simplicity. This equation highlights the effect of the locational externality $G(x)$ on willingness to pay $R_i(x)$ once distance to the CBD is controlled.

The equilibrium in an urban area X composed of discrete locations at a distance x from the CBD is a set of rental prices $R(x)$ and neighborhood externality $G(x)$ for $x \in X$, for which:

- Residents maximize their utility and in equilibria no one wants to change their location.
- Absentee monopolistic landlords maximize their total profit by renting each location x in the urban area at $R(x)$.

We assume that rental contracts only last for one period and that constructing or demolishing building is costless and immediate. Let us assume that in each period t the landlord will rent the set of locations available in the urban area. Let us also assume that the landlord has perfect knowledge about the wages and commuting costs in the area, but not about the exact number of people that want to reside there. The landlord only knows the probability that a given location will be rented by H workers or L workers which is related to the number of people wanting to reside in the city, that is, the incoming and outgoing flows of workers from and to other cities. This probability can be understood as the degree of attractiveness of the urban area compared to other urban areas.

Under uncertainty about the arrival and exit rate of workers to and from the urban area (due to worker's imperfect information about wages and commuting costs in other places or about the locational externality), it is reasonable to think that the urban configuration will change. If the probability that L workers will move to the urban area is higher than in the case of H workers, a monopolistic landlord should assess if such a difference makes it profitable to reconfigure the urban area to increase the space rented to L workers provided a given location x is not occupied. In this case, it is reasonable to think that the landlord can afford an opportunity cost, c .

If the probability that a location x will be rented is μ_H for H workers and μ_L for L workers, the landlord will pay cost c with probability $(1 - \mu_H)$ if she assigns x to H workers (setting a rental price for location x that low-income agents cannot afford), and with probability $(1 - \mu_L)$ if x is assigned to L workers. This cost should be understood as the cost to the landlord of wrongly sizing the urban area.

In this model setting, the resulting urban configuration is, to some extent, like the leapfrog urban configuration in (Markusen and Scheffman, 1978) where there may be unoccupied land inside the city area. In

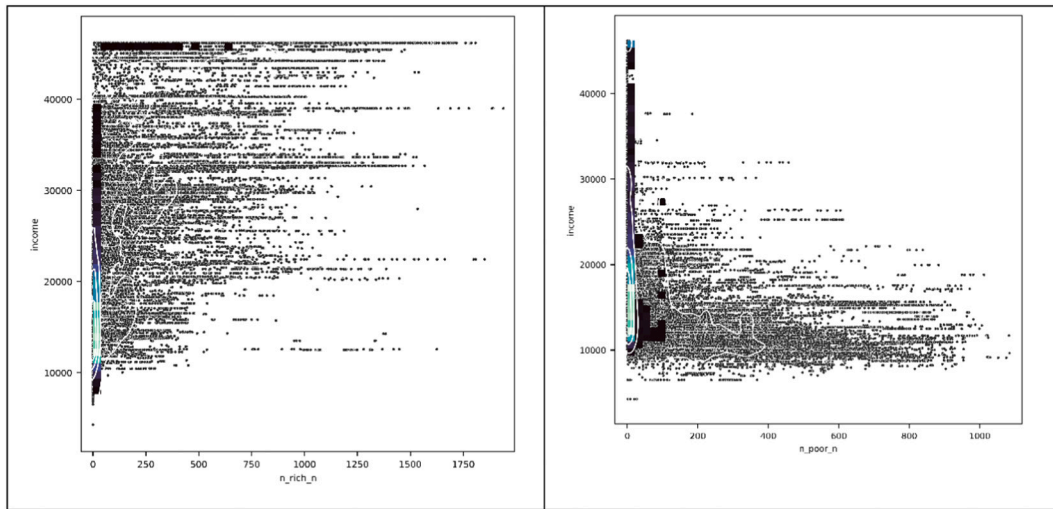


Fig. 10. Distribution plot of number of rich (left) and poor (right) buildings in an area with a radius of 1500 m and per capita income (2019). Source: own elaboration. Number of buildings of each type in a radius of 1500 m.

the model considered here, “gaps” are random, as some locations will be not occupied with probabilities $(1 - \mu_H)$ in the case of H workers' neighborhoods and $(1 - \mu_L)$, in the case of L workers' neighborhood.

We will denote the city size as x_M and the size of the H neighborhood by x^* . As is immediate, the size of L neighborhood will be equal to $x_M - x^*$. In the equilibrium (see Appendix) defined above, as in the Alonso–Muth–Mills monocentric urban equilibria, the city is stratified by income since the H workers will be located near the CBD and the L workers will be located in the outskirts, with $R(x)$ decreasing in x . Wages w_H, w_L and commuting costs w_H, w_L, r_H, r_L will determine the city size and the size x^* of the H neighborhood.

The locational externality,

$$G(x) = G(x) = \begin{cases} 0 & x > x^* \\ (x^* - x)\mu_H & x \leq x^* \end{cases} \quad (4)$$

will be decreasing with distance, thus satisfying $G(0) = x^* \mu_H$. In other words, the higher the number of H workers in the city, the higher the value of the locational externality.

The spatial equilibrium leads to a city size given by

$$x_M = \frac{w_L - c \frac{(1-\mu_L)}{\mu_L}}{r_L} \quad (5)$$

As regards the city area occupied by H workers, x^{**} , in equilibrium we will have,

$$x^* = \frac{w_H - gw_L + c(1-g)}{r_H - gr_L} \quad (6)$$

where $g = \frac{\mu_L}{\mu_H}$.

When there are no uncertainty, city size will be given by $x_M = \frac{w_L}{r_L}$. High-wage workers will be located at shorter distances than $x^* = \frac{w_H - w_L}{r_H - r_L}$, while low wage workers will be located at longer distances but inside the city limits. When $g = 1$, there are no differences with respect to the non-uncertainty version of the model.

