WORKING PAPER 25-06

Accessibility and population patterns: A study of the expansion of the rail network in Madrid (1998-2020).

Diego Buitrago-Mora
Anna Matas

Departament d'Economia Aplicada
Edifici B
Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona, Spain.
d.econ.aplicada@uab.cat
www.uab.cat/en/applied-economics
@ecapuab.bsky.social

UAB
Universitat Autònoma
de Barcelona

Accessibility and population patterns: A study of the expansion of the rail network in Madrid (1998–2020).

Diego Buitrago-Mora[2]

Anna Matas [3]

May 2025

Abstract: This study examines the relationship between the expansion of the rail system in the Madrid metropolitan area and patterns of population redistribution between 1998 and 2020. We focus on two major Metro expansion phases, which together account for approximately 40% of the current network. To capture the effects of improved infrastructure, we construct an accessibility measure based on the increase in job opportunities resulting from reductions in travel time. To address potential endogeneity in this key variable, we propose an instrument based on subsoil permeability, which serves as a novel proxy for tunneling feasibility. We estimate an elasticity of population with respect to travel time of 0.7, which remains stable across varying magnitudes of travel time reductions. The effect is stronger for individuals aged 40 to 59 and for the Spanish-born population.

Key words: Accessibility, population patterns, suburbanization, rail expansion.

[2]Department of Urbanism, TU Delft, Faculty of Architecture and the Built Environment, Julianalaan 134, 2628 BL Delft (e-mail: dbuitragomora@tudelft.nl; phone: +31 620593827; website: https://diegob901.wixsite.com/diegobuitragomora. This research has benefitted from funding from Minciencias Colombia Convocatoria Exterior 860.

[3]Department of Applied Economics, Universitat Autònoma de Barcelona, Edifici B, Facultat d'Economia i Empresa, 08193 Cerdanyola del Vallès, Spain (e-mail: anna.matas@uab.cat; phone: +34 93 581 4584; website: https://ieb.ub.edu/en/researcher/matas-prat-anna/. This research has benefitted from funding from Spains Ministerio de Ciencia y Universidades under project RTI2018-097434-B-I00.)

1. Introduction

In this study, we investigate the relation between the rapid growth of the rail network and population intra-urban mobility patterns in the Madrid Metropolitan Area (MMA)[1].

We look at two expansion plans for the light rail and Metro of Madrid, which account for roughly 40% of the total network, as well as the changes in the network of suburban train. We measure this growth by observing the variation in accessibility caused by the opening of new stations and whether that variation is a determining factor in population mobility patterns inside the region, which increased by more than 1.7 million inhabitants between 1998 and 2020. We use census tract-level data (3,908 per year). Although our main results focus on the MMA, we also refer to the *Region of Madrid* (Autonomous Community of Madrid), which is the larger administrative region encompassing the MMA, the *Municipality of Madrid* (the central city of the metropolitan area), and the *districts*, which are the internal administrative subdivisions of the Municipality of Madrid.

This study adds to a large body of literature that investigates the relation between an expansion of the transportation infrastructure and the relocation of the population (or employment). The study of this relation has important implications for policymakers. First, when planning the construction of highways or urban trains, it is critical to understand the medium- and long-term impact of such projects, as well as their potential to generate decision patterns in housing sorting among the population. Second, to investigate the changes in transportation and commuting costs, which are regarded as one of the main causes of population decentralization [8], since these changes are an important driver of the increase in agglomeration benefits. Third, in order to plan and build more sustainable cities with a higher quality of life, it is necessary to have adequate urban transportation planning. Furthermore, understanding where the population will (potentially) be located and which areas will be more appealing than others is critical to addressing the issues related to the provision of housing. Finally, understanding the characteristics of the population groups that are relocating, such as age groups, as well as whether they are immigrants or not, is a cross-cutting goal in this field of study.

There has been a substantial amount of research investigating the relationship between highways and suburbanization. [1] analyzes the impact of new commuting highways in U.S. metropolitan areas and concludes that highway construction accounts for nearly 28% of the population decline in the average city center.[7] also show that the stock of highways in the United States contributed to urban population growth between 1980 and 2020. [10] investigate highway expansion in Spain and its association with population decline in central business districts (CBDs) between 1960 and 2011. More recently, [31], analyzing 579 cities across 29 European countries, estimate that each additional highway ray decreased the central city population share by an average of 5%. Hence, although the present study focuses primarily on trains, it considers the expansion of the main highway network to be a fundamental control.

In contrast, the literature on the relationship between rail investment and the spatial distribution of urban populations is less abundant and less conclusive. [21] demonstrate that the extension of the Paris regional train system led to a 9% increase in employment, without significantly affecting population growth. Also in the Paris Metropolitan Area, [29] corroborate the decisive role of

[1]Unlike other European metro areas, the Madrid Metropolitan Area (MMA) has no official boundaries. The most recent official definition dates from 1964. As a result, in this study, we will define the MMA using the definition proposed in [4], which includes twenty-eight municipalities. The municipalities that make up the MMA are: Madrid, Alcobendas, San Sebastián de los Reyes, Colmenar Viejo, Tres Cantos, Pozuelo de Alarcón, Majadahonda, Las Rosas, Boadilla del Monte, Villaviciosa de Odon, Villanueva de la Cañada, Villanueva del Pardillo, Brunete, Alarcón, Leganés, Getafe, Móstoles, Fuenlabrada, Parla, Pinto, Coslada, San Fernando de Henares, Torrejón de Ardoz, Alcalá de Henares, Paracuellos de Jarama, Mejorada del Campo, Velilla de San Antonio, and Rivas-Vaciamadrid.

the RER network in driving employment growth and the emergence of employment subcenters. [9] shows that improvements to highways and railroads stimulate suburban population growth in Barcelona. [24] examine, among other factors, how improvements in transportation infrastructure affect the spatial distribution of populations, presenting evidence at various spatial scales both within and across cities. [13] explore the link between subway systems, population, and spatial configuration. Using nighttime lights as indicators of centralization for 138 subway systems worldwide, they find that while subways decentralize cities, the effect is small. [31] conclude that, unlike highways, railways have no average impact on suburbanization across the 579 cities studied. However, when disaggregating by city type, they find that railways do contribute to suburbanization in smaller, more centralized, inland cities that are warmer, more rugged, and located in Mediterranean regions.

The large scale expansions of Madrid's metro network has led to various studies analyzing its impact on the population, land prices, and the location of economic activities. [2] analyze the impact of the expansion of two metro lines on residential developments and population growth over an 11-year period. They find more dynamic urbanization and population settlement near new stations compared to areas without urban rail service or with older stations. Population growth was especially strong near stations located in the outer areas of Madrid, with a much smaller effect in the Madrid city center. Additionally, they find that the intensity of the effect diminishes as the distance to the station increases. [28, 27] evaluate the effects of the new Metrosur line on real estate prices and on firm location, respectively. They find that improved rail accessibility positively influenced property values, with suburban rail having a greater impact than Metrosur. Municipalities furthest from Madrid, with lower incomes and higher transport fares, benefited the least. Regarding firm location, they show that the patern of economic activity location is closely linked to accessibility. Finally, [26] compare London's Jubilee Line Extension and Madrid's Metrosur, identifying key preconditions and factors that help explain the likelihood of success of an investment in terms of economic impact.

Our data and methodological approach contribute significantly to the literature in several ways. First, and as a key differentiating feature from previous studies, our conclusions are based on a major expansion of the metro network. Over a relatively short period of 20 years, the network almost doubled in size and incorporated a tramway system into the broader rail network. Therefore, our findings reflect substantial, system-wide changes, rather than isolated improvements to individual metro lines.

Second, the impact is analyzed within a region characterized by high levels of development and a dense concentration of economic activity in its central area. As such, our findings offer valuable insights for policymakers focused on urbanization processes in large European metropolitan areas.

Third, while most prior studies measure connectivity differences using distance to the nearest access point (e.g., ramps or train stations), we employ an opportunity-based accessibility measure. This metric captures not only local effects but also changes in accessibility brought about by system-wide alterations. Previous research has used similar measures to assess changes in accessibility to employment [14, 20] or education [25] in evaluating the effects of transport infrastructure on suburbanization. We adopt the accessibility measure proposed by [16], quantifying the attractiveness of each census tract by the number of jobs reachable in the municipality of Madrid (where over 60% of total employment is concentrated) via the rail system. Our primary independent variable is the change in accessibility between 1998 and 2020, specified in first-difference form. This offers a comprehensive view of how the entire rail network influences job accessibility and intra-urban spatial patterns.

Fourth, this is the first study to consider the expansion of all Metro, light rail, and suburban train lines in conjunction, supported by precise data on connections between the three modes, made possible through the introduction of our accessibility measure and the use of General Transit Feed

Specification (GTFS) files.

Finally, to address potential endogeneity, we rely on geological variables that affect the feasibility of tunnel construction—a critical element in the network's expansion. The use of continuous geological subsoil data as instrumental variables enhances the robustness of our analysis by mitigating concerns about endogenous accessibility variation. For these reasons, our study provides robust evidence on the impact of the urban rail network on population growth. The results show that there is a positive and significant impact but limited to municipalities outside the central area, suggesting that the metro contributes to suburbanization. Thus, our findings from a large-scale metro expansion align with results reported for cities such as Paris and Barcelona. Moreover, we differentiate the impact by age group and country of origin. Overall, this study offers valuable insights into transportation infrastructure planning and its implications for urban dynamics.

After this introductory section, the subsequent sections of this study are structured as follows: Initially, the data section introduces the primary sources used and explains the construction of the model variables, including the accessibility measures. Notably, it focuses on population dynamics and the expansion of the rail system. Subsequently, we present the empirical strategy section, which is structured into two main components: the formulation of the empirical model and the selection of instruments. This is followed by a section dedicated to presenting the main results, including a marginal effects analysis and robustness checks. Finally, we conclude with a section summarizing our findings.

2. Data

This section presents the variables included in the analysis, including both the dependent and explanatory variables. We provide information on their original sources, how these variables were constructed and relevant descriptive statistics.

2.1. Population patterns in Madrid

From 1998 to 2020, the total population of Spain increased by 17%, while the population of the Region of Madrid, increased by 35% during the same time, reaching 6,779,888 people. The Madrid Metropolitan Area (MMA) is the second-largest metropolitan agglomeration in the European Union, behind Paris but ahead of Berlin, Barcelona, and Rome.

As shown in Figure 1, both the foreign-born and local populations have driven the growth. However, the variability has been more pronounced among the foreign-born, who represented 14% of the total population in 2020. Madrid's most important foreign-born populations are from South America and the European Union.

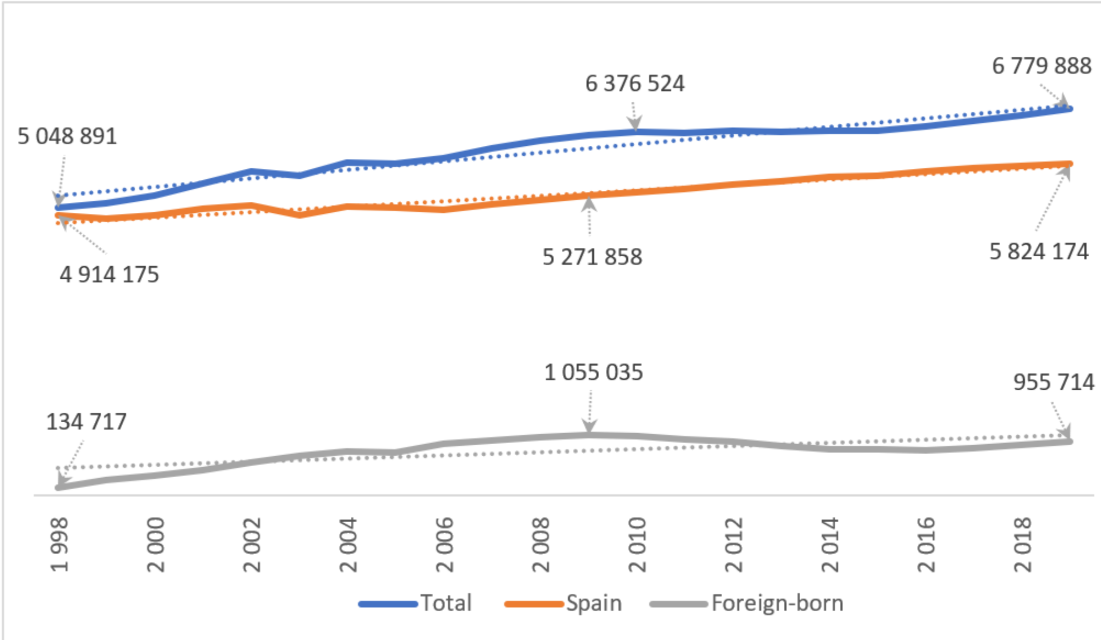
Our variable of interest is the population of the Metropolitan Area of Madrid (MMA) at the census tract level. For the period from 1998 to 2020, we obtained data from the Madrid Institute of Statistics' annual register (padrón) for each of the 3,908 census tracts. The data include the number of residents, their country of birth and their age.

It is important to note that in Madrid, the boundaries of census tracts are updated annually, which gives rise to two issues. The first concerns the accuracy of the maps, as GIS analyses have improved in precision over the years, and the latest maps include elements such as platforms and vehicular roundabouts. The second, and more significant, issue is the change in the size of census tracts due to the creation of new ones. To address these differences, we follow the methodology of [18] and [19], using an aerial interpolation process[2].

Figure 2 depicts the change in population between 1998 and 2020 at the census tract level (the darker the census tract, the greater the variation). When comparing the first and last years of

[2]In Appendix A

Figure 1: Population trends in Madrid between 1998 and 2020.



Notes: Total population based on the annual official register (padrón). Own development using data of the Spanish Institute of Statistics (INE).

our study, the raw data suggests that the population has moved towards the periphery of the metropolitan area, the boundaries of the Municipality of Madrid, and outside of it. Some areas in the south, north, and northeast were empty in 1998 and have since then been transformed into areas that accommodate a large proportion of the total population in 2020. It is also worth noting that there was negative growth of population in some census tracts, implying that population relocation was significant in some areas. In this way, we are capturing not only the expansion in some census tracts, but also the heterogeneous contraction in several parts of the metro area.