As a first interesting result, we can conclude that city size is affected by the probability μ_L . In particular, the city size is smaller when μ_L is low, or in other words, when the landlord believes that the inflow of low-wage workers is low. Conversely, an increase in the μ_H value makes x^* higher, since the landlord will reduce the urban space to L workers due to the relatively higher rate of arrival of H workers. An increase in the μ_H value also affects the value of the locational externality, $G(x)$, in the locations.

As an interesting result, if μ_H increased when μ_L remains unchanged the size of the urban area is not affected but the space assigned to H workers increases with a higher value of x^* . This is similar to a gentrification process affecting the urban space bordering H and L workers in the original spatial equilibrium with a lower μ_H value.

5. Testing the effect of the neighborhood externality

In the theoretical framework analyzed above, the size of the urban area will increase, and areas more distant from the city center will be occupied, when there is a strong entry of low-skilled/low income workers, that is, a high probability of occupation if the location x is assigned to L workers. Since the 1970s, along with lower transportation costs due to improvements in the transportation network and the increased use of private vehicles, a similar process has occurred in Madrid as the city's suburbs and surrounding municipalities have grown. These developments took place given the existing urban configuration inherited from the post-Civil War period, but without any natural geographic landforms, lakes, or rivers that might limit the expansion of the urban area. More interesting is how the socio-spatial stratification has evolved. As discussed in the theoretical analysis, individuals' willingness to pay depends not only on their proximity to the center or CBD, but also on the characteristics of the location, $G(x)$. One of the characteristics analyzed in the theoretical framework is the type of neighborhoods that surround a given location, which determines the value of $G(x)$ and therefore the willingness to pay to occupy a given location. In the model, the neighborhood characteristics, $G(x)$, depend on both the number of H workers and L workers, since the number of H workers increases the value of $G(x)$, while L workers will decrease the value, with a minimum of $G(x_M) = 0$.

Under this specification, the value of $G(x)$ depends on the number of type H and type L agents surroundings location x . This simple result allows us to consider an indirect method to approximate the value of $G(x)$ by considering the number of each type of agent residing near a given location. We have proceeded first by assigning each residential building (a total of 403,813 buildings) an income level equal to that of the census tract to which the building is assigned (INE data corresponding to 2019) and then counting the number of high-income buildings (per capita income ≥ 40 K euros) and low-income buildings (per capita income < 10 K euros) in an area with a radius of 1.5 km (the average size of the census tracts in the region of Madrid is 1.9 km^2 , although they are considerably smaller in some urban areas). The idea behind this procedure is to check whether the composition of a given neighborhood favors the occupation

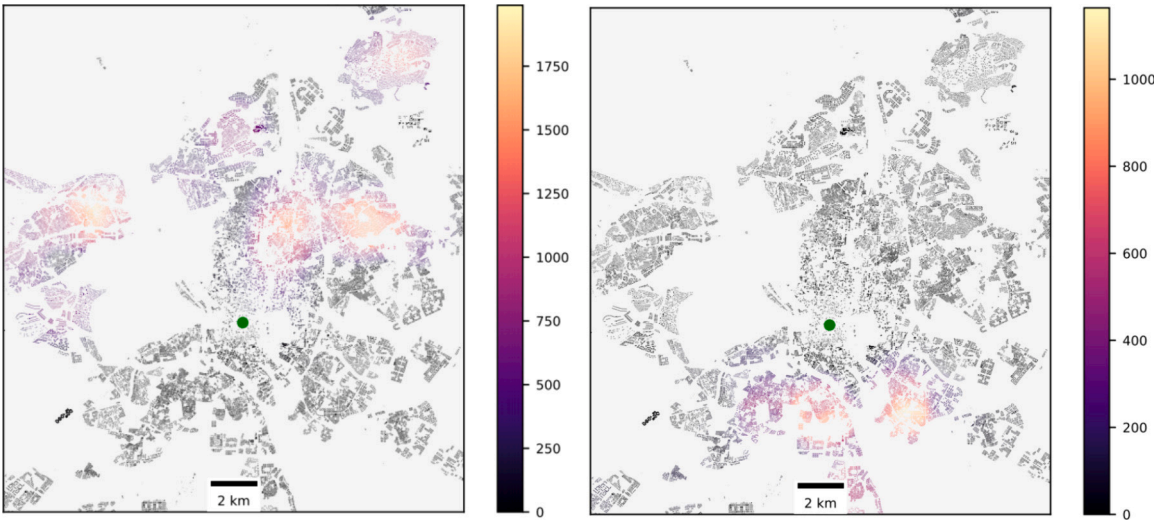


Fig. 11. City center (green point) and spatial distribution of the number of rich (left) and poor (right) buildings in a radius of 1500 m. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
Source: own elaboration.

Table 2
Estimation of neighborhood effects.

	OLS	P-value	Robust least squares	P-value	Generalized linear model	P-value
	Parameter		Parameter		Parameter	
Const	23,328.2 (150.7395)	0.0000	19,478.92 (0.7075)	0.0000	10.0409 (0.004)	0.0000
Pop. density	−187,710.7 (3548.673)	0.0000	−138,587.4 (23.1655)	0.0000	−9.2746 (0.132)	0.0000
Building surface/n° dwellings	3.058883 (0.402554)	0.0000	12.59917 (0.000964)	0.0000	0.0003 (0.00000554)	0.0000
Distance (minimum to CBD)	−0.242675 (0.00347)	0.0000	−0.195479 (0.0000249)	0.0000	−0.00001324 (0.000000142)	0.0000
N° rich buildings in 1500 m. radius	16.5282 (0.289822)	0.0000	17.58315 (0.001136)	0.0000	0.0006 (0.00000649)	0.0000
N° poor buildings in 1500 m. radius	−6.59442 (0.216056)	0.0000	−5.230237 (0.002011)	0.0000	−0.0005 (0.0000115)	0.0000
Cov type:	HAC		Huber Type I Standard Errors & Covariance		HAC	
GLS model family					Gamma (log)	

HAC: heteroskedasticity and autocorrelation consistent (Newey West automatic bandwidth).
Note: the dependent variable is the mean per capita income at building level. A total of 404,309 residential units are considered. The income assigned to each building is the per capita income in year 2019 at census tract level. Population density is also measured at census tract level. Data on buildings refer to 2020. Similar results are obtained if distance (minimum distance to CBD) is replaced by distance to the city center (Puerta del Sol).

of adjacent areas by individuals with similar income levels. The result of this analysis is shown in Fig. 10. As can be seen in the figure, most observations correspond to per capita incomes in the 10 k–25 k range and are located at low values for both the number of rich “neighbors” and for the number of poor “neighbors.” At the extremes of income, the expected stratification is clearly observed. Poor neighbors are surrounded by poor neighbors and rich neighbors by rich neighbors in very specific areas of the urban area of Madrid: a rich area in the north and a poor area in the south (Fig. 11).

Since high rental price locations are assigned to higher income individuals in equilibrium, income can be used as a proxy for the cost of accommodation and an expression such as

$$R_i = \mu + x_i\beta + \theta_1 dmin_i + \theta_2 n_rich_i + \theta_3 n_poor_i + \varepsilon_i$$

where R_i is the rental value of location i proxied by the income assigned to location i (all the buildings in the same census tract will be assigned the same per capita income), $dmin$ is the minimum distance to the CBD (attraction areas), n_rich and n_poor are the number of “rich” buildings and “poor” buildings, respectively, in a 1500 m radius of location i and x_i

are other variables affecting rental value (floor area of dwelling and population density of the census tract to which location i belongs).

To check the robustness of the results, we test a slightly different specification where we test how per capita income in each location depends on the average distance to the neighborhoods⁶ with the highest value of the locational externality (places belonging to the 95 percentile of the n_rich variable) and on the average distance to the neighborhoods with the lowest value of the locational externality (places belonging to the 95 percentile of the n_poor variable). The results in Table 3 show a negative effect on income of distance to neighborhoods with a higher concentration of wealthy individuals, where the value of the location externality is higher. At the same time, the estimated effect of distance to

⁶ Buildings in the 0.95 percentile are grouped by the HBDSCAN algorithm with a minimum cluster size of 50. This gives a total of 9 clusters in the case of affluent neighborhoods (those with the highest values of the variable n_rich) and 8 clusters for the poorest neighborhoods (highest values of the variable n_poor). We compute the average distance of each building to each centroid once the centroids of the clusters for rich and poor neighborhoods are determined.

Table 3
Effect of distance to neighborhood type on income.

	OLS	P-value	Robust least squares	P-value	Generalized linear model	P-value
	Parameter		Parameter		Parameter	
Const	29,609.51 (258.2355)	0.0000	21,355.04 (0.916179)	0.0000	10.37151 (0.013197)	0.0000
Pop. density	-260,197.8 (4582.21)	0.0000	-155,597.7 (23.115)	0.0000	-13.16389 (0.224724)	0.0000
Building surface/n° dwellings	4.466138 (0.572139)	0.0000	16.36831 (0.00099)	0.0000	0.000465 (0.0000263)	0.0000
Distance to centroid of the area with highest n _{rich}	-0.488878 (0.005969)	0.0000	-0.321608 (0.00003)	0.0000	-0.0000285 (0.000000329)	0.0000
Distance to centroid of the area with highest n _{poor}	0.125108 (0.002908)	0.0000	0.096942 (0.0000238)	0.0000	0.00000585 (0.000000167)	0.0000
Cov type: GLS model family	HAC		Huber Type I Standard Errors & Covariance		HAC Gamma (Log)	

Note: the dependent variable is the mean per capita income at building level. The income assigned to each building is the per capita income in year 2019 at census tract level.

Table 4
Income variations and neighborhood composition.

	OLS (**)	P-value	GMM (*)	P-value
	Parameter		Parameter	
Const	875.1896	0.00000	390.3208	0.00000
Income year 2015	0.113185	0.00000	0.123865	0.00000
Var. in N° poor buildings in 1500 m. radius	-0.761532	0.00010	-1.072249	0.00000
Var in N° rich buildings in 1500 m. radius	1.805071	0.00000	1.326099	0.00000
Cov type:	HAC		HAC	

HAC: heteroskedasticity and autocorrelation consistent (Newey West automatic bandwidth).

The dependent variable is the variation (2019–2015) in the average per capita income in the census section where the residential unit (building or set of buildings) is located. A total of 404,306 residential units are considered.

(*) Instruments: Population density at census level, Average surface per dwelling, Income in 2015, and variations in the number of rich and poor in a 1500 m. radius.

(**) Included dummies for construction year.

poor neighborhoods on income is positive. These results seem to confirm the coexistence of a positive externality associated with proximity to high-income neighborhoods and a negative externality associated with proximity to low-income neighborhoods. Although low-income neighborhoods are not particularly disadvantaged in terms of accessibility to the CBD, the location anomaly in the Madrid case mentioned above, they have a sort of negative effect on the residential choices of middle- and high-income individuals.

Based on the simple model estimated in Table 2, it is possible to propose a retrospective exercise in which the values of the variables are again quantified considering the date of construction of the buildings t , which would reflect a situation similar to that existing on date t . In this way, instead of the variable n_{rich_i} we will consider a variable $n_{rich_{it}}$ that would represent the number of buildings with high-income individuals (according to income criteria for the year 2019) that surround (within a radius of 1500 m) building i built in period t , only considering buildings built in period t or earlier. Similarly, $n_{poor_{it}}$ would be defined for the case of low-income individuals.