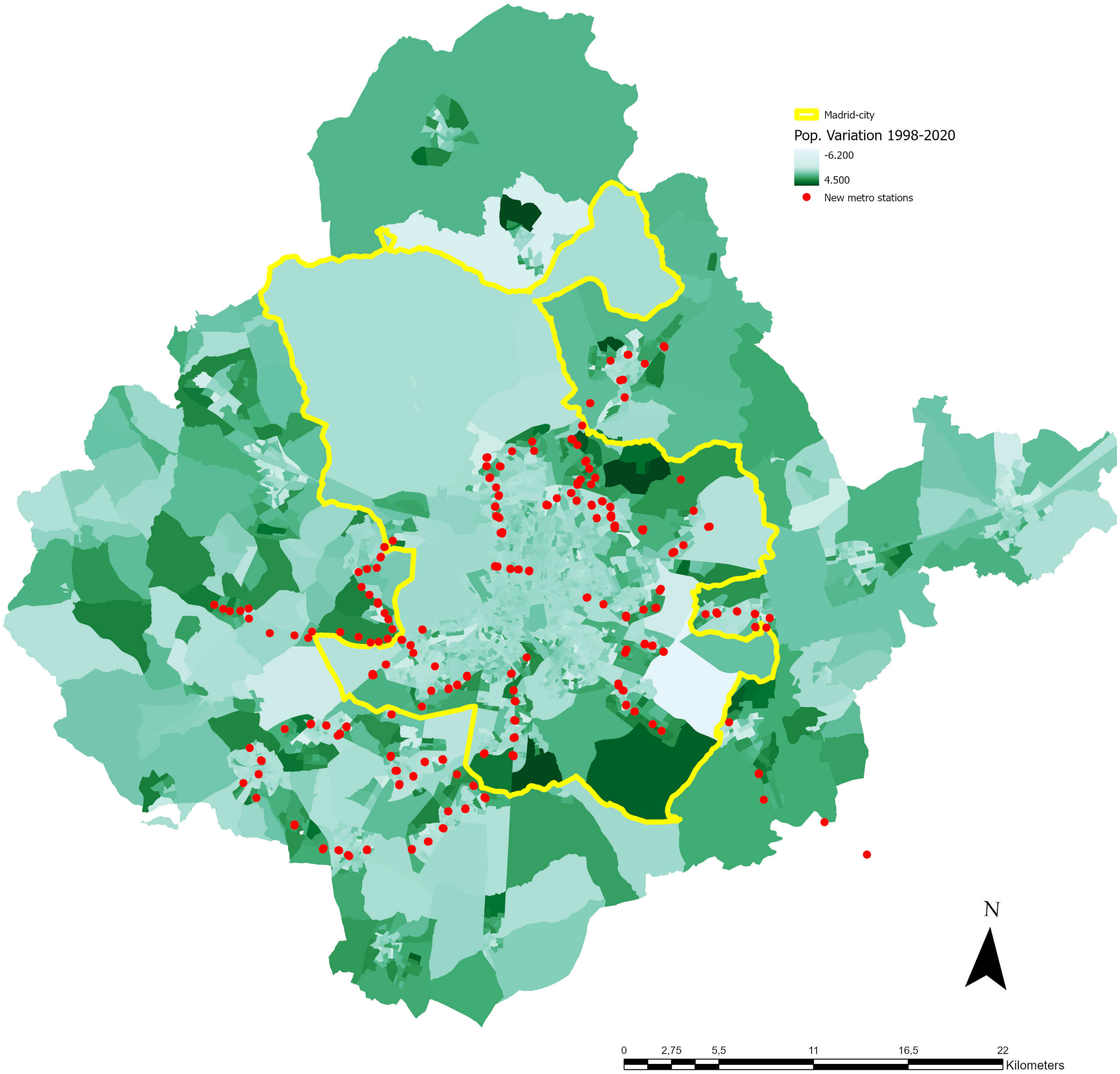
2.2. *Expansion of the rail system in Madrid between 1998 and 2020.*

Madrid's railway network includes 361 kilometers of suburban trains, 289 kilometers of Metro and, 36 kilometers of light rail. The municipality of Parla has an approximately 8.2-kilometer long light-rail line. This line is included in Table 1, but not in the study, since it only provides local service. The most ambitious expansion plans in the history of the Madrid railway system were carried out between 1998 and 2020, with the main emphasis on the expansion of the Metro network. Table 1 shows how, in 1998, the Metro only served the central municipality; by 2020, it had expanded to 11 municipalities, and a light rail system had also emerged. Although the suburban train system expanded, it did not do so at the same rate, and the majority of the new stations were in municipalities that already had a connection to the system. The train system is fully integrated, which means that if a family lives near one Metro, light rail, or suburban train station, they live near the entire system. In this section, we will go over the two Metro expansion plans (which include light rail) in detail, as well as mention the expansion of the suburban train system. This analysis is critical in the context of this research because the accessibility measure we propose takes such changes into account.

2.2.1. *Metro and light metro of Madrid*

The Madrid metro was inaugurated in 1919 with a length of 4 km. Since then, the network has grown to become one of the largest in the world (ranked eighth), with many stations (ranked fifth) and escalators (ranked fourth). The Metro is a vast network of mostly underground tunnels

Figure 2: Madrid Metropolitan area and the Municipality of Madrid: Δ(2020-1998) total population, by census tract.



Notes: Own development from data from INE.

through which nearly 2,400 trains pass daily; it is the backbone of the transportation system of Madrid.

Table 2 shows that the network of 1995 only amounted to 42% of the 2020 system, implying that more than half of the 2020 system was constructed in the last three decades.

Following two ambitious plans of expansion (one from 1995 to 2003 and the second from 2003 to 2007), Madrid's subway network expanded from 121 kilometers to 325 kilometers and from 164 to 322 stations between 1995 and 2020. Two entirely new lines, as well as three new light rail lines, were built.

Although the expansion plans we consider here had been in place since 1995, we will evaluate the impact on the population since 1998 due to the availability of data at the census tract level. Figure 3 shows that prior to the expansion plans, the Metro network did not extend beyond the Municipality of Madrid limits (green zone). Buses and the suburban train system connected it with the MMA (yellow area). After the expansion, a large part of the region was connected with the

Table 1: Rail mode by municipality in 2020. Metropolitan Area of Madrid (MMA).

Rail mode	Metro		Light rail		Suburban	
Municipality	1998	2020	1998	2020	1998	2020
Alcalá de Henares					✓	✓
Alcobendas		✓				✓
Alcorcón		✓			✓	✓
Boadilla del Monte				✓		
Brunete						
Colmenar Viejo						✓
Coslada		✓			✓	✓
Fuenlabrada		✓			✓	✓
Getafe		✓			✓	✓
Leganés		✓			✓	✓
Madrid	✓	✓		✓	✓	✓
Majadahonda					✓	✓
Mejorada del Campo						
Móstoles		✓			✓	✓
Paracuellos de Jarama						
Parla				✓*	✓	✓
Pinto					✓	✓
Pozuelo de Alarcón				✓		
Rivas-Vaciamadrid		✓				
Rozas de Madrid (Las)					✓	✓
San Fernando de Henares		✓				
San Sebastián de los Reyes		✓				✓
Torrejón de Ardoz					✓	✓
Tres Cantos					✓	✓
Velilla de San Antonio						
Villanueva de la Cañada						
Villanueva del Pardillo						
Villaviciosa de Odón						

Notes: * The light rail line of Parla was excluded. However, this municipality is integrated into the system through the suburban train. Several new suburban train stations were opened as alternatives in municipalities where another station already existed; in this table, we compare municipalities with and without a station in 1998 and 2020.

Metro [3].

2.2.2. *Suburban trains in Madrid*

In 1851, the first line of the Madrid suburban train network opened between Atocha (Madrid's downtown and then most important transportation hub) and the municipality of Aranjuez. There are currently 11 lines with 95 stations, 15 of which were opened between 1995 and 2020, but coverage did not change much by municipality during our study period, and some stations were even closed. In Table 1, we see that suburban trains serve 17 of the MMA's 28 municipalities.

Though its expansion was less ambitious than that of the Metro (in part because it was already the most extensive rail mode in the MMA), we included it in this study because we use an accessibility metric (described in the following section), and thus we consider the overall effect of the system's network. This means that municipalities where there was no variation in the suburban train system (i.e., they already had a station and no new one was opened) will in any case experience a change in accessibility if new stations are opened in another municipality.

[3] In Appendix 2, we detail the general and specific objectives of the plans, as well as the timeline of construction.

Table 2: Expansion of the Public Transport Network in Madrid (1995–2020)

Year	Nº lines	Stations (net/line)	Length (km)	Nº lines	Nº stations	Length (km)	Length (2020 %)
		Metro			**Light Rail**		**Total**
1995	11	n.a./164	121				42
1998	11	n.a./182	135				47
2003	12	188 / 237	227				79
2005	12	188 / 237	227				79
2007	12	230 / 285	278	4	56	36	97
2010	12	233 / 286	281	4	56	36	98
2015	12	236 / 286	289	4	56	36	100
2020	12	237 / 287	289	4	56	36	100
		Suburban Rail (Renfe)					
1995	8	80 / n.a.	281				
1998	11	78 / n.a.	280				
2000	11	80 / n.a.	315				
2003	12	97 / n.a.	337				
2005	12	99 / n.a.	340				
2007	9	99 / n.a.	337				
2010	9	98 / 155	387				
2015	9	94 / 169	391				
2020	11	95 / 203	361				

Notes: In 2020, the network accounting criteria for the suburban rail (Renfe) were modified and the data are not comparable with previous years.
Source: Consorcio de Transportes de Madrid, Annual Reports.

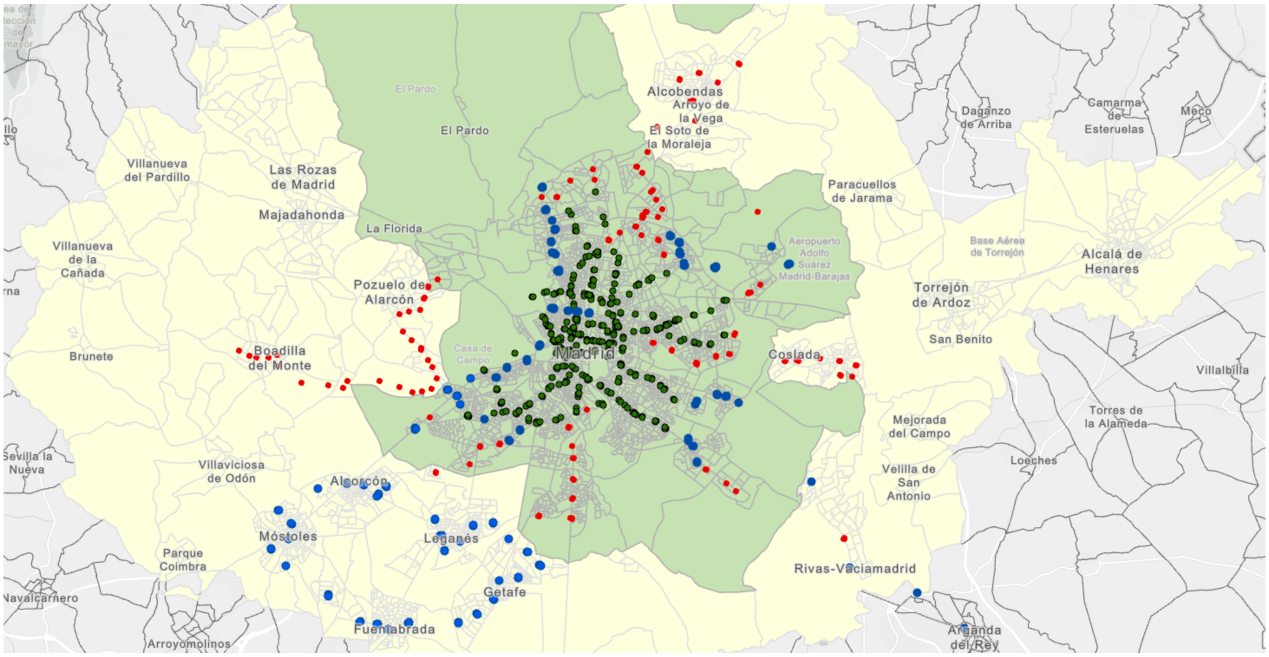
2.3. *Accessibility Measures*

We aim to see if the variation in job accessibility caused by the expansion of the rail system influences the patterns of redistribution of the population in Madrid. Due to the centralized structure of the MMA (with the central municipality accounting for roughly 60% of total employment), we focus on the variation of accessibility from the peripheral municipalities to the Municipality of Madrid between 1998 and 2020. Using an accessibility measure has an advantage over traditional methods such as Euclidean distances or binary variables, because it includes the effect of the entire network, beyond the local effect of the new stations alone [30].

The accessibility measure we propose is a gravity-based measure that follows [16] and [12]. It seeks to assign a weight or importance to each area before and after the transportation innovation in order to measure its attractiveness in comparison with the other ones. Such a weight is derived from the number of opportunities available, which, in our case, is proportional to the number of jobs to which residents of a census tract have access. Territorially disaggregated employment data is available from the decennial censuses and only at the district level, not at the level of census tracts. For our sample period, we have data for the years 2001, 2011, and 2021 for the 21 districts of the city of Madrid. Since our objective is to measure the effect of changes in accessibility resulting from variations in the rail network—and not from changes in employment levels—we chose to keep the number of jobs fixed. This approach also avoids inconsistencies in the variation of employment composition across different census years. We selected 2011 because it is the midpoint of our sample period. Appendix B shows that the distribution of employment across districts has remained very stable across the three census years for which data is available. Therefore, we understand that the choice of a specific year does not affect the results.

The accessibility measure takes the following form:

Figure 3: Entrances Metro of Madrid 2020.



Notes: Own development from data from INE and Madrid´s Consortium of transport. Green: Metro network in 1998. Blue: Expansion plan between 1995 and 2003. Red: expansion plan between 2003 and 2007 (finished in 2020). Green area: Municipality of Madrid. Yellow area: MMA.

$$A_{it} = \sum_j E_j exp(-\beta d_{ij,t}) \tag{1}$$

For the above, we define the accessibility A as the sum of jobs E that can be reached from the ith census tract (centroid) in the metropolitan area of Madrid, to the jth district in the Municipality of Madrid. The d_{ij}s are the travel times from the origin–destination matrix that includes the three existing train modes (Metro, light rail, and suburban train) as well as walking times to the system (We expand on the method to compute d_{ij} at the end of this section.).

Using a gravity equation makes it possible to define a distance decay parameter with β. That is, the jobs closest to the ith centroid have a higher weight, which decreases as we move away from the ith centroid. Besides, the exponential equation easily accommodates the area´s own employment when computing total accessibility. When $d_{ii} = 0$, $exp(-\beta d_{ii,}) = 1$ and therefore the area´s entire employment is included in the measure.