With these new variables we can determine which areas exhibit greater variations in the location externality because of the arrival of new neighbors (new buildings) and whether these arrivals entail a greater number of rich neighbors or a greater number of poor neighbors (i.e., neighborhood composition in 2019 vs. neighborhood composition in the year the building was constructed). These new variables allow us to check (see Table 4) if these arrivals are related to variations in rental prices proxied by per capita income of these areas over the period

2015–2019. As the theoretical model predicts, the higher the inflow of H workers, the higher the locational externality and hence rental prices proxied by income.

Although the period considered is relatively short (there are no previous data with the same level of disaggregation), it shows how changes in neighborhood characteristics affect rental prices. In general, areas with the highest growth in rental prices (per capita income) are also those in which there has historically been a greater concentration of high-income neighbors. This effect is maintained even after controlling by level of income in 2015. Therefore, the results seem to indicate that variations in the level of income not only depend on the starting point, but there is also a positive locational effect linked to the place where individuals reside. Not only do wealthy neighborhoods attract wealthier people, but the mean income of wealthy areas increases to a greater extent as the number of wealthier neighbors increases, thus reinforcing urban stratification.

6. Concluding remarks

Socio-spatial stratification is a common feature of large cities. Neighborhoods of extreme income, where high- and low-income people live, are easily identifiable as clusters in the urban structure. The configuration of these spatial clusters is not only a consequence of income inequality, but is also influenced by the preferences and choices of residents. High-income individuals tend to concentrate in spatial clusters within affluent neighborhoods, reflecting a propensity to live near others with similar income levels. This clustering of high-income residents contributes to the distinctive socio-spatial stratification observed in urban environments. Notably, this phenomenon is not limited to high-income households. Low-income individuals also show spatial grouping within certain neighborhoods, not because of a desired outcome, but rather because they have a limited range of options.

Theoretical models based on the classical Alonso-Mills-Muth model can explain the observed stratification in cities. However, urban areas are fundamentally dynamic, and these models are entirely static. Urban areas expand through sprawl processes, but the intensity of land use can also change within existing neighborhoods. As an urban area expands and more people move in and out of it, the spatial structure of the urban area can change over time. How these changes affect the socio-spatial stratification in the metropolitan area is the question we are interested in. In this article, we explore the characteristics of the expansion process of the urban area of Madrid, and how this process affects its socio-spatial stratification. The significant variation in population during the 1960s can be explained by internal migration flows from rural areas to industrial centers; however, the baby boom was a major factor in the increase in population and urban size of the 1970s. During the 2000s, both the Madrid region and the country as a whole experienced the most

significant population growth in the last five decades. This demographic expansion is due to the significant entry of foreign migrants. Madrid has seen an important inflow of low-skilled labor, along with an important inflow of high-skilled labor. This concentration of highly skilled workers is due to a “headquarters” effect, as Madrid is Spain's administrative capital and the city where Spain's largest companies are based, but also to the positive effect such a move can have on their careers.

The urban growth pattern observed in the Madrid metropolitan area shows an initial expansion in areas close to the original Madrid urban area, but that in recent years has been observed in both distant and closer areas. The replacement of old buildings close to the city center, as well as the filling of gaps in the expansion process itself, may be behind this process. The Madrid metropolitan area shows a clear socio-spatial stratification in terms of income distribution. Households with higher incomes tend to be located in the northern and western sectors of the Madrid metropolitan area, while households with the lowest incomes are concentrated in the southeastern and outermost regions of the metropolitan area. The data show that both low-income and high-income individuals can be found in areas close to the city center, which is not fully consistent with the predictions of the monocentric Alonso-Mills-Muth model. This apparent anomaly can be partially explained by the “age” of these low-income urban areas, since the buildings are generally <60 years old, and by the large flow of immigrants since 2000, which has been particularly concentrated in these areas. In addition, the proximity of these areas to existing urban transport infrastructures (i.e. railways, high-capacity roads, airports) means that the disadvantages outweigh the advantages of being close to the city center. More interesting is the peculiar configuration in the spatial distribution of the newer neighborhoods occupied by people with high income levels, which are close to other older neighborhoods occupied by people with high income levels.

This article presents a urban equilibrium model with commuting costs and a neighborhood composition externality, in which the spatial configuration of the city is explained by the different arrival intensities of residents, which affect both the size of the urban area and the stratification of the city. In this model, the composition of neighborhoods in terms of income is important because it drives a composition/location externality for which individuals are willing to pay higher housing prices. The location externality affects how social stratification evolves through attraction (homophily) and repulsion (heterophobia) effects. New rich neighborhoods are located close to existing rich neighborhoods and far from poor neighborhoods. This attraction-repulsion effect affects the entire urban area, so that the spatial heterogeneity of income in the city can be modeled using a simple model in which the composition effect, together with the distance to the CBD, influences the location decision of agents. Income is positively affected in

neighborhoods surrounded by other rich neighborhoods, and negatively affected if poor neighbors reside in the surrounding neighborhood.

Gentrification processes can reduce the amount of urban space occupied by low-income people if high-income people are more likely to move into the city (e.g., due to the growth of sectors that require high-skilled jobs). The replacement of low-income residents by wealthier residents increases the value of the locational externality in previously wealthy neighborhoods and reinforces urban stratification. This effect can be observed in the case of Madrid, as the areas that have experienced the highest growth in per capita income are those that have historically had a higher concentration of high-income neighbors. This simple model explains, with a considerable degree of precision, the observed variation in the spatial distribution of income in the urban area based on the variation in the number of rich and poor neighbors in the vicinity. These forces of attraction and repulsion could explain why certain areas experience gentrification and others do not, simply because of their location and proximity to wealthy neighborhoods.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Julián Moral-Carcedo: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

Author declares no potential conflicts of interest.

Data availability

Datasets are publicly available:

Spanish National Statistical Institute (INE)

Income data: https://www.ine.es/experimental/atlas/experimental_atlas.htm.

Residential variations: https://www.ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736177013&menu=resultados&idp=1254734710990.

Mobility data can be found in https://www.ine.es/experimental/movilidad/experimental_em.htm

Buildings dataset, Dirección general de Catastro: Ministerio de Hacienda. Spanish government. <http://www.sedecatastro.gob.es/>.

Appendix A. Spatial equilibrium

Spatial equilibrium without uncertainty: $\mu_H = 1$, $\mu_L = 1$

As it is assumed that housing demand is set to 1, distances x_M (city size) and x^* (length of the H neighborhood) will determine the population of each type that will live in the urban area. In a linear city, maximum population will be given by area x_M where $N_H^{max} = x^*$ will be the maximum number of H workers, and $N_L^{max} = x_M - x^*$ the maximum number of L workers.

The bid-rent curves of both agents are decreasing with distance x , moreover following Eq. (3), $R_i(x) = w_i - r_i x + \beta G(x) - \bar{U}_i$, $i = H, L$. As the bid rent of H workers is steeper due to their higher commuting costs⁷ bid rent curves will cross. Assuming land value for other uses equal to zero, city limits x_M are determined by the maximum distance for which rental prices are equal or higher than zero. Taking bid rent for L workers and imposing $\bar{U}_i = 0$, $G(x_M) = 0$, we get city size from condition $R_L(x_M) = 0$, that is $x_M = \frac{w_L}{r_L}$. It is easy to see from this equation that city size will be higher when wages are higher and commuting costs are lower. Note that we are setting $G(x_M) = 0$ as by definition no agent of any type will be located at distances higher than x_M .

⁷ Both bid-rent curves will intersect in location $x' = \frac{w_H - w_L}{r_H - r_L}$.

Monopolistic landlord will assign the set of x locations to the agent with the highest willingness to pay, given by the bid rent in Eq. (3). Therefore, in equilibrium rental prices will be,

$$R(x) = \max[R_H(x), R_L(x)] \quad (\text{A.1})$$

Denoting as x^* the location where $R_H(x^*) = R_L(x^*)$ it can be concluded that $x^* = \frac{w_H - w_L}{r_H - r_L}$ and therefore,

$$R(x) = \begin{cases} w_H - r_H x + \beta G(x); & 0 \leq x \leq \frac{w_H - w_L}{r_H - r_L} \\ w_L - r_L x + \beta G(x); & \frac{w_H - w_L}{r_H - r_L} < x \leq \frac{w_L}{r_L} \end{cases} \quad (\text{A.2})$$

To close the model $G(x)$ should be determined. As the neighborhood externality is linked to the number of H workers in the environment of location x , the role of landlord establishing rental prices is fundamental in order to understand the configuration of the urban area.

From the landlord point of view that seeks to maximize her profits, as land and building costs are zero, her profit renting the locations set $X = [0, x_M]$ will be given by

$$V(x, G(x)) = \sum_{x=0}^{x_M} R(x, G(x)). \quad (\text{A.3})$$

Introducing the indicator variable u_x that takes value 1 if location x is rented at $R^H(x, G(x))$ prices and value 0 if x is rented at $R^L(x, G(x))$ prices, the landlord problem can be reformulated as determining u_x in order to maximize

$$V(x, G(x)) = \max_{u_x \in \{0, 1\}} \sum_{x=0}^{x_M} [R^H(x, G(x))u_x + R^L(x, G(x))(1 - u_x)] \quad (\text{A.4})$$

As $G(x)$ is related to the number of H workers between the city limits and x , we will have $G(x_M) = 0$. As locations are discrete of size equal to one, $G(x)$ will evolve as

$$G(x) = G(x-1) - u_x; 0 < x \leq x_M - 1 \quad (\text{A.5})$$

The transition equation in (A.5) allows to reformulate the maximization of Eq. (A.4) in terms of the Bellman equation as

$$V(x, G(x)) = \max_{u_x \in \{0, 1\}} [R^H(x, G(x))u_x + R^L(x, G(x))(1 - u_x)] + V(x+1, G(x+1)) \quad (\text{A.6})$$

for x in $[0, x_M]$, with the transition equation, $G(x+1) = G(x) - u_x$

As the set of x locations is discrete in the range $[0, x_M]$, the Bellman equation is well defined. Assuming a terminal value $V(x_{M+1}, G(x_{M+1})) = 0$, and $G(x_M) = 0$, the values u_x , and $G(x)$ for $x < x_M$ may be determined recursively by repeated application of Bellman's equation starting at final x_M .

When $x_M = \frac{w_L}{r_L}$, then, $w_L - r_L x_M = 0$. By assumption $\frac{r_H}{w_H} > \frac{r_L}{w_L}$, and therefore $u_{x_M} = 0$, as $w_L - r_L x_M + \beta G(x_M) > w_H - r_H x_M + \beta G(x_M)$.

For $x = [x^*, x_M]$ where $x^* = \frac{w_H - w_L}{r_H - r_L}$, $G(x)$ can be updated according to $G(x-1) = G(x) + u_x$, with $u_x = 0$, having in mind that $R_H(x) < R_L(x)$ for $x^* < x \leq x_M$. At x^* we will get $G(x^*) = G(x^* + 1) = G(x^* + 2) = \dots = G(x_M) = 0$

For $x = [0, x^*]$, as $R_H(x) > R_L(x)$, $u_x = 1$, and $G(x) = x^* - x$ will be uniformly decreasing, satisfying $G(0) = x^*$.

Spatial equilibrium under uncertainty

Monopolistic landlord have imperfect knowledge about housing demand. The probability that a location x will be occupied is μ_H for H workers and μ_L for L workers. We will assume an opportunity cost c that the landlord should afford when the location x is not occupied, what happens with probability $(1 - \mu_H)$ if x is assigned to H workers and with probability $(1 - \mu_L)$ if x is assigned to L workers.