In order to compute the accessibility measure, we need to approximate the value of β. To do so, following [23], we look for a value of β that agrees with the observed behavior of commuters in the MMA. From Equation 1 we can express the proportion of employment (Q) that is counted into the ith area in the accessibility measure as:

$$exp(-\beta d) = Q \tag{2}$$

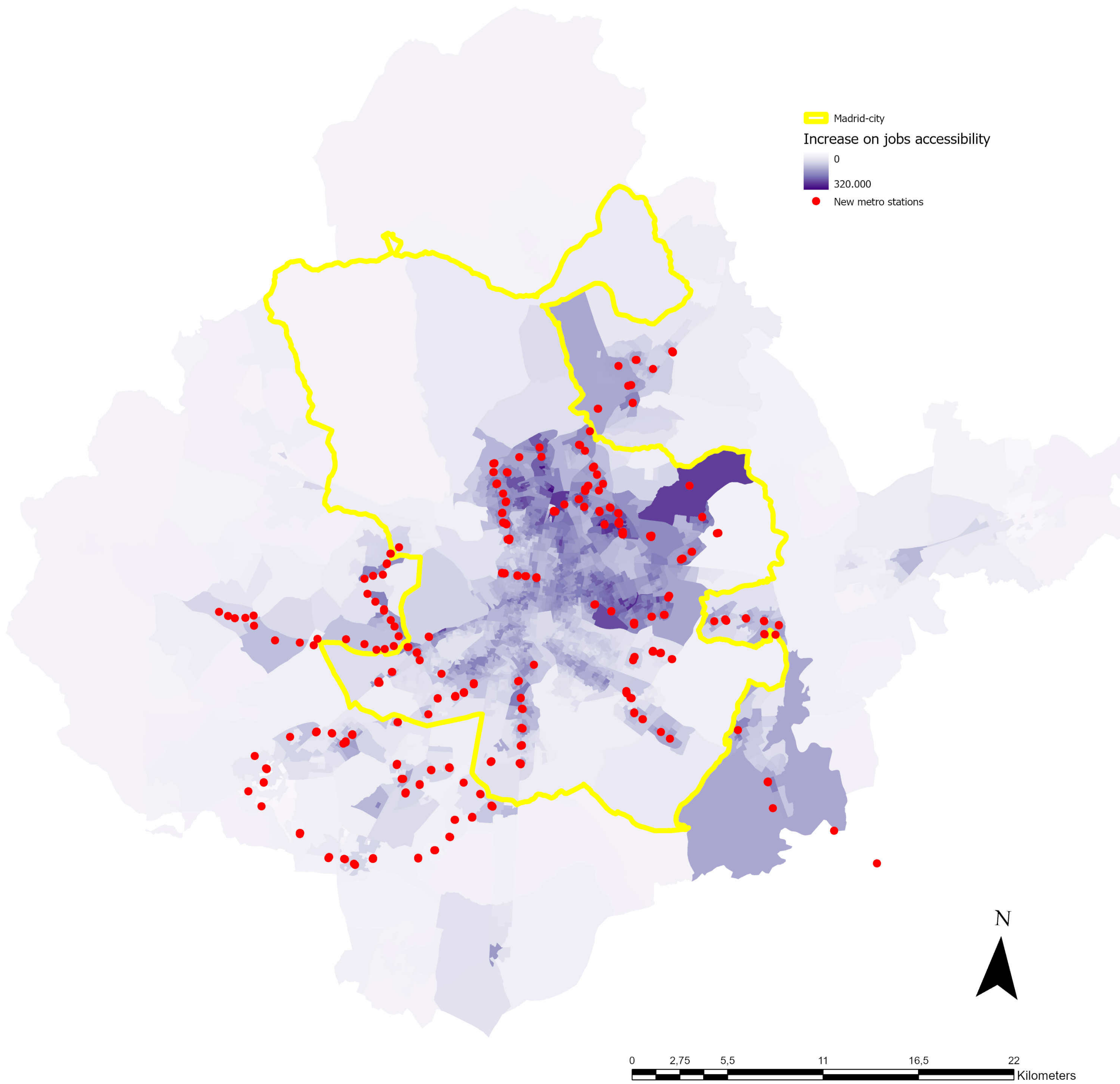
Taking logarithms, rearranging terms, and explicitly solving for β results in

$$\beta = -\frac{ln(Q)}{d} \tag{3}$$

Therefore, we can calculate β fixing Q and d. We evaluated β at the maximum commute time observed in MMA. We use real data on commute times generated by the INE based on the decennial survey of *Essential Characteristics of Population and Housing 2021* (Encuesta de Características Esenciales de la Población y las Viviendas) which complements the 2021 *Population and Housing*

Census. We filter by municipalities in the Madrid metropolitan region (excluding the Municipality of Madrid itself) and by those who have to commute to work to a different municipality (we exclude people working from home or in the same municipality). The sample has around 660,000 workers. According to this sample, the maximum round trip commuting time is two hours, or one hour each way. This maximum time was reported by 11% of the commuters. As a result, we define $Q_{ij,max}$ = 0.11 and $d_{ij,max}$ = 60 (minutes). Thus, we found a β of -0.0367 from Equation 3. In map 4 we show the findings of the differences in accessibility between 1998 and 2020, the darkest areas represent census tracts with the greatest increase in accessibility.

Figure 4: Madrid Metropolitan area and the Municipality of Madrid: Δ(2020-1998) Accessibility, by census tract.



Notes: Own development from data from INE.

Additionally, we have approximated accessibility using the concept of *effective density*. This concept was developed by [14] and applied in the Spanish context by [17] and [20].

$$ED_{it} = \sum_{j=1}^{21} \frac{E_j}{d_{ij,t}} \tag{4}$$

where ED is the Effective Density at census tract i and period t (1998 or 2020), E denotes for total employment from the 2011 census in district j, while the $d_{ij,t}$ are the travel times from the origin–destination matrix between each census tract and each district´s centroid in the central municipality (the Municipality of Madrid).

Origin–destination matrix (GTFS files)

To compute accessibility using Equations 1 and 4 , the matrix of origin–destination times ($d_{ij,t}$) between each census tract (centroid) and the 21 districts of Madrid must be generated. The matrix will contain the time (in minutes) that a passenger takes in the rail system to travel between all feasible combinations of census tract and district. We use the General Transit Feed Specification (GTFS) files for the computation, in addition to the actual information on the network of the streets along which to walk to get to the railway station.

GTFS files are created by public transportation operators in order to include geo-referenciated and homogeneous data structures into a spatial analysis using a standard vocabulary. The GTFS includes the structure of schedules, frequencies, fares, routes, stops, calendars, and trips for each mode of transport. In our case, the files for the three modes are published by the Madrid Regional Public Transport Consortium.

We include the network of streets where people may move using a GIS management software. Finally, we run simulations at 9 a.m. It is important to note that we assumed that the speed and service frequency remain constant over time. This assumption is necessary due to the unavailability of GTFS files for 1998. As a result, our analysis focuses solely on the effect of network expansion. The assumption of stability in speed and frequency is supported by available data on the railway network. Appendix D shows that, for each metro line, variations in these two variables have been low during the sample period. The only exception is Line 8, which exhibits more significant changes; however, these are confined to the early years of the sample, specifically 1998 to 2000.

2.4. Additional explanatory variables

In addition to our main variable of interest, rail accessibility, we include as control explanatory variables the following: connectivity to highways, distance to the CBD, the area of the census tract, and population in the initial year.

The Madrid transport consortium provided us with geo-referenced information on Metro, light rail, and suburban train entrances for the year 2020. We built the network in 1998 based on this information.

During our sample period, there was also a significant expansion of the system of highways connecting municipalities in the metropolitan area of Madrid. To capture the impact of this expansion, a new variable has been created that measures the change in the distance between the centroid of each census tract and the nearest highway access point. The variation has been calculated between the years 1990 and 2017, the period during which the highway network expansion took place.

$$2017 - 1990\Delta\text{distance(highway ramp access)} = distance_{2017} - distance_{1990} \tag{5}$$

That is, for each census tract (centroid) we compute the change in distance to the nearest highway access ramp between 2017 and 1990. Using geo-referenced data published by the Madrid Statistics Institute, we were able to construct the 1990 road network taking the network in 2017 (which is the last year with information available, including the latest expansion plan.) as the starting point and subsequently excluding the highways built in that period.

The distance to the CBD of the metropolitan region (that we set to be the *Plaza del Sol* in Madrid, which also serves as the zero point of the radial highway system), and the area of each census tract are the result of our own calculations using geo-referenced information from the census tracts and with the support of GIS packages.

Table 3: Statistics of the main variables and data sources.

	Metro Area			Metro area (exclu. the Munici. of Madrid)		
Variable	Obs.	Min.	Max.	Obs.	Min.	Max.
Δ Tot. Pop. 2020–1998	3,246	-6,213	2,963	1,070	-6,213	2,963
Δ Spa. Pop. 2020–1998	3,246	-6,355	2,705	1,070	-6,205	2,705
Δ Fore. Pop. 2020–1998	3,246	-321	987	1,070	-161	987
Δ Accessibility 2020–1998 [c]	3,246	-19,692	335,986	1,070	-19,692	207,983
Distance to CBD [a]	3,246	0	34,681	1,070	6,647	34,681
Δ Distance (hwy. ramp) 2017–1990 [a]	3,246	-18,298	0	1,070	-18,298	0
Area census tract [b]	3,246	0.01	9.83	1,070	0.01	9.83
Population 1998 (initial)	3,246	3	8,381	1,070	3	7,612

Database	Source	Period	Spatial level
Tracts	Madrid Statistics Institute	1998–2020	Census tracts
Employment	National Statistics of Spain (INE)	2011	Districts
GTFS travel times	Madrid Transport Consortium	2020	Metro area
Highway ramp	Madrid Statistics Institute	1990–2017	Metro area
Land Use	CORINE	2020	Metro area
Population	Madrid Statistics Institute	1998–2020	Census tracts
Lithology	Geological institute of Spain (IGME)	2020	Metro area: continuous data.

Notes: [a] Variable in meters. [b] Variable in square kilometers. Distance hwy. ramp = distance to the closest highway ramp access. Tot. pop. = Total population. Spa. pop. = Spanish population. For. pop. = Foreign-born population. [c] Number of jobs accessible in the central municipality (defined in Section 2.3).

Finally, data from the European Land Cover Project (CORINE) allowed us to identify the percentage of residential land use at the lot level. Based on this information, we excluded from our analysis any census tracts with more than 90% non-residential use. We had 3,693 census tracts at the end of the process. We then exclude census tracts with zero population in 1998 (we give further explanation in the subsequent section). At this point we have 3,246 census tracts. There are 2,176 that correspond to the Municipality of Madrid and 1,070 that correspond to the MMA outside the Municipality of Madrid. We obtain information by country of birth and age group from the same database.

In Table 3. The upper panel shows descriptive statistics for the main variables used in this study for the Madrid metropolitan area, both including and excluding the Municipality of Madrid. The lower panel presents the main data sources.

3. Empirical strategy

3.1. *Empirical model*

The objective of this study is to evaluate the impact of metro network expansion on residential population patterns. It is worth noting that during the period analyzed, the length of the network doubled and extended beyond the central municipality into less densely populated areas. The scale of this expansion allows us to analyze its consequences by distinguishing the impact within the central municipality—an area that was densely populated in 1998 and already had high metro availability—from that in adjacent municipalities, areas where the metro was introduced for the first time. The impact of the opening of a new metro station on location decisions is not immediate; instead, a minimum number of years is required. The overlap of network changes during the period analyzed, along with the challenges of capturing the dynamics of these changes, has led us to estimate the impact based on the variation in population between the first and last year of the sample. Accordingly, the estimated equation captures the change in population and accessibility between 1998 and 2020, a period of 23 years.

The estimated equation takes the following form:

$$\Delta Pop_{it} = \beta_0 + \beta_1 \Delta(A_{it}) + \beta_2 \Delta(ramp_{it}) + X'\beta + \eta_{it} \tag{6}$$

Where

- $\Delta\ Pop_{it} = Pop_{i,2020} - Pop_{i,1998}$
- $\Delta(A_{it}) = accessibility_{i,2020} - accessibility_{i,1998}$ (Alternatively, accessibility could take the form of ΔED_{it}, as presented in Section 2.3.)
- $\Delta(ramp_{it}) = ramp_{i,2017} - Ramp_{i,1990}$ (distance to the closest ramp entrance)
- X' = Includes: initial population, area of the census tract and distance to the CBD.

Over this extended period, urban developments were carried out in previously depopulated areas of the Madrid metropolitan region (in 1998), which subsequently attracted population. These developments were often accompanied by metro extensions into these areas. To avoid clearly endogenous relationships, all areas without population in the first year of the sample were excluded. Likewise, areas without residential land use were also excluded.

To address the identification problem that arises since improvements in rail and road systems are expected to be endogenous to total population growth [1, 9, 10], we employ a Two-Stage Least Squares method with the introduction of instrumental variables. First, we calculate the value of the variation in accessibility with Equation 7.

As stated in the literature [1, 9, 10], improvements in rail and road systems may be potentially endogenous to total population growth. Endogeneity concerns can arise from reverse causation or measurement errors in the endogenous explanatory variables. In terms of reverse causation, policymakers may choose to invest in regions with rising projected population growth, or alternatively, in areas with poor prospects. Measurement errors in the accessibility variable can also introduce bias, particularly when walking distances are excluded from the calculations, frequencies and speeds are assumed to be constant, or when errors may exist in the GTFS files. If such measurement errors are correlated with the error term, estimates from a standard OLS model will be biased downward. To address this potential issue, we employ the Two-Stage Least Squares (2SLS) method, introducing instrumental variables. In the first stage, we estimate the variation in accessibility using Equation 7.

$$\Delta(A_{it}) = \theta_0 + \theta_1 Z_{1,i} + \theta_2 Z_{2,i} + \theta_3 \Delta(ramp_{it}) + X'\beta + \mu_{it} \tag{7}$$

Here, $Z_{1,i}$ and $Z_{2,i}$ denote the two instruments used, which are discussed in the following subsection. Then, using Equation 8, we estimate the relationship between the variation in total population and the predicted values of accessibility ($\widehat{\Delta(A_{it})}$) and distance to highway access points ($\Delta\widehat{(ramp_{it})}$).

$$\Delta Pop_{it} = \beta_0 + \beta_1 \widehat{\Delta(A_{it})} + \beta_2 \Delta\widehat{(ramp_{it})} + X'\beta + \eta_{it} \tag{8}$$

3.2. Instruments

The primary instrument is based on land permeability, reflecting construction feasibility and cost considerations during the expansion of the rail network. In addition, to address the potential endogeneity of highway expansions, we include an instrument grounded in historical geography: the radial post-road system initiated in 1760 under the Bourbon monarchy, which still influences the layout of modern highways in Spain.