The expected profit of the landlord will be now in terms of the Bellman equation,

$$V(x, G(x)) = \max_{u_x} [\mu_H R^H(x, G(x)) - c(1 - \mu_H)]u_x + [\mu_L R^L(x, G(x)) - c(1 - \mu_L)](1 - u_x) + E[V(x+1, G(x+1))]; x = 0, 1, \dots, x_M \quad (\text{A.7})$$

with the transition equation, $G(x+1) = G(x) - \mu_H u_x$

The monopolistic landlord will determine the urban configuration, that is, she define the city limits x_M' and how the x locations are assigned between workers types setting $u_x = 1$ if x is assigned to H workers, and $u_x = 0$ if the landlord assigns x to L workers. As the locations are occupied only with probabilities μ_H, μ_L , it will be possible that the location x remain unoccupied.

We will consider an equilibrium where both H and L workers are in the city area but changing accordingly the probabilities μ_H, μ_L can lead to an equilibria where the locations are assigned only to L workers or only to H workers.

As before, city limit x_M' are determined setting $\mu_L R^L(x_M', G(x_M')) - c(1 - \mu_L) = 0$ under the condition $G(x_M') = 0$. In this case we get

$$x_M' = \frac{w_L - c \frac{(1 - \mu_L)}{\mu_L}}{r_L} \quad (\text{A.8})$$

To determine x^{**} , that is, the locations assigned to H workers, we determine the value of x for which the landlord expected profit for both type of workers are equal,

$$\mu_L R^L(x^{**}, G(x^{**})) - c(1 - \mu_L) = \mu_H R^H(x^{**}, G(x^{**})) - c(1 - \mu_H) \quad (\text{A.9})$$

It is easy to show that when $\mu_L > \mu_H$, then $x^{**} < x^*$, where $x^* = \frac{w_H - w_L}{r_H - r_L}$ is defined in the non-random version of the model. In other words, the urban area assigned to H workers is reduced when the rate of arrival of L workers is higher than for H workers.

Using the transition equation $G(x) = G(x+1) + \mu_H u_x$ together with the assumption $G(x_M) = 0$ and given that L workers have higher willingness to pay for distant locations, for $x \geq x^{**}$ we get $G(x) = 0$, and therefore⁸

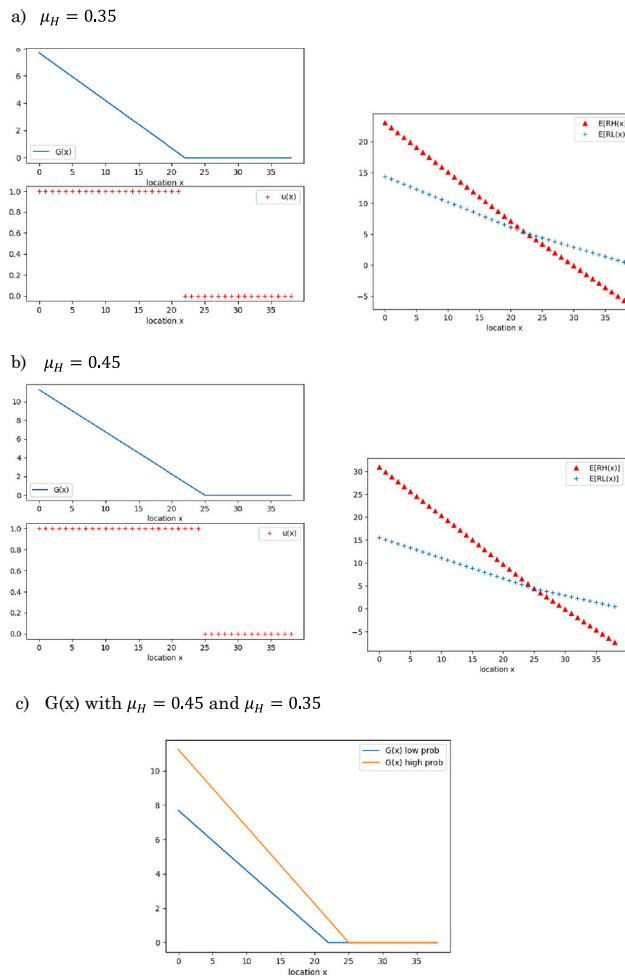
$$\mu_H(w_H - r_H x) - (1 - \mu_H)c = \mu_L(w_L - r_L x) - (1 - \mu_L)c$$

$$x^{**} = \frac{[w_H - gw_L] + (1 - g)c}{r_H - gr_L}$$

where $g = \frac{\mu_L}{\mu_H}$.

If $g = 1$, then $x^{**} = x^*$, and if g increases the value of x^{**} will be lower, as according Eq. (A.9) the landlord will reduce the urban space assigned to H workers in favor of L workers due to the relative higher rate of arrival of L workers.

For $x < x^{**}$, $u_x = 1$, and $G(x) = (x^{**} - x)\mu_H$. In the CBD, $x = 0$, we have $G(0) = x^{**}\mu_H$, that is, $G(0)$ is the expected number of locations occupied by H workers in the urban area. It can be concluded that the expected value of the neighborhood externality is lower due to the uncertainty in the H workers residing in the urban area. As μ_H increases Fig. A.1, under condition $\mu_L > \mu_H$, the value of x^{**} will converge to x^* , and $G(x)$ will increase.



	Model parameter values								Spatial equilibrium		
	β	w_H	w_L	r_H	r_L	μ_H	μ_L	c	x^*	x_M	$G(0)$
Upper graph	0.8	60	30	2	0.75	0.35	0.40	0.5	22	39	7.7
Lower graph	0.8	60	30	2	0.75	0.45	0.40	0.5	25	39	11.25

Fig. A.1. Spatial equilibria simulation after an increase in μ_H . Left: $(G(x)$ and $u_x)$. Right: $(E[R_H(x)], E[R_L(x)])$.