3.2.1. Subsoil permeability

Lithology and tunnel construction

Modern tunnel construction is a complicated engineering operation that must take into consideration technical features of the subsoil, such as rock strength, water inflow, and ground support needs, in addition to other variables, such as archaeology, risk management, and so on.

The development in Spain of tunneling has been inextricably linked to the expansion of the country's rail network. Initially, with the extension of freight trains from Madrid to the ports; later, in the early twentieth century, with the establishment and quick expansion of urban rail networks [22]. The Madrid Metro is a prime illustration of this connection. Tunnels account for 78% of the total length of the system, or approximately 230 kilometers. As a reference, in New York, it is 59%, while in London, it is 45%. Several underground parts of the Madrid Metro system are among the longest of its kind in Europe (Lines 12 and 7, with 42 and 32 kilometers of tunnels, respectively).

We propose employing geological characteristics to capture the likelihood of building tunnels for the reasons stated above. By considering the types of rock (lithology) present in the subsoil, we incorporate parameters linked to water input and rock strength. More specifically, we use the soil permeability classification, which also allow us to take into account the stability of the ground. The Geological and Mining Institute of Spain (IGME) supplies vector data for this, which sets us apart from other studies that use raster data (for example, [3] use a 1-km raster for France). Specifically, we were provided access to continuous lithostratigraphic and permeability maps. As far as we know, this information has never been used in this context.

Rock permeability

As mentioned, our instrument incorporates rock permeability. The permeability of a rock indicates how quickly a fluid can penetrate it. If the permeability is high, rainwater will easily permeate the pores. If the permeability is low, however, rainwater will tend to accumulate on the surface or, if the terrain is sloped, move along it.

Table 4: Lithology and subsoil permeability. IGME classification.

	Permeability				
Lithology	Highly Permeable	Permeable	Medium	Low	Very Low
Carbonated	C-HP	C-P	C-M	C-L	C-VL
Detrital (Quaternary)	Q-HP	Q-P	Q-M	Q-L	Q-VL
Detrital	D-HP	D-P	D-M	D-L	D-VL
Volcanic (pyroclastic and lava)	V-HP	V-P	V-M	V-L	V-VL
Meta-Detrital	M-HP	M-P	M-M	M-L	M-VL
Igmeas	I-HP	I-P	I-M	I-L	I-VL
Evaporitic	E-HP	E-P	E-M	E-L	E-VL

As shown in Table 4, the IGME lithological classification divides rock types based on their permeability, from highly permeable to very low permeability. The less permeable the rock found in the subsoil, the more likely it is to build a tunnel. The more permeable the rock, the more challenging it is to construct a tunnel due to the increased risk of infiltration. In other words, the closer a census tract is to a low or very low permeable rock system, the more likely it is that a subway will be constructed due the relation of this with stability, costs, duration of work, etc.

The fact that we chose the low and very low permeable systems of rocks, means that we consider rocks such as ampelitic slates, hercynian plutonic rocks, quartzites, among others, that have lower levels of permeability. In following IGME´s classification, we choose all the rocks contained in the following groups (see Table 4 and Appendix E): C-L, Q-L, D-L, V-L, M-L, I-L, E-L, C-VL, Q-VL, D-VL, V-VL, M-VL, I-VL, and E-VL.

In map 5, we depict the MMA divided by municipalities (with the Municipality of Madrid in the middle) and the spatial distribution of the system of rocks of low and very low permeability, according to the classification introduced above. The subsoil with lower permeability is mostly distributed between the southeast and the northwest of the Metro Area, though there are areas inside the Municipality of Madrid with such characteristics. However, most of the area of the Municipality of Madrid is actually built above soils with a medium level of permeability, making it a relatively stable subsoil for tunneling. In 5, the white categories are the remain ones[4]. In fact, since the medium level is the most predominant in the MMA, it was necessary to select either the higher or the lower categories of permeability. Nevertheless, the higher categories also could be related with the presence of bodies of water (both under and on the ground) which would have the opposite effect as an instrument since it would predict a positive relation with urban agglomerations while we would expect an opposite effect from the permeability levels (as a high permeability makes it less likely to build a tunnel). That's why we chose the lower permeability categories of rocks.

Our instrument is the distance between each census tract's centroid and the nearest low and very low permeable rock system; we expect a positive relation as this facilitates nearby tunneling. We therefore assume that the smaller the distance, the greater the likelihood that a tunnel will be constructed in that tract (and in its region of influence). When analyzing census tracts, the instrument captures an individual effect that, when added, tells us where a tunnel could be constructed due to rock permeability conditions.

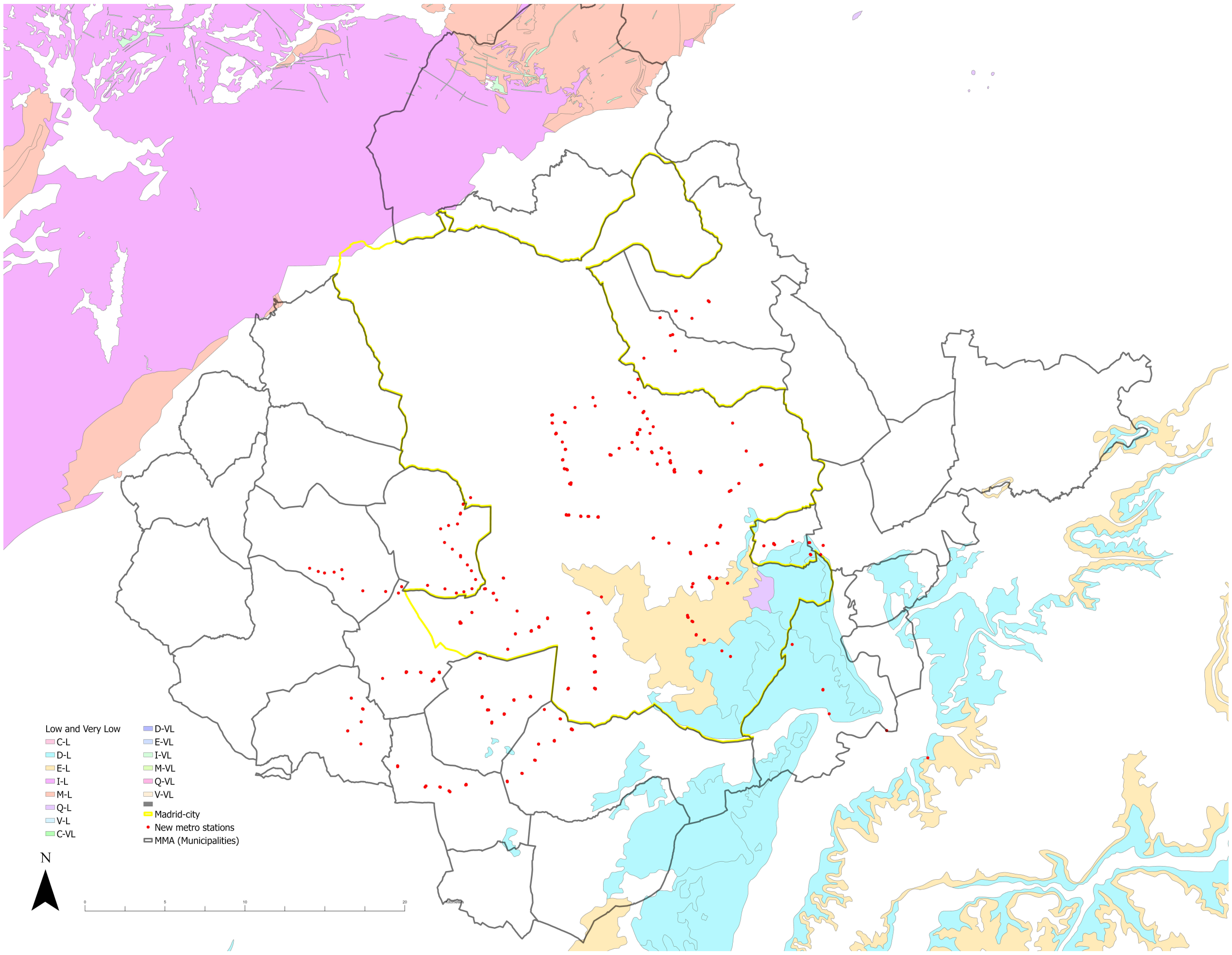
Validity of the instrument

To maintain the validity of our geological instrument, we must ensure that the distance between each census tract and the less permeable zones explains tunnel construction. Given that geological instruments have already been employed to quantify agglomeration economies [3, 6], it is vital to elaborate on what sets this apart. Recall that in this scenario, we are focusing on geological characteristics that directly explain tunneling, as these structures are the axis of the expansion of the rail system that we are analyzing. So, in our study, we use the characteristics of the rocks for different reasons than we would if we were explaining (for example) housing, total employment, or the location of firms. We specifically consider the presented instrument to be valid because:

1. We did not include bodies of water, which have been used to explain human density or as a control variable in agglomeration economics and urban form ([7], for example, use under groundwater). That is, we directly omitted all highly permeable terrain. This is crucial because highly permeable soil can be linked to water sources. In other words, urban centers may have emerged near local water sources and our instrument just consider the less permeable parts (see map E.8).
2. The relevance of the instrument, which we split into two categories: scale relevance and historical relevance. The former refers to the fact that the soil's geological factors such as

[4]In Appendix Appendix E, we provide a complimentary map (E.8) that presents the spatial distribution of the different levels of permeability (highly permeable, permeable, medium, low and, very low).

Figure 5: Lithology in the Madrid Metropolitan Area: Only Low (L) and Very Low (VL) permeable rocks.



Notes: Own development from continuous vector data from IGME.

average soil richness, presence of water, average stability indicators, and so on, are frequently used as aggregated values. We are using continuous data in our scenario, and each census tract has a unique value (the average length of a census tract in the MMA is 322 meters). This is important since, according to [15], aggregate instruments (including geological instruments), may be less useful for desegregated work evaluating urbanization at smaller geographical scales. [3], for example, use changes in mineral content to forecast population density since mineral-rich soil enhances food production, which is critical for human life. Although the assumption is important for measuring aggregate population density (e.g., municipal), it is ineffective for our geographical scale of research.

Second, in terms of historical relevance. The historical context suggests that geology (and more specifically, lithology) was much less understood at the time the municipalities of the MMA were established [5]. For this reason, the expansion of the population between 1998 and 2020 in the towns surrounding Madrid has already internalized the "established" conditions. In contrast, it is highly likely that zones with low levels of permeability were taken into account for in the planning and final placement of the tunnels (planned in the 1990s).

[5]The majority of the municipalities in the Madrid region have evidence of human settlement in the modern, middle, and even ancient periods. The only exceptions are two completely contemporaneous municipalities. First, the 1980-founded municipality of Tres Cantos, which is not served by the Metro or light rail, and second, the municipality of Rivas Vaciamadrid, which is connected to the Metro via external trains at ground level (i.e., without tunnels).

3. Local versus regional. We may deduce from map 5 that soil features have a regional influence that extends beyond the local dimension. That is, if soils containing less or highly permeable rocks cross a census tract, it has no direct effect on the population growth because (ceteris paribus) you can always decide to build in another part of the same town, because there is an incentive to build close to an already established city. However, the closeness to a less permeable soil system does influence the choice to build a tunnel, given the regional implications such tunnels have (recall that most of the metro expansions and tunnels we are evaluating connect several municipalities, covering regional spaces).

3.2.2. 1760 Bourbon Road Network.

The second instrument builds on the 1760 Bourbon road network, designed to connect all regions of Spain radially through Madrid. These post roads, established over 250 years ago, formed the first centralized infrastructure system in Spain and shaped subsequent national transportation planning. The instrument was first developed for Spain by [11], who showed that modern highways in Spain largely follow this historical structure: each additional 1760 Bourbon road ray from Madrid significantly predicts the presence of a modern highway, particularly in central cities and surrounding municipalities.

This instrument is considered exogenous for two key reasons. First, the 1760 roads were planned based on administrative centralization rather than demographic or economic forecasts, decoupling them from current population trends. Second, they predate contemporary urban expansion and transportation investment by more than two centuries, satisfying the exclusion restriction when controlling for geography and long-run historical population trends.

We constructed a variable that captures the distance from each census tract to the nearest segment of the historical 1760 Bourbon road network, using the dataset developed by [11]. The resulting variable (the Euclidean distance in meters from the tract centroid to the nearest route) has been used in previous studies to examine the effects of road expansion. In our case, it serves as a measure of baseline connectivity to the historical road system.

4. Results

4.1. OLS results

Table 5 presents the main results from estimating Equation 6, using both our custom-developed measure of accessibility and, for comparison, the Effective Density measure. As stated previously, the OLS results serve as a reference point; however, due to potential correlation between accessibility and the error term introduces a risk of bias in our coefficient of interest. Columns 1 and 6 present the results for the whole metropolitan area, columns 2 and 7 for the Municipality of Madrid, and the rest of the columns for the metro area excluding the Municipality of Madrid.

Regarding our preferred accessibility measure (columns 1–5), the coefficients are positive in every case. However, they are significant only when we omit the effect on the Municipality of Madrid and focus on the effect on the rest of the metro area. In column 5 we present the results adding the tracts where population was zero in 1998: the coefficient for accessibility increases, suggesting a potential problem of endogeneity. As mentioned before, we exclude non-residential areas from the analysis since the population dynamics may respond to other causes, such as illegal settlements or slow cadastral updates. Nonetheless, we checked whether the exclusion of these zones has any impact on the results. Comparing columns 3 and 4 (and columns 8 and 9), we can observe that including non-residential areas does not significantly change the estimated coefficient for the accessibility measure.

Furthermore, when using effective density (columns 6 to 10), the results are confirmed. The coefficients are always positive, and the changes go in the same direction. However, since accessibility is defined differently, the coefficients are not directly comparable.

Table 5: Madrid, 2020-1998 Δ total population. OLS results.

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Δ accessibility		**.00007**	**.00006**	**.000728**	**.000732**	**.00088**					
	a	*(.00011)*	*(.00013)*	*(.000327)*	*(.000330)*	*(.00034)*					
	b	*.570*	*.631*	*.026*	*.027*	*.010*					
Δ ED							**.00204**	**.00054**	**.01187**	**.01181**	**.01578**
	a						(.00122)	(.00131)	(.00355)	(.00359)	(.00379)
	b						*.095*	*.684*	*.001*	*.001*	*.000*
Initial population		x	x	x	x	x	x	x	x	x	x
Distance to CBD		x	x	x	x	x	x	x	x	x	x
Distance to ramp		x	x	x	x	x	x	x	x	x	x
Area census tract		x	x	x	x	x	x	x	x	x	x
M.Area (not Madrid-Muni.)		x		x	x	x	x		x	x	x
Municipality of Madrid		x	x				x	x			
Only Residential Land		Y	Y	Y	N	Y	Y	Y	Y	N	Y
Population>0 in 1998		Y	Y	Y	Y	N	Y	Y	Y	Y	N
adjusted-R2		0.38	0.300	0.493	0.483	0.721	0.38	0.300	0.490	0.486	0.723
No. of observations		*3,246*	*2,176*	*1,070*	*1,093*	*1,333*	*3,246*	*2,176*	*1,070*	*1,093*	*1,333*

Notes: Distance to ramp = 1990–2017 variation of the distance to the closest highway access ramp. M. area= metropolitan area of Madrid. [a] = Standard deviation [b]= *P*-value. We use decimal points. Appendix F.18 contains more detailed results.

4.2. TSLS results

Table 6 shows the main results for the estimated Equation 8 using a TSLS method. We use the instruments presented in Section 3.2. With a similar structure as that of Table 5, we compare the results of accessibility and ED as well as an equation that includes non-residential census tracts in column 4 (and 9) and another one that includes census tracts with population zero in 1998 in column 5 (and 10).

As in the OLS, the changes in accessibility have a positive and significant impact on the population growth when the Municipality of Madrid is excluded from the sample. The difference in the coefficient, in our preferred specification in column 3, from its OLS counterpart (0.0042 and 0.0007 respectively) suggests that the expansion of the rail system in Madrid is endogenous, which is corroborated by the difference between the effective density coefficients (TSLS of 0.04614 and OLS of 0.01187).

The estimated coefficient for our preferred specification — column 3 — is higher than the one obtained using OLS, which suggests that the OLS estimate is biased downward. This finding is further supported when the initial definition of accessibility is replaced with that of effective density — column 8. As discussed in Section 3.1, the increase in the coefficient's magnitude when using instrumental variables is likely driven by measurement error in the accessibility variable. In this context, TSLS corrects the downward bias inherent in the OLS estimates.

Moreover, regressing changes in accessibility on changes in population reveals a negative correlation: census tracts with lower population growth tend to receive more rail stations. This result is consistent with the findings of the studies by [7] and [9], which analyzed the effects of highway expansions in U.S. cities and Barcelona, respectively, using a similar empirical framework.

Another reason supporting the notion that OLS underestimates the true effect is the potential omission of relevant variables — such as municipal amenities or local regulations — that are negatively correlated with our regressor of interest.

Overall, our findings align with prior studies that show the estimated coefficient increases when an exogenous source of variation (i.e., instrumental variables) is used to address potential biases in OLS estimates. That is the case with [9], who observed that the difference between OLS and TSLS is between 2.5 and 3.7 times larger in railroad and highway coefficients in non-central areas when explaining population growth in Barcelona between 1991 and 2001. Furthermore, [1] demonstrates that the coefficients are between 1.7 and 2.8 times larger when studying the variation of growth of the population in the central city due to changes in the system of highways in the United States between 1950 and 1990.

The results of the first-stage regressions (following Equation 7) are presented in Table 7. Both instruments are statistically significant and exhibit the expected sign. Since the model is exactly identified, weak identification critical values are not applicable. Nonetheless, the reported first-stage F-statistics confirm that the instruments are strongly correlated with their respective endogenous variables. As before, we use the ED results as a robustness check for our main accessibility measure.

In line with the OLS results, when we consider non-residential land (columns 4 and 8) there is no statistical difference from the results that did not consider them.

Table 6: Madrid, 2020–1998 Δ total population. TSLS results.

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Δ accessibility		**-.00187**	**.00017**	**.00425**	**.00428**	**.00541**					
	a	*(.00105)*	*(.00055)*	*(.00220)*	*(.00219)*	*(.00191)*					
	b	*.074*	*.761*	*.053*	*.051*	*.005*					
Δ ED							**-.01962**	**.00176**	**.04614**	**.04644**	**.05606**
	a						*(.01100)*	*(.00577)*	*(.02354)*	*(.02343)*	*(.01944)*
	b						*.075*	*.761*	*.050*	*.047*	*.004*
IV: Permeability		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IV: 1760 Roads		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Initial population		x	x	x	x	x	x	x	x	x	x
Distance to CBD		x	x	x	x	x	x	x	x	x	x
Distance to ramp		x	x	x	x	x	x	x	x	x	x
Area census tract		x	x	x	x	x	x	x	x	x	x
Metro Area		x		x	x	x	x		x	x	x
Muni. of Madrid		x	x				x	x			
Only Residential Land		Y	Y	Y	N	Y	Y	Y	Y	N	Y
Population>0 in 1998		Y	Y	Y	Y	N	Y	Y	Y	Y	N
No. of observations		*3,246*	*2,176*	*1,070*	*1,093*	*1,333*	*3,246*	*2,176*	*1,070*	*1,093*	*1,333*

Notes: Distance to ramp = 1990–2017 variation of the distance to the closest highway access ramp. Permeability = closest distance in meters to *Low* and *Very Low*. Metro area = metropolitan area of Madrid. First-stage *F*-statistics are reported for each endogenous variable in Table 7. [a] = Standard error. [b] = *P*-value. We use decimal points. Appendix G.19 contains more detailed results.

4.2.1. *Results by country of birth and age group*

Up to this point, we have estimated the average impact for the whole population. However, it is interesting to look for potential heterogeneity between population groups. Hence, we test whether different impacts are found according to age or country of birth. According to age, we distinguish between three groups: people between the ages of 20 and 39, 40 and 59, and over 60. The countries

Table 7: Madrid*, 2020-1998 Δ total population. TSLS. First-Stage results.

Dependent	1998–2020 Δ accessibility		1998–2020 Δ ED	
	[1]	[2]	[3]	[4]
Low Permeability (meters)	**36.525**	**36.198**	**3.290**	**3.582**
^{a}	*(6.704)*	*(6.710)*	*(551)*	*(474)*
^{b}	*0.000*	*0.000*	*0.000*	*0.000*
^{c}	*20.00*	*7.56*	*25.34*	*39.07*
Historic Routes (meters)	**0.5552**	**0.5417**	**0.5552**	**0.5437**
^{a}	*(0.0778)*	*(0.0778)*	*(0.0545)*	*(0.0694)*
^{b}	*0.000*	*0.000*	*0.000*	*0.000*
^{c}	*35.55*	*30.21*	*35.55*	*42.74*
Initial population	x	x	x	x
Distance to CBD	x	x	x	x
Distance to ramp	x	x	x	x
Area census tract	x	x	x	x
Non-Residential Land		✓		✓
No. of observations	*1,070*	*1,333*	*1,070*	*1,333*

Notes: * Madrid metropolitan area excluding the Municipality of Madrid. Distance to ramp: 1990–2017 variation of the distance to the closest highway access ramp. Permeability: closest distance in meters to *Low* and *Very Low* soil. ^{a} = Standard deviation. ^{b}= *P*-value. ^{c}= F test of excluded instrument.

of birth under consideration are Spanish (native) and foreign born (non-native). We had access to detailed information at the census tract level in both cases, as explained in Section 2.

Table 8 summarizes the main findings of our study on the heterogeneous effects of changes in accessibility in intra-urban mobility patterns of the population. We estimate Equation 8 for different groups of population within the census tracts. Columns 1 to 5 display the results for accessibility, while columns 6 to 9 (lower part of the table) for ED.

We found that the population between 40 and 59 years of age drives the average results from Table 6, followed by the population between 20 and 40 years old (although not statistically significant). We found no evidence that the variation in accessibility causes people over the age of 60 to relocate within the metropolitan area, which suggests that population mobility due to an expansion of a transport system has a limit in terms of the life cycle, and younger people are more willing to relocate.

In terms of country of birth, we discovered that the Spanish population grows the most when accessibility changes, implying that the expansion of the rail network has an impact on housing sorting for this group. On the other hand, despite the fact that the number of foreign-born people increased the most during the study period, we found no evidence that the location of this group was influenced by the increase in accessibility. One possible explanation is that immigrants' residences are more volatile. Furthermore, the developments surrounding the new metro stations may be prohibitively expensive for this demographic group. We do not provide causal evidence for this behavior, but keep in mind that we measure accessibility in terms of reachable job opportunities, so it is expected that foreign-born people would wish to live closer to a metro station; since we do not observe this, this means that the expansion of the rail network in Madrid did not promote their accessibility, forcing them to seek housing solutions far from the trains.

4.2.2. *Marginal Effects*

The functional form of the two accessibility measures used makes it difficult to interpret the magnitude of the corresponding estimated coefficients. In order to provide an order of magnitude for the effect of increased accessibility we compute the marginal effects. To do so, we extend the

Table 8: Madrid*, 2020–1998 Δ total population. TSLS results. Heterogeneous results by age group and country of birth.

Age-group / Birth	20–39	40–59	>60	Spain	Foreign
	[1]	[2]	[3]	[4]	[5]
Δ accessibility	**.00104**	**.00300**	**.00005**	**.00463**	**-.00039**
a	*(.00082)*	*(.00098)*	*(.00057)*	*(.00208)*	*(.00045)*
b	*.206*	*.002*	*.936*	*.026*	*.390*
IV: Permeability	✓	✓	✓	✓	✓
IV: 1760 Roads	✓	✓	✓	✓	✓
Initial population	x	x	x	x	x
Distance to CBD	x	x	x	x	x
Distance to ramp	x	x	x	x	x
Area census tract	x	x	x	x	x
No. of observations	*1,070*	*1,070*	*1,070*	*1,070*	*1,070*
Age-group / Birth	**20–39**	**40–59**	**>60**	**Spain**	**Foreign**
	[6]	[7]	[8]	[9]	[10]
Δ ED	**.01127**	**.03257**	**.00050**	**.05033**	**-.00419**
a	*(.00880)*	*(.01050)*	*(.00619)*	*(.02214)*	*(.00482)*
b	*.200*	*.002*	*.936*	*.023*	*.385*
IV: Permeability	✓	✓	✓	✓	✓
IV: 1760 Roads	✓	✓	✓	✓	✓
Initial population	x	x	x	x	x
Distance to CBD	x	x	x	x	x
Distance to ramp	x	x	x	x	x
Area census tract	x	x	x	x	x
No. of observations	*1,070*	*1,070*	*1,070*	*1,070*	*1,070*

Notes: * Madrid metropolitan area excluding the Municipality of Madrid. Distance to ramp = 1990–2017 variation of the distance to the closest highway access ramp. Permeability = closest distance in meters to *Low* and *Very Low* soil. First-stage *F*-statistics are reported for each endogenous variable in Table 7. [a] = Standard error. [b] = *P*-value.

analysis by simulating various travel times to get a sense of the marginal effects of our model.

We simulate changes in travel times, reducing them by 1%, 2%, 5%, and 10%. Then, we replace them in Equations 1 and 4. Following that, we compare each of these simulated accessibilities to the original (with actual travel times) to determine the impact of changes in accessibility on the number of inhabitants.

Table 9 displays the marginal effects for the total population. We prefer to rely on our main measure of accessibility because it reduces the weight of employment with distance; however, for the same reason, ED may be considered an upper limit for the marginal effect. According to our estimates, a 1% decrease in travel times on Madrid's rail system (between suburban municipalities and the Municipality of Madrid) would result in an increase in total population by from 11 to 56 inhabitants per census tract. A 10% reduction in travel times would result in an increase in population from 108 to 147 by census tract. The elasticity of population with respect to time is equal to 0.7 (for the exponential accessibility) and remains almost constant when the magnitude of the change in travel time varies[6].

Next, we calculate the marginal effects of reducing travel times by 1% for the previously studied heterogeneous groups. Table 10 presents the results by age group and country of birth. The Spanish

[6]We calculate it as a ratio, the percentage variation of population/percentage change of time, for the sample averages.

Table 9: Marginal analysis: 2020–1998 Δ of total population with different travel times. TSLS results with geological IV.

	Δ accessibility		Δ ED	
Travel time reduction	Mean [1]	Std. Dev. [2]	Mean [1]	Std. Dev. [2]
1%	10.40	4.30	52.61	21.00
2%	20.88	8.72	61.62	23.55
5%	52.67	22.75	89.49	32.89
10%	108.49	48.11	137.50	51.47

population would grow by from 11 to 57 people as a result of changes in access to job opportunities in Madrid. This is higher than the overall population average and vastly different from the situation faced by the non-Spanish.

When it comes to housing sorting near a train station, we found that the marginal effects for the group between 40 and 59 years old could increase by between 7 and 36 the number of people living in a census tract if travel times were reduced by 1% due to improvements in the rail infrastructure.

Table 10: Marginal analysis: 2020-1998 Δ of heterogeneous population groups with Δ of 1% of the travel times. TSLS results with geological IV.

	Δ accessibility		Δ ED	
Age group / Birth	[1] Mean	[2] Std. Dev.	[3] Mean	[4] Std. Dev.
20–39	2.54	1.04	13.06	5.00
40–59	7.29	3.05	36.31	15.28
>60	0.11	.04	0.58	0.21
Spain	11.33	4.67	57.26	22.70
Foreign	-.94	.38	-4.94	1.83

4.2.3. Robustness

Density of population

We propose measuring population density as an alternative dependent variable to complement the understanding of the population dynamics. The density in fixed-area zones is expected to increase as the total population grows. An increased density can also be an indicator of productivity, innovation, access to goods and services, and, in general, an increase in the benefits of agglomeration economies, in addition to providing strong support for our main result [7].

Table 11 summarizes the main findings. Columns 1 through 3 present the average results. In line with the results in Tables 5 and 6, we get the expected effect and the difference between the two coefficients (OLS and TSLS) suggests that there is an downward bias to be considered, which in any case remains when we replace the dependent variable.

When distinguishing between age groups and country of birth (columns 4 to 7), the coefficients of the accessibility variables are positive and statistically significant. This means that improved accessibility not only explains population growth, but also increases the population density near

[7]This study does not intend to investigate the effects of an expansion of transportation infrastructure on productivity or agglomeration economies, but it does open the door for future research in the case of Madrid's expansion of its rail system.

the train stations. This tendency is mostly evident among the Spanish population: In the main specification (Table 8, column 4), the coefficient is not statistically different from the total population coefficient (Table 6, column 3), but is twice as large if we make the same comparison for population density (Table 11 columns 2 and 4). As mentioned by [5], anything that makes a city more appealing (such as increased accessibility) attracts people from other places, putting upward pressure on house prices, resulting in higher land prices that many foreign-born families cannot afford.

Table 11: Madrid*, 2020-1998 Δ density of population. TSLS results with geological IV.