Appendix B. CBD identification based on cell phone data

Mobility data based on cell phones positioning published by the Spanish National Statistical institute (INE) for a representative week of November

⁸ Note that the value x^{**} should be rounded as the dynamic programming problem only considers a discrete number of locations x .

2019 is used to identify the city areas that “attract” people in a working day, as these areas can be associated to theoretical “central business districts” (CBD). The daily mobility data provides the number of cell phones that move from the area of “residence” of the telephone to the destination area. The area of “residence” of each mobile phone is the one where the mobile phone is located for the longest time between 22:00 h the previous day and 06:00 h during the observed day. The destination area is that area (it can be the same as the area of residence) in which the terminal is located for the longest time during the hours of 10:00 to 16:00 of the observed day, and remain at least 2 h in the destination area.

Based on this data, the CBD is defined as an area with a population of >10,000, in which the balance between inflows (people living in other areas and moving to the area) and outflows (people living in the area and moving to other areas) is >50 % of the resident population. Sparsely populated areas are excluded to prevent them from being classified as CBD despite having insignificant movements of people. With this limit, it is guaranteed that they are areas in which there are net entries of at least 5000 people. Fig. B.1 shows the different identified CBDs in the urban area, together with their corresponding centroids and the location of the city center (black point, Puerta del Sol). As can be seen, the CBDs are located fundamentally in the northern part of the urban area, with a particularly significant area on the border between the municipalities of Alcobendas and Madrid, which more than doubles its population during working hours. Comparing this map with the spatial distribution of income there is a high correspondence between the CBDs and high-income areas, which seems to confirm the importance of proximity to work centers in location decisions.

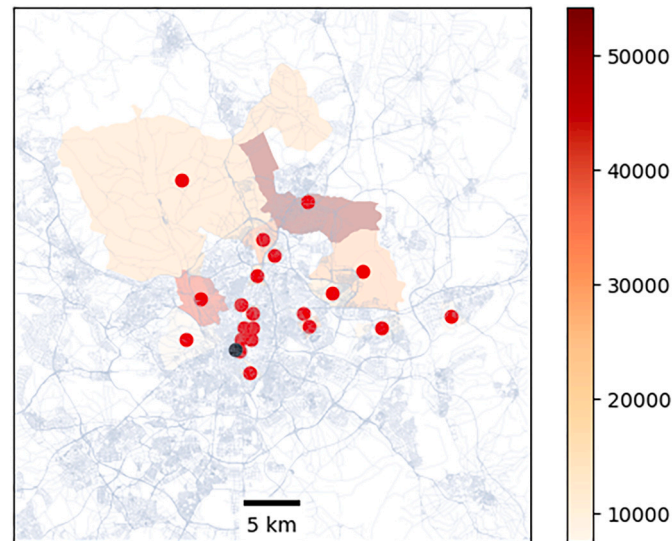


Fig. B.1. Location of the areas with the highest intra daily population variation (population entering minus population leaving) and comparison with the city center (black point). November 2019.

Source: own elaboration with INE data. Only areas in which the balance between inflows and outflows is >50 % of the resident population in that area (only areas with population higher than 10,000 are considered).

References

- Ahlfeldt, G. M. (2013). If we build it, will they pay? Predicting property price effects of transport innovations. *Environment and Planning A: Economy and Space*, 45-8, 1977–1994.
- Autor, D. H., & Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, 1553–1597.
- Baum-Snow, N., & Pavan, R. (2013). Inequality and city size. *Review of Economics and Statistics*, 95, 1535–1548.
- Bayer, P., Ferreira, F., & McMillan, R. (2007). A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy*, 115, 588–638.
- Bayer, P., & McMillan, R. (2012). Tiebout sorting and neighborhood stratification. *Journal of Public Economics*, 96, 1129–1143.
- Bayona-Carrasco, J., & Gil-Alonso, F. (2011). Suburbanisation and international immigration: The case of the Barcelona Metropolitan Region (1998–2009). *Tijdschrift voor Economische en Sociale Geografie*, 103(3), 312–329.
- Berkes, E., & Gaetani, R. (2023). Income segregation and the rise of the knowledge economy. *American Economic Journal: Applied Economics*, 15-2, 69–102.
- Blázquez, M., Llano, C., & Moral-Carcedo, J. (2010). Commuting times: Is there any penalty for immigrants? *Urban Studies*, 47(8), 1663–1686.
- Bothner, M. S., Smith, E. B., & White, H. C. (2010). A model of robust positions in social networks. *American Journal of Sociology*, 116-3, 943–992.
- Brueckner, J. K., & Rosenthal, S. S. (2009). Gentrification and neighborhood housing cycles: Will America's future downtowns be rich? *The Review of Economics and Statistics*, 91(4), 725–743.
- Brueckner, J. S. (1987). The structure of urban equilibrium: A unified treatment of the Muth-Mills Model. In , Vol. 11. *Handbook of regional and urban economics*. s.l.: Elsevier.
- Cao, J., & Smith, E. B. (2020). Why do high-status people have larger social networks? Belief in status-quality coupling as a driver of network-broadening behavior and social network size. *Organization Science*, 1–22.
- Chiu On Ki, W. M. J. (2010). The effects of urban redevelopment on neighbourhood housing prices. *International Journal of Urban Sciences*, 14-3, 276–294.
- Dauth, W., Findeisen, S., Moretti, E., & Südekum, J. (2019). *Matching in cities*. IZA discussion papers 12278. Institute of Labor Economics (IZA).
- Díaz-Palacios-Sisternes, S., Ayuga, F., & García, A. I. (2014). A method for detecting and describing land use transformations: An examination of Madrid's southern urban-rural gradient between 1990 and 2006. *Cities*, 40(part A), 99–110.
- Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, 79, 1407–1440.
- Edward L. Glaeser, M. E. K. (2004). Sprawl and urban growth. In J. V. Henderson, & J. Thisse (Eds.), vol. 4. *Handbook of regional and urban economics* (pp. 2481–2527). s. l.:s.n.
- Gallego-Valadés, A., Ródenas-Rigla, F., & Garcés-Ferrer, J. (2021). Socio-spatial stratification and social services in a small-sized city: The case of torrent (Spain). *Cities*, 111, 1–13.
- Guerrieri, V., Hartley, D., & Hurst, E. (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics*, 100, 45–60.
- Gutiérrez, E., Moral-Benito, E., & Ramos, R. (2020a). Tendencias recientes de la población en las áreas rurales y urbanas de España. In *Bank of Spain, Occasional Paper*. N.º 2027.
- Gutiérrez, E., Moral-Benito, E., & Ramos, R. (2020b). The spatial distribution of population in Spain: An anomaly in European perspective. In *Bank of Spain Working Paper* n.º 2028.
- Hill, R., & Dunbar, R. (2003). Social network size in humans. *Human Nature*, 14, 53–72.
- Jayantha Wadu Mesthrige, T. M. (2022). Transport infrastructure, accessibility and residential property values: Evidence from Hong Kong. *Built Environment Project and Asset Management*, 12-2, 163–179.
- Jordahl, K., et al. (2020). *Geopandas/geopandas: v0.8.1 (Version v0.8.1)*. <https://doi.org/10.5281/zenodo.3946761>
- Kain, J. F. (1968). Housing segregation, negro employment, and metropolitan decentralization. *Quarterly Journal of Economics*, 82(2), 175–197.
- Kaufman, H. F. (1952). An approach to the study of urban stratification. *American Sociological Review*, 17-4, 430–437.
- Kidokoro, T., Sho, K., & Fukuda, R. (2023). Urban suburbia: Gentrification and spatial inequality in workers' communities in Tokyo. *Cities*, 136, 1–14.