Birth/Age-group	Total	Total	Total	Spain	20–39	40–59
	[1] OLS	[2] TSLS	[3] TSLS	[4] TSLS	[5] TSLS	[6] TSLS
Δ accessibility	**.01646**	**.06054**		**.13864**	**.06399**	**.04028**
a	*(.00794)*	*(.03212)*		*(.04281)*	*(.02645)*	*(.00853)*
b	*.039*	*.059*		*.001*	*.016*	*.000*
Δ ED			**.65764**			
a			*(.34182)*			
b			*.054*			
IV: Permeability		✓	✓	✓	✓	✓
IV: 1,700 Routes		✓	✓	✓	✓	✓
Initial population	x	x	x	x	x	x
Distance to CBD	x	x	x	x	x	x
Distance to ramp	x	x	x	x	x	x
Area census tract	x	x	x	x	x	x
No. of observations	*1,070*	*1,070*	*1,070*	*1,070*	*1,070*	*1070*
R2	0.219	N/A	N/A	N/A	N/A	N/A

Notes: * Madrid metropolitan area excluding the Municipality of Madrid. Distance to ramp= 1990–2017 variation of the distance to the closest highway access ramp. Permeability= closest distance in meters to *Low* and *Very Low* soil. First-stage *F*-statistics are reported for each endogenous variable in Table 7. *a* = Standard deviation *b*= *P*-value. We did not consider the area of the census tract as a control variable.

Connectivity vs. accessibility: Euclidean distance

Connectivity to a transportation network is typically measured as the distance in meters (Euclidean distance) to the closest station, stop, terminal, etc. ([1, 9] provide examples of its application). Consequently, in our study we also use variations in the Euclidian distance to the nearest train station as a robustness test.

Thus, in this subsection, we show the results of replacing our primary right-hand variables (accessibility or ED) with the change in the Euclidean distance (connectivity). Table 12 displays the main results.

The coefficients are negative, indicating that a reduction in the distance to the nearest station produces an increase of population in the census tract. Overall, the results are consistent with those found in our main specification: total population coefficients are significant and driven by the Spanish population and the age group of 40 to 59. We found a significant coefficient at 90% for the population aged 20–39.

The findings support our main hypothesis: whether measured by changes in connectivity or accessibility, improvements in the rail infrastructure in the Madrid Metropolitan area resulted in an increase in total population in the suburbs of the region.

Table 12: Madrid*, 2020–1998 Δ total population. Euclidean distance, TSLS results with geological IV.

Birth/Age-group	Total	Total	Spain	For.	20–39	40–59	>60
	[1] OLS	[2] TSLS	[4] TSLS	[5] TSLS	[6] TSLS	[7] TSLS	[8] TSLS
Δ Connectivity	**-.04863**	**-.04852**	**-.05292**	**.00441**	**-.01185**	**-.03425**	**-.00053**
a	*(.01060)*	*(.02232)*	*(.02014)*	*(.00498)*	*(.00866)*	*(.01022)*	*(.00652)*
b	*.000*	*.030*	*.009*	*.376*	*.171*	*.001*	*.936*
IV: Permeability	N/A	✓	✓	✓	✓	✓	✓
IV: 1760 Roads	N/A	✓	✓	✓	✓	✓	✓
Initial population	x	x	x	x	x	x	x
Distance to CBD	x	x	x	x	x	x	x
Distance to ramp	x	x	x	x	x	x	x
Area census tract	x	x	x	x	x	x	x
No. of observations	*1,070*	*1,070*	*1,070*	*1,070*	*1,070*	*1,070*	*1,070*
R^2	0.507	N/A	N/A	N/A	N/A	N/A	N/A

Notes: * Madrid metropolitan area excluding the Municipality of Madrid. Δ Connectivity is measured using Euclidean distance (Euc_98_20_distance). Distance to ramp = 1990–2017 variation of the distance to the closest highway access ramp. Permeability = closest distance in meters to *Low* and *Very Low* soil. ^{a} = Standard deviation, ^{b} = *P*-value. First-stage *F*-statistics are reported for each endogenous variable in Table 7.

5. Conclusions

Our study has demonstrated that the expansion of the rail system in the Madrid metropolitan area has had a significant impact on its population dynamics. The method employed is a long-run difference (between 1998 and 2020) which uses accessibility measures based on General Transit Feed Specification (GTFS) files (real-time travel matrix) and an instrumental variable based on geological features. This framework allows us to identify a positive relation between improvement in accessibility and population patterns.

This study relies on the use of a geological instrument grounded in subsoil permeability, offering a novel and highly relevant proxy for tunneling feasibility. This methodological innovation strengthens the causal identification strategy by directly tying the instrument to engineering constraints in metro construction.

Specifically, our results suggest that a 1% reduction in travel times (due to the opening of new train stations) increases the total number of residents in a census tract by 10 over the 23-year study period. A 10% increase, in turn, leads to 108 additional residents. (On average, a census tract outside the Municipality of Madrid had 1,541 residents in 2020.) This corresponds to a population elasticity with respect to travel time of 0.7.

Further analysis of demographic heterogeneity shows that a 1% improvement in travel times leads to an increase of seven residents aged 40–59, and three residents aged 20–39. We also find evidence that Spanish-born individuals tend to locate near new train stations: on average, eleven Spanish-born individuals moved into the neighborhood of a new station for every 1% improvement in travel times. This pattern likely reflects greater access to jobs in the Municipality of Madrid. In contrast, increased accessibility does not attract foreign-born residents. In fact, we find a small negative effect.

We complemented this analysis with a study of population density and found that improvements in accessibility affect Spanish-born residents more than the population overall—more than doubling the effect. This suggests that agglomeration dynamics are particularly strong among this group, who prefer to reside close to their counterparts.

In line with previous studies, we also confirm that the change in distance to the closest train station is a valid measure of transportation connectivity. However, this measure mainly captures local effects, unlike the accessibility metric, which accounts for the influence of the entire network.

Our results are necessarily based on a series of assumptions that mainly affect the construction of the data. Specifically, the variables of train speed and frequency are held constant; non-residential areas and those with no population in the initial year have been excluded; and aerial interpolation was required to standardize the spatial units. Additionally, although the estimation has accounted for potential endogeneity issues related to the accessibility variable by using instruments that meet commonly accepted first-stage tests, the possibility that some estimation bias remains cannot be entirely ruled out.

From a policy perspective, these findings have crucial implications for urban planning and transport infrastructure development. Understanding how accessibility improvements shape population growth can inform decisions related to rail system expansion, housing policy, and city development. Policymakers should consider the distinct needs and preferences of different demographic groups to guide inclusive urban growth and avoid reinforcing disparities. Although we do not provide direct evidence of segregation, our results show that Spanish-born residents cluster near new train stations, while the fastest-growing segment—foreign-born residents—does not.

References

[1] Nathaniel Baum-Snow, *Suburbanization and Transportation in the Monocentric Model*, Journal of Urban Economics, Vol. 62, No. 3, pp. 405–423, November 2007.

[2] Francisco Calvo, Juan de Oña, and Fernando Arán, *Impact of the Madrid subway on population settlement and land use*, Land Use Policy, Vol. 31, pp. 627-639, 2013.

[3] Pierre-Philippe Combes, Gilles Duranton, Laurent Gobillon, and Sébastien Roux, *Estimating agglomeration economies with history, geology, and worker effects*, Agglomeration Economics, University of Chicago Press, pp. 15–66, 2010.

[4] Comunidad de Madrid, *Atlas de la comunidad de Madrid en el umbral del siglo XXI. Imagen socioeconómica de una región receptora de inmigrantes*, Editorial Complutense, S.A., 1st edition, 2002.

[5] Gilles Duranton and Diego Puga, *The economics of urban density*, Journal of Economic Perspectives, Vol. 34, No. 3, pp. 3–26, 2020.

[6] Gilles Duranton and Matthew A. Turner, *The Fundamental Law of Road Congestion: Evidence From US Cities*, American Economic Review, Vol. 101, No. 6, pp. 2616–2652, October 2011.

[7] Gilles Duranton and Matthew A. Turner, *Urban Growth and Transportation*, Review of Economic Studies, Vol. 79, No. 4, pp. 1407–1440, 2012.

[8] Masahisa Fujita and Hideaki Ogawa, *Multiple equilibria and structural transition of non-monocentric urban configurations*, Regional Science and Urban Economics, Vol. 12, No. 2, pp. 161–196, 1982.

[9] Miquel-Àngel Garcia-López, *Urban Spatial Structure, Suburbanization and Transportation in Barcelona*, Journal of Urban Economics, Vol. 72, No. 2–3, pp. 176–190, September–November 2012.

[10] Miquel-Àngel Garcia-López, Albert Solé-Ollé, and Elisabet Viladecans-Marsal, *Does zoning follow highways?*, Regional Science and Urban Economics, Vol. 53, pp. 148–155, July 2015.

[11] Miquel-Àngel García-López, Adelheid Holl, and Elisabet Viladecans-Marsal, *Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities*, Journal of Urban Economics, Vol. 85, pp. 52–67, 2015.

[12] Karst T. Geurs and Bert van Wee, *Accessibility evaluation of land-use and transport strategies: review and research directions*, Journal of Transport Geography, Vol. 12, No. 2, pp. 127–140, 2004.

[13] Marco Gonzalez-Navarro and Matthew A. Turner, *Subways and urban growth: Evidence from earth*, Journal of Urban Economics, Vol. 108, pp. 85–106, 2018.

[14] Daniel J. Graham, *Agglomeration, Productivity and Transport Investment*, Journal of Transport Economics and Policy, Vol. 41, No. 3, pp. 317–343, 2007.

[15] Daniel J. Graham, Patricia S. Melo, Piyapong Jiwattanakulpaisarn, and Robert B. Noland, *Testing for causality between productivity and agglomeration economies*, Journal of Regional Science, Vol. 50, No. 5, pp. 935–951, 2010.

[16] Walter G. Hansen, *How Accessibility Shapes Land Use*, Journal of the American Institute of Planners, Vol. 25, No. 2, pp. 73–76, 1959.

[17] Adelheid Holl, *Market Potential and Firm-Level Productivity in Spain*, Journal of Economic Geography, Vol. 12, No. 6, pp. 1191–1215, 2012.

[18] John R. Logan, Zengwang Xu, and Brian Stults, *Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database*, The Professional Geographer, Vol. 66, No. 3, pp. 412–420, July 2014.

[19] David Martin, Danny Dorling, and Richard Mitchell, *Linking Censuses through Time: Problems and Solutions*, Area, Vol. 34, No. 1, pp. 82–91, 2002.

[20] Anna Matas, Josep-Lluis Raymond, and Josep-Lluis Roig, *Wages and Accessibility: The Impact of Transport Infrastructure*, Regional Studies, Vol. 49, pp. 1236–1254, 2015.

[21] Thierry Mayer and Corentin Trevien, *The impact of urban public transportation: Evidence from the Paris region*, Journal of Urban Economics, Vol. 102, pp. 1–21, 2017.

[22] Ministerio de Fomento, *Tuneles de Espana. De la ingeniería medieval a los modernos túneles del siglo XXI.*, Revista del Ministerio de Fomento, Vol. 586, No. 6, 2009.

[23] Morton E. O'Kelly and Mark W. Horner, *Aggregate accessibility to population at the county level: US 1940-2000*, Journal of Geographical Systems, Vol. 5, No. 1, pp. 5, 2003.

[24] Stephen J. Redding and Matthew A. Turner, *Transportation Costs and the Spatial Organization of Economic Activity*, Handbook of Regional and Urban Economics, Vol. 5, pp. 1339–1398, 2015.

[25] Ignacio Tiznado-Aitken, Juan Carlos Muñoz, and Ricardo Hurtubia, *Public transport accessibility accounting for level of service and competition for urban opportunities: An equity analysis for education in Santiago de Chile*, Journal of Transport Geography, Vol. 90, pp. 102919, 2021.

[26] Lucia Mejia-Dorantes and Karen Lucas, *Public transport investment and local regeneration: A comparison of London's Jubilee Line Extension and the Madrid Metrosur*, Transport Policy, 35, 2014, pp. 241–252.

[27] Lucia Mejia-Dorantes, Antonio Paez, and José M. Vassallo, *Transportation infrastructure impacts on firm location: The effect of a new metro line in the suburbs of Madrid*, Journal of Transport Geography, 22, 2012, pp. 236–250.

[28] Lucia Mejia-Dorantes, Antonio Paez, and José M. Vassallo, *Analysis of house prices to assess economic impacts of new public transport infrastructure: Madrid Metro Line 12*, Transportation Research Record, 2245(1), 2011, pp. 131–139.

[29] Alexander Salov and Ekaterina Semerikova, *Transportation and urban spatial structure: Evidence from Paris*, Environment and Planning B: Urban Analytics and City Science, 51(6), 2024, pp. 1248–1273.

[30] Siqi Zheng and Camila Ramos, *Subway expansion, job accessibility improvements, and home value appreciation in four global cities: Considering both local and network effects*, Transport Policy, 116, 2022, pp. 11–21.

[31] Miquel-Àngel Garcia-López, Ilias Pasidis, and Elisabet Viladecans-Marsal, *Suburbanization and transportation in European cities*, Journal of Economic Geography, Vol. 24, No. 6, pp. 843–869, 2024.

Appendix A. Aerial Interpolation

When working with spatial data, and particularly when studying a dynamic effect, it is necessary to deal with the disparities that emerge due to changes of the boundaries of the administrative units, in this case, the census tracts. This is a common situation faced by researchers that use spatial information, yet the treatment depends on each case [18]. In this study, it is necessary to homologate the census tracts to the delimitation of 2020.

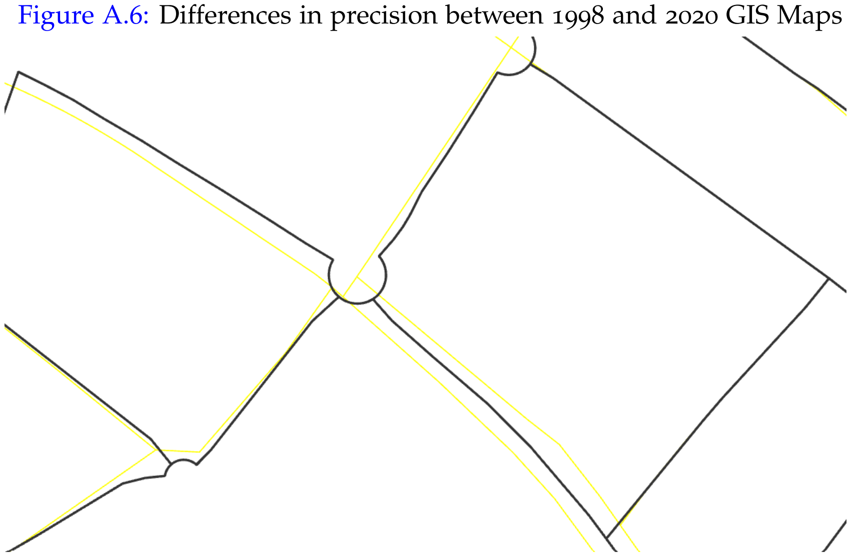
First, we compare each census tract to its boundaries in 2020 and apply an initial filter by area. If the difference in the area is greater than 2%, the census tract is interpolated. Otherwise, we assume that it is unchanged and that no interpolation is required. Second, we run the interpolation with area – weights, as in [18], which interpolates the census tracts of the decennial census of the United States from 1970 to the boundaries of 2010 using a combination of aerial and population interpolation.

As a result, to estimate the population in 1998 by tract, we directly assigned the official register (padrón) if the census tract has the same area as in 2020 (at least a 98% coincidence). We use the 95% and 90% cut-offs as a robustness test. The number of tracts did not change). Then, with the remaining census tracts, we conduct an aerial interpolation as follows:

$$InterpolatedPopulation_{i,1998} = \sum_{i=1}^{n} (\frac{area_\gamma}{area_i} \times Pop_{i,1998}) \tag{A.1}$$

Here, i indicates the census tract with 2020 boundaries, and γ indicates a tract in 1998 that intersects i. The population of the ith section is the sum of the parts in 1998 that make up that section. We run the process by year between 1998 and 2019, in order to have a picture of the evolution of the sizes; however, we only employ the result for 1998. In Table Appendix A this evolution is shown. In 1998, the number of new census tracts, compared to 2020, was 20% less, and 3% changed their size. We are aware that an aerial interpolation does not take into account the internal distribution of the population. However, we are using the smallest administrative unit available.

Figure A.6: Differences in precision between 1998 and 2020 GIS Maps



Notes: Own development from data from INE. Black lines correspond to 2020 boundaries and the yellow lines to 1998.

Table A.13: Region of Madrid: Aerial Interpolation. Census tracts between 1998 and 2020.

Year	Number of census tracts			%		
	Change size	Match	New	Change size	Match	New
1,998	149	3,406	862	0.03	0.77	0.20
1,999	150	3,503	764	0.03	0.79	0.17
2,000	150	3,503	764	0.03	0.79	0.17
2,001	149	3,565	703	0.03	0.81	0.16
2,002	135	3,578	704	0.03	0.81	0.16
2,003	135	3,578	704	0.03	0.81	0.16
2,004	126	3,646	645	0.03	0.83	0.15
2,005	125	3,685	607	0.03	0.83	0.14
2,006	114	3,806	497	0.03	0.86	0.11
2,007	91	3,889	437	0.02	0.88	0.10
2,008	87	3,937	393	0.02	0.89	0.09
2,009	80	3,992	345	0.02	0.90	0.08
2,010	71	4,094	252	0.02	0.93	0.06
2,011	64	4,159	194	0.01	0.94	0.04
2,012	64	4,163	190	0.01	0.94	0.04
2,013	54	4,221	142	0.01	0.96	0.03
2,014	41	4,261	115	0.01	0.96	0.03
2,015	20	4,305	92	0.00	0.97	0.02
2,016	16	4,309	92	0.00	0.98	0.02
2,017	15	4,313	89	0.00	0.98	0.02
2,018	4	4,409	4	0.00	1.00	0.00
2,019	1	4,416		0.00	1.00	0.00

Notes: Change in size: if the census tract's area differed by more than 2% from the previous year. We use 1997, which is why we have 1998 results. New: If there was no register of that census tract the previous year, usually due to a subdivision.

Figure A.7: Differences in the internal distribution of the same census tract

Notes: Own development from data from INE. Black lines correspond to 2020 boundaries and the yellow lines to 1998. Left: just 1998 boundaries. Right: 2020 boundaries on top of 1998 boundaries.

Appendix B. Employment Share by District

Table B.14: City of Madrid

District	2001	2011	2021
01. Centro	0.04596	0.05028	0.04862
02. Arganzuela	0.04846	0.05307	0.05202
03. Retiro	0.04211	0.03862	0.03603
04. Salamanca	0.04691	0.04747	0.04482
05. Chamartín	0.04551	0.04620	0.04545
06. Tetuán	0.04798	0.04942	0.05120
07. Chamberí	0.04840	0.04609	0.04424
08. Fuencarral - El Pardo	0.07101	0.07476	0.07403
09. Moncloa - Aravaca	0.03741	0.03654	0.03585
10. Latina	0.08392	0.07064	0.06730
11. Carabanchel	0.07180	0.07377	0.07439
12. Usera	0.03691	0.03897	0.03932
13. Puente de Vallecas	0.07184	0.06644	0.06717
14. Moratalaz	0.03564	0.02802	0.02646
15. Ciudad Lineal	0.07509	0.06696	0.06437
16. Hortaleza	0.05143	0.05579	0.05926
17. Villaverde	0.04094	0.04092	0.04364
18. Villa de Vallecas	0.02061	0.03103	0.03748
19. Vicálvaro	0.01899	0.02204	0.02375
20. San Blas	0.04519	0.04754	0.04895
21. Barajas	0.01388	0.01546	0.01567

Notes: Share of employment in each district as a proportion of total employment in the Municipality of Madrid, according to the 2001, 2011, and 2021 censuses.

Appendix C. Expansion Plans of the Metro of Madrid 1995–2020

This expansion occurred over two consecutive investment plans, with the following goals in mind:

1. To connect densely populated suburbs of Madrid to the subways.
2. To provide access to strategic areas (the existing airport, the Olympic Village related to the failed bid for the Olympic Games, or a new airport never constructed).
3. Construct a circular line with 27 stations connecting major suburban towns south of Madrid.
4. The extension to municipalities outside Madrid without a train connection.
5. The extension of the Madrid subway system to peripheral suburbs and relatively small towns near Madrid (with a population between 40,000 and 113,000).
6. The connection of new urban residential developments to the network and the construction of three intermediate stations on new lines.
7. The connection of the new residential developments located in the north and southeast of the Municipality of Madrid.

First Plan 1995–2003

Between 1995 and 2003, the first extension plan was implemented in two subperiods. The first, which took place between 1995 and 1999, included 56 new kilometers, 38 new stations, and 223 new rolling stock (in Figure 3, represented by blue dots). The second period was from 1999 to 2003, when 54.6 new kilometers and 36 new stations (419 rolling stock) were built. The objectives of these two successive plans were as follows: the extension of the Madrid subway to densely populated suburbs (extensions of lines 1, 4, 7, 9, and the new line 11); the improvement of the structure of the subway and connections (central portions of lines 7 and 10); the improvement of accessibility to strategic areas (line 8 to the airport and trade fair venue); and Metrosur.

The timeline of the construction between 1999 and 2003 was:

- 1994 Line 1: Puente de Vallecas – Miguel Hernández, five new stations.
- 1995 Line 6: closes the circle, extension from Laguna to Ciudad Universitaria, six new stations, May.
- 1996 Line 10: Plaza de España – Príncipe Pío, one station.
- 1998 Line 8: Mar de Cristal – Campo de las Naciones, June. Line 9: Pavones – Vicálvaro, December. Line 10: Alonso Martínez – Nuevos Ministerios (two stations); it continues as far as Fuencarral, taking advantage of the previous L8 (which changed its name to L10). Line 11: Plaza Elíptica – Pan Bendito, new line, three stations, November. Line 4: Canillas – Mar de Cristal (one station), April. Line 4: Mar de Cristal – Parque de Santa María (2 stations), December. Line 7: Avenida de América – Canal, three new stations, March and October.
- 1999 Line 8: Campo de las Naciones – Aeropuerto T1, T2 and T3, June. Line 8: Aeropuerto – Barajas, September. Line 9: Vicálvaro – Arganda del Rey, September, 18 km. Line 1: Miguel Hernández – Congosto (Vallecas), three new stations, 2.7 km, March. Line 5: New station, Eugenia de Montijo, built between Aluche and Carabanchel. Line 7: Extension from Canal to Pitis, ten stations.
- 2002 Line 8: Nuevos Ministerios – Colombia- Mar de Cristal. Line 10: extended from Casa de Campo to Colonia Jardin.

Second Plan 2003–2007

The metro added 59 kilometers of new lines, 28 kilometers of light rail, and 81 new stations (see red dots in Figure 3).

The timeline of the construction between 1999 and 2003 was:

- 2003 Line 12: Circle line around municipalities south of Madrid, April. Line 10: extension from Colonia Jardín - Cuatro Vientos – Puerta del Sur (connected to L12). Line 5: extension from Aluche to Casa de Campo (three stations), October.
- 2006 Line 10, new station Aviación española, built between Colonia Jardín and Cuatro Vientos. Line 11: extension from Pan Bendito to La Peseta, three new stations (Carabanchel), 2.7 km. Line 5: extension from Canillejas to Alameda de Osuna, two stations, November, 2.4 km.
- 2007 Line 8: Nuevos Ministerios – Aeropuerto T4; last extension in 2007. Line 8: new station Pinar del Rey, built between two existing stations. Line 8: extended from Barajas to Airport T4, one station, 2.4 km. Line 10: From Fuencarral to Hospital Infanta Sofia, 15.8 km, incorporating eleven new stations. Line 3: Extension from Legazpi to Villaverde Alto, seven new stations, 8.4 km. Line 2: Extension from Ventas to Elipa (one station), 1.5 km. Line 1: Extension from Congosto to Valdecarros (three stations), South, 5.3 km. Line 1: Extension from Plaza Castilla to Pinar de Chamartín (three stations), North. Line 4: Extension from Parque de Santa María to Pinar de Chamartin (three stations). Line 6: New station, Arganzuela-Planetario, built between Mendez-Alvaro and Legazpi. Line 7: Extension from Las Musas to Henares, seven new stations. Line 1: Light Rail: nine stations, 5.4 km, Pinar de Chamartín-Las Tablas. Line 2: Light Rail: thirteen stations, 8.4 km, Colonia Jardín – Estación de Aravaca. Line 3: Light Rail, 16 stations, 13.5 km, Colonia Jardín – Puerta de Boadilla.
- 2008 Line 9, new station Rivas Futura, built between Rivas – urbanizaciones and Rivas Vaciamadrid. Line 7: extension from Henares to Hospital de Henares, one station, February.
- 2010 Line 11: La Peseta – La Fortuna (Leganés), one station (last extensión), October.
- 2011 Line 9: extension from Herrera Oria to Mirasierra, one station. Line 2: extension from Elipa to Las Rosas, four stations.
- 2015 Line 9: extension from Mirasierra to Paco de Lucia (one station).
- 2019 Line 7: extension from Lacoma to Pitis (one station).

Appendix D. Region of Madrid: Operating Speed and Frequencies by Metro Line (1998–2020)

Table D.15: Speed

Line	1998	1999	2000	2010	2020
1	19.05	21.3	20.1	21.4	22.8
2	18.28	19.3	18.2	18.5	23.2
3	19.79	19.3	18.2	24.7	23.9
4	21.3	21.3	18.93	21.8	21.7
5	21.96	18.8	18.98	23.5	22.6
6.1	28.04	24.7	26.22	24.6	24.3
6.2	27.4	27.3	27.85	24.9	25.1
7	26.42	26.8	25.6	28.6	27.0
7b				35.7	32.7
8	21.8	37.0	37.06	42.1	35.5
9	28.68	26.8	26.9	26.7	25.4
10	27.79	27.4	26.35	30.9	29.9
10b					30.0
11	17.35	18.2	17.4	23.6	28.4
12.1				41.6	40.6
12.2				41.0	40.6
Ramal	12.91	13.3	11.63	11.8	11.6
TFM				57.4	
Metro Norte				40.3	

Notes: Average operating speed in km/h by line and year. Blank cells indicate the line was not yet in operation or data was unavailable for that year.

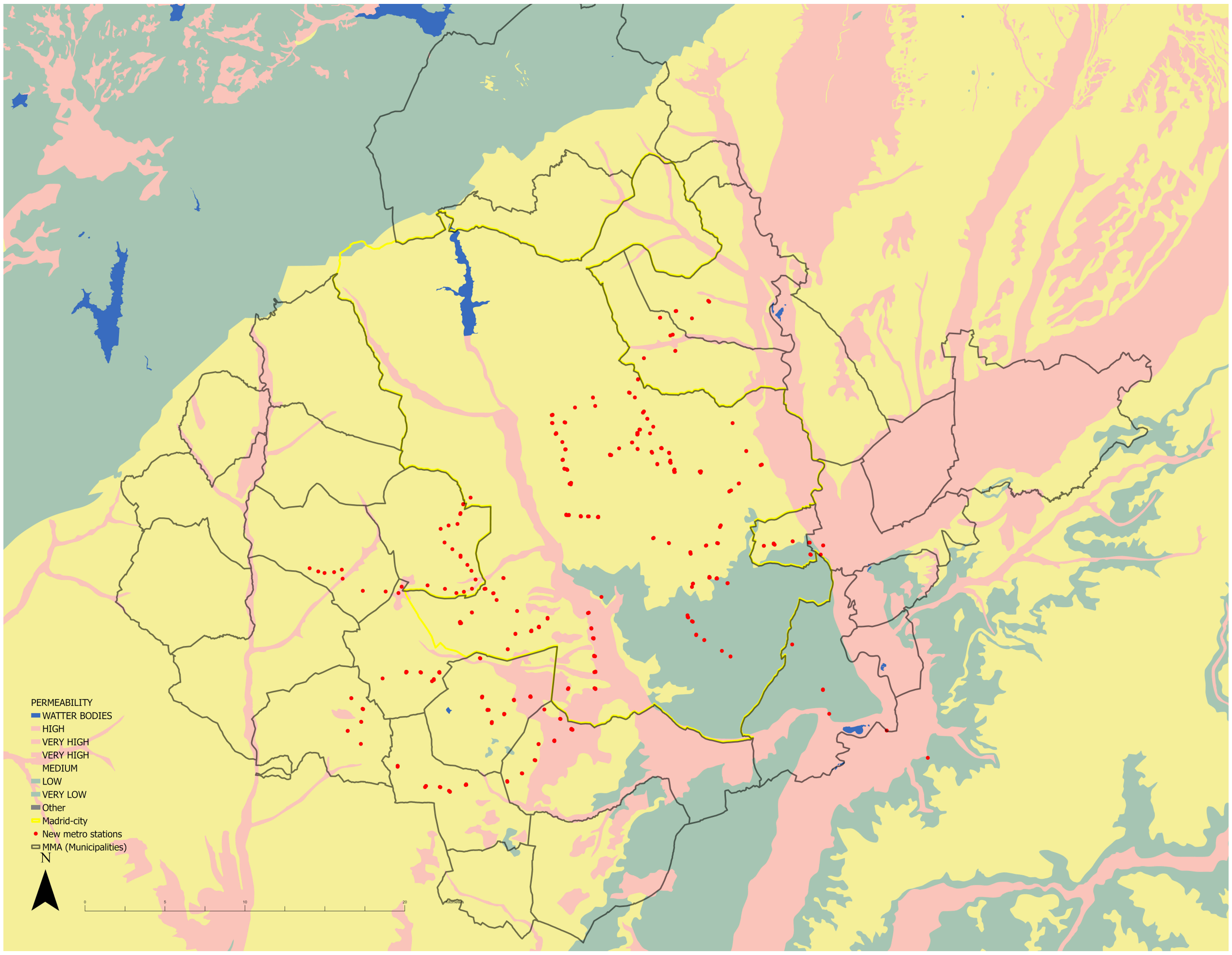
Table D.16: Frequencies

Line	1998	1999	2000	2010	2020
1	2.7	2.4	2.4	2.9	3.2
2	3.8	3.7	4.2	3.5	3.6
3	2.3	2.3	2.2	2.5	2.5
4	3.2	3.2	3.1	3.0	3.0
5	3.7	3.7	3.4	3.4	3.7
6.1	3.0	2.9	2.5	3.1	2.9
6.2	3.8	3.8	3.3	4.0	4.1
7	4.1	4.6	4.5	4.0	4.3
7b				4.9	5.8
8	5.8	5.1	4.3	3.5	3.9
9	4.0	3.7	3.1	3.6	4.3
10	2.9	2.8	2.5	2.9	2.8
10b				3.9	4.0
11	5.9	5.8	5.3	5.6	5.6
12.1				5.3	5.9
12.2				5.3	5.9
Ramal	4.5	4.4	4.4	5.0	5.0

Notes: Average operating frequencies in minutes by line and year. Blank cells indicate the line was not yet in operation or data was unavailable for that year.