- Kleit, R. G. (2008). Neighborhood segregation, personal networks, and access to social resources. In N. K. K. J. H. Carr (Ed.), *Segregation: The rising costs for America* (pp. 237–260). s.l.:s.n.
- Lehmer, F., & Möller, J. (2010). Interrelations between the urban wage premium and firm-size wage differentials: A microdata cohort analysis for Germany. *The Annals of Regional Science*, 45, 31–53.
- Libertad Gonzalez, F. O. (2013). Immigration and housing booms. Evidence for Spain. *Journal of Regional Science*, 53-1, 37–59.
- Lindley, J., & Machin, S. (2014). Spatial changes in labour market inequality. *Journal of Urban Economics*, 79, 121–138.
- Markusen, J. R., & Scheffman, D. T. (1978). Ownership concentration and market power in urban land markets. *Review of Economic Studies*, 45(3), 519–526.
- McCormick, B. (1986). Employment opportunities, earnings, and the journey to work of minority workers in Great Britain. *Economic Journal*, 96(382), 375–397.
- Morteza Moallemi, D. M. (2020). The impact of immigration on housing prices in Australia. *Papers in Regional Science*, 99-3, 773–786.
- Mussa, A., Nwaogu, U. G., & Pozo, S. (2017). Immigration and housing: A spatial econometric analysis. *Journal of Housing Economics*, 35, 13–25.
- Ponce, J. (2010). Affordable housing and social mix: Comparative approach. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, 2(1), 31–41.
- Ponce, J. (2019). Urban planning and legal framework for sustainable communities: Affordable housing, social cohesion and ghettos. *International Journal of Legal Information*, 42(1), 75–120.
- Roca, J.d.l., & Puga, D. (2017). Learning by working in big cities. *Review of Economic Studies*, 84, 106–142.
- Saiz, A. (2007). Immigration and housing rents in American cities. *Journal of Urban Economics*, 61-2, 345–371.
- Sambricio, C. (1999). La vivienda en Madrid, de 1939 al Plan de Vivienda Social, en 1959. In *La vivienda en Madrid en la década de los cincuenta: el Plan de Urgencia Social*. Madrid: s.l.:Electa.
- Sánchez-Moral, S., Arellano, A., & Díez-Pisonero, R. (2018). Interregional mobility of talent in Spain: The role of job opportunities and qualities of places during the recent economic crisis. *Environment and Planning A: Economy*, 50(4), 789–808.
- Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, 59-2, 488–493.
- Smith, E. B., Menon, T., & Thompson, L. (2011). Status differences in the cognitive activation of social networks. *Organization Science*, 23-1, 67–82.
- Stopczynski, A., Pentland, A., & Lehmann, S. (2018). How physical proximity shapes complex social networks. *Scientific Reports*, 8(17722), 1–10.
- Su, Y. (2022). The rising value of time and the origin of urban gentrification. *American Economic Journal: Economic Policy*, 14(1), 402–439.
- Turner, M. A. (2008). Residential segregation and employment inequality. In N. K. K. J. H. Carr (Ed.), *Segregation: The rising costs for America* (pp. 151–196). s.l.:s.n.
- Valenzuela-Rubio, M. (1974). *Iniciativa oficial y crecimiento urbano en Madrid (1939–1973)*. Issue 137. Estudios Geográficos.
- Yankow, J. J. (2006). Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. *Journal of Urban Economics*, 60, 139–161.
- Yates, S. R., & Cowart, L. B. (2019). The impact of shuttered golf courses on property values. *Journal of Sustainable Real Estate*, 11.
- Zenou, Y. (2009). *Urban labor economics*. s.l.:Cambridge University Press.
- Zhu, Y., Fu, Q., & Ye, C. (2022). State-embedded gentrification in China. *Cities*, 131, 1–13.