Appendix E. Permeability

Figure E.8: Madrid region, permeability levels



Notes: Own development from continuous vector data from IGME.

Table E.17: Lithology of the subsoil in Spain classified by permeability levels

Permeability level	**Lithology**	**Types of rocks in the subsoil**
Highly Permeable	Q-HP	DETRITICAL (QUATERNARY)-VERY HIGH: Gravels, sands, silts (alluvial deposits, valley bottoms and low terraces in main rivers)
Permeable	Q-P	DETRITICAL (QUATERNARY)-HIGH: Blocks, stones, gravels, sands, silts and clays (glacis deposits, foothills and surfaces, hillside deposits, colluvium, moraines, glacial deposits, screes, mountain colluvium)
Permeable	Q-P	DETRITICAL (QUATERNARY)-HIGH: Gravels, sands, silts and clays (Deposits of medium and high terraces)
Permeable	Q-M	DETRITICAL (QUATERNARY)-MEDIA: Blocks, stones, gravels, sands, silts and clays (glacis deposits, foothills and surfaces, hillside deposits, colluvium, moraines, glacial deposits, screes, mountain colluvium)
Medium	C-M	CARBONATED-MEDIUM: Gastropod limestones, marls, margolimestone, and sometimes conglomerates and gypsum
Medium	C-M	CARBONATED-MEDIUM: Reddish ocher limestones, limestones, dolomites, marls (Dol. Villa de Ves, Caballar, F. Tabladillo, Hontoria; Mgas. Chera) and white marls
Medium	C-M	CARBONATED-MEDIUM: Limestone, marly dolomite limestone, marl, white marl, sometimes with flint nodules and lignite
Medium	C-M	CARBONATED-MEDIUM: Limestone, marl, margolimestone, dolomite, calcareous and oncolithic breccias ("Tierra de Campos" terminal limestones, east of Burgos, "intra-Cuestas" limestones, Arévalo limestones, Páramo 1 and 2 limestones)
Medium	C-M	CARBONATED-MEDIUM: Marbles, limestone and calcium silicate rocks
Medium	D-M	DETRITICAL-MEDIA: Clays, silts and sands, with intercalations of sandstones and limestones
Medium	D-M	DETRITICAL-MEDIA: Arcoses and thick arkoses (sometimes with pebbles, conglomerates, clays and arkosic muds)
Medium	D-M	DETRITICAL-MEDIA: Sand and gravel. Variolated sludge
Medium	D-M	DETRITICAL-MEDIA: Sands, clays, gravels and conglomerates (F.Utrillas), sometimes with dolomites (F. Ar. and Arc. Segovia)
Medium	D-M	DETRITICAL-MEDIA: Arenis.quarc., arenis.y microcongl.silíc., with rubefacc. (Siderolitic Series, Ar. de Salamanca, Ud. Torneros) silic gravels, conglom., pink clays, limestones, siltstones, red clays (Sigüenza Zone), clays (Ar. de Cabrerizos and Villamayor) yav
Medium	D-M	DETRITICAL-MEDIA: Sandstones and conglomerates (with clays and silts), limestones, shales and sometimes marls
Medium	D-M	DETRITICAL-MEDIA: Conglomerates, sandstones and shales
Medium	D-M	DETRITICAL-MEDIA: Conglomerates, quartzite conglomerates, gravels, sands, siliceous sands, silts and clays (Rañas and other finineogenic alluvial and fluvio-lacustrine facies)
Medium	D-M	DETRITICAL-MEDIA: Marls, margolimestone, limestone, bioclastic and/or sandy limestone, calcarenites, clays, dolomites, sandstones, siliceous sands, shales
Low	D-L	DETRITICAL-LOW: Shales, sandstones, white marls, gypsum marls, gypsum, with clays, sands and flint. Red clays and sandstones. Conglomerates and limestones, subordinate
Low	D-L	DETRITICAL-LOW: Marls, clays, gypsum, conglomerates, sandstones, shales, red shales and ocher sands
Low	D-L	DETRITICAL-LOW: Flint and sepiolite levels
Low	D-L	DETRITICAL-LOW: Gypsum (sometimes with gypsum clays, loams, clays and sandstones), marls, clays, limestones and shales
Low	E-L	EVAPORITE-LOW: Shales, sandstones, white marls, gypsum marls, gypsum, with clays, sands and flint. Red clays and sandstones. Conglomerates and limestones, subordinate
Low	E-L	EVAPORITE-LOW: Gypsum (sometimes with gypsum clays, loams, clays and sandstones), marls, clays, limestones and shales

Permeability level	Lithology	Types of rocks in the subsoil
Low	I-L	IGNEOUS-LOW: Hercynian plutonic rocks (granites, granodiorites, quartzdiorites, gabbros, diorites, tonalites, ultramafic rocks)
Low	M-L	META-DETRITICS-LOW: Black schists, slates, black quartzites, migmatites, amphibolites, paragneisses, acidic volcanic rocks, shales, sandstones, conglomerates, quartzites, micaschists, phyllites and meta-sandstones, arkosic sandstones, schists, conglomerate, volc.meta rocks
Low	M-L	META-DETRITICS-LOW: Ampelitic slates, sandstones, quartzites, shales, limestones, conglomerates, vulcanites, metavulcanites
Low	M-L	META-DETRITICS-LOW: Metamorphized rocks and peridotites (metabasites, amphibolites, serpentinites, orthogneisses, migmatites, glandular gneisses, metarriolites, peralkaline gneisses)
Low	Q-L	DETRITICAL (QUATERNARY)-LOW: Clays, silts and stones, peat (Deposits from endorheic areas, lacustrine deposits, peat bogs)
Very Low	I-VL	IGNEOUS-VERY LOW: Hercynian vein rocks (porphyries, aplites) or basic/intermediate rocks (diabases/ lmprofhides), quartz and gabbro-diabase veins from the Plasencia dam
Very Low	I-VL	IGNEOUS-VERY LOW: Hercynian plutonic rocks (granites, granodiorites, quartzdiorites, gabbros, diorites, tonalites, ultramafic rocks)
Very Low	M-VL	META-DETRITICS-VERY LOW: Ampelitic slates, sandstones, quartzites, shales, limestones, conglomerates, vulcanites, metavulcanites

Appendix F. OLS results

Table F.18: Madrid, 2020-1998 Δ total population. OLS results. All coefficients

		[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Δ accessibility		**.00007**	**.00006**	**.00073**	**.00073**	**.00088**					
	a	*(.00011)*	*(.00013)*	*(.00033)*	*(.00033)*	*(.00034)*					
	b	*.570*	*.631*	*.026*	*.027*	*.010*					
Δ ED							**.00204**	**.00054**	**.01187**	**.01181**	**.01578**
	a						(.00122)	(.00131)	(.00355)	(.00359)	(.00379)
	b						*.095*	*.684*	*.001*	*.001*	*.000*
Initial population		**-.58191**	**-.49141**	**-.65828**	**-.6755**	**-.95095**	**-.58124**	**-.4915**	**-.66026**	**-.67734**	**-.94954**
	a	*(.03086)*	*(.44167)*	*(.04304)*	*(.04237)*	*(.01915)*	*(.03089)*	*(.04429)*	*(.04299)*	*(.04232)*	*(.01911)*
	b	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*
Distance to CBD		**.00681**	**-.00543**	**.019016**	**.019438**	**.01974**	**.00745**	**-.00522**	**.01977**	**.020164**	**.021087**
	a	*(.00150)*	*(.00317)*	*(.00301)*	*(.00302)*	*(.00308)*	*(.00151)*	*(.00316)*	*(.00290)*	*(.00291)*	*(.00295)*
	b	*.000*	*.086*	*.000*	*.000*	*.000*	*.000*	*.098*	*.000*	*.000*	*.000*
Distance to ramp		**.00953**	**-.01664**	**.01901**	**.02034**	**.01991**	**.00975**	**-.01653**	**.02028**	**.020165**	**.0210871**
	a	*(.00480)*	*(.01024)*	*(.00542)*	*(.00538)*	*(.00456)*	*(.00481)*	*(.01015)*	*(.00542)*	*(.00291)*	*(.00295)*
	b	*.047*	*.104*	*.000*	*.000*	*.000*	*.043*	*.104*	*.000*	*.000*	*.000*
Area census tract		**93.556**	**29.091**	**113.35**	**109.705**	**105.986**	**93.712**	**28.572**	**112.453**	**108.831**	**105.205**
	a	*(19.011)*	*(22.909)*	*(20.102)*	*(19.459)*	*(17.733)*	*(18.979)*	*(22.909)*	*(19.990)*	*(19.356)*	*(17.667)*
	b	*.000*	*.204*	*.000*	*.000*	*.000*	*.000*	*.212*	*.000*	*.000*	*.000*
MMA (excluding the Mun.of Madrid)		x		x	x	x	x		x	x	x
Municipality of Madrid		x	x				x	x			
Only Residential Land		Y	Y	Y	N	Y	Y	Y	Y	N	Y
Pop.>0 in 1998		Y	Y	Y	Y	N	Y	Y	Y	Y	N
R2*		0.38	0.300	0.493	0.483	0.721	0.38	0.300	0.490	0.486	0.723
No. of observations		*3,246*	*2,176*	*1,070*	*1,093*	*1,333*	*3,246*	*2,176*	*1,070*	*1,093*	*1,333*

Notes: Distance to ramp= 1990–2017 variation of the distance to the closest highway access ramp. Permeability= closest distance in meters to *Low* and *Very Low* soil. [a] = Standard deviation [b]= *P*-value.

Appendix G. TSLS results

Table G.19: Madrid, 2020-1998 Δ total population. TSLS results with geological IV.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
Δ accessibility	**-.00187**	**.00017**	**.00425**	**.00428**	**.00541**					
a	*(.00105)*	*(.00055)*	*(.00220)*	*(.00219)*	*(.00191)*					
b	*.074*	*.761*	*.053*	*.051*	*.005*					
Δ ED						**-.01962**	**.00176**	**.04614**	**.04644**	**.05606**
a						*(.01100)*	*(.00577)*	*(.02354)*	*(.02343)*	*(.01944)*
b						*.075*	*.761*	*.050*	*.047*	*.004*
IV: Permeability	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IV: 1760 Roads	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Initial population	**-.5921**	**-.4901**	**-.6636**	**-.6872**	**-.9470**	**-.5916**	**-.4898**	**-.6660**	**-.6900**	**-.9470**
a	*(.03359)*	*(.04455)*	*(.04427)*	*(.04358)*	*(.01994)*	*(.03337)*	*(.04472)*	*(.04435)*	*(.04363)*	*(.01994)*
b	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*	*.000*
Distance to CBD	**-.0038**	**-.0048**	**.0338**	**.0351**	**.0342**	**-.0030**	**-.0042**	**.0305**	**.0318**	**.0342**
a	*(.00576)*	*(.00399)*	*(.00997)*	*(.00986)*	*(.00688)*	*(.00536)*	*(.00424)*	*(.00821)*	*(.00813)*	*(.00688)*
b	*.512*	*.234*	*.001*	*.000*	*.000*	*.571*	*.326*	*.000*	*.000*	*.000*
Distance to ramp	**.01288**	**.01718**	**.01836**	**.03217**	**.02370**	**.00511**	**.01574**	**.01633**	**.03002**	**.02370**
a	*(.01582)*	*(.00860)*	*(.01913)*	*(.01956)*	*(.01673)*	*(.01407)*	*(.00843)*	*(.01899)*	*(.01743)*	*(.01673)*
b	*.415*	*.046*	*.337*	*.100*	*.157*	*.717*	*.062*	*.390*	*.085*	*.157*
Area census tract	**80.91**	**33.23**	**127.25**	**113.37**	**112.21**	**89.84**	**32.16**	**122.55**	**108.33**	**112.21**
a	*(23.83)*	*(28.23)*	*(27.57)*	*(26.68)*	*(23.42)*	*(22.21)*	*(26.87)*	*(26.16)*	*(25.25)*	*(23.42)*
b	*.001*	*.239*	*.000*	*.000*	*.000*	*.000*	*.231*	*.000*	*.000*	*.000*
Metro Area	x		x	x	x	x		x	x	x
Muni. of Madrid	x	x				x	x			
Only Residential Land	Y	Y	Y	N	Y	Y	Y	Y	N	N
Population>0 in 1998	Y	Y	Y	Y	N	Y	Y	Y	Y	Y
No. of observations	*3,246*	*2,176*	*1,070*	*1,093*	*1,333*	*3,246*	*2,176*	*1,070*	*1,093*	*1,333*

Notes: Distance to ramp = 1990–2017 variation of the distance to the closest highway access ramp. Permeability = closest distance in meters to *Low* and *Very Low*. Metro area = metropolitan area of Madrid. First-stage *F*-statistics are reported for each endogenous variable in Table 7. [a] = Standard deviation [b] = *P*-value.