Departament d'Història Econòmica, Institucions, Política i Economia Mundial
Av. Diagonal, 690
08034 Barcelona

DOCTORAL STUDENT SEMINAR 2020-21

PHD IN ECONOMIC HISTORY

DEPARTAMENT D'HISTÒRIA ECONÒMICA, INSTITUCIONS, POLÍTICA I ECONOMIA MUNDIAL

UNIVERSITAT DE BARCELONA

Guillermo Esteban Oliver (Universitat de Lleida)

The curse of geography? Railways and growth in Spain 1877-1930

Very preliminary version. Please do not quote

Date: Thursday, February 18, 2021
Hour: 14:00 hours
Place: Webinar. Link: https://eu.bbcollab.com/guest/da338575022d4b898a4a7cbc2567dc19

***The curse of geography?* Railways and growth in Spain 1877-1930**

Guillermo Esteban-Oliver

Department of Geography and Sociology, University of Lleida, Lleida, Spain

guillermo.esteban@udl.cat

Abstract

In this study, we explore the relationship between rail accessibility and municipal population growth in Spain from 1877 to 1930. To carry out this analysis we introduce a novel database, which combines census data with the geo-location of access points (stations and stops). Then, and in order to establish causality, we use a Least Cost Path (LCP) instrument. Our results suggest that municipalities with *direct access* (less than 1-hour walking distance to the nearest station or stop) experienced more rapid growth. The findings are robust to several checks and point to the transformative power of transport infrastructure, especially developing economies with an unforgiving geography.

Resumen

En este trabajo exploramos la relación entre el acceso al ferrocarril y el crecimiento de la población municipal en la España de entre 1877 a 1930. Para realizar este análisis hemos utilizado una nueva base de datos que combina la información censal con la geolocalización de los puntos de acceso a la red (estaciones y apeaderos). En el apartado metodológico, y para establecer causalidad, utilizamos el instrumento "Least Cost Path" (LCP). Nuestros resultados indican que los municipios con acceso directo a la red (a menos de una hora de distancia de camino hasta la estación o apeadero más cercano) experimentaron, de media, un crecimiento significativamente superior. Los resultados son robustos a diversos controles y revelan la capacidad de transformación que las infraestructuras de transporte pueden tener, especialmente en economías en desarrollo y de complicada orografía.

JEL: N7, N9, R00

Keywords: Railways, population, growth

1. Introduction

Advances in transport technology are at the core of socioeconomic change. Prior to the development of steam-powered rail transport, waterways and ocean shipping outcompeted any form of overland transport. Consequently, railways brought about a profound transformation, not just by reducing transport costs but also by stimulating political integration. National and global market integration and the articulation of State structures thus owed a great deal to rail transport which, in a way, fuelled the transition into modern economies and societies during the 19th century.

Still, in order to assess how a novel transport technology (railways, driverless cars…) transforms our daily lives it is critical to understand the mechanisms through which economy and society are affected. This is far from trivial since there are different modes of transportation and implications. In the case of railways, there is an extensive literature that first described its origins and expansion, and then analysed its socioeconomic impact[1]. For instance, the effect of railways on economic growth has been quantified following a *social savings* approach (Fogel, 1964; Herranz-Loncán, 2007).

More recently, and as a result of the development of Geographic Information Systems (GIS), it has been possible to delve deeper into the subject (Atack, 2013; Martí-Henneberg, 2013). Using counties as units of analysis, Atack *et al* (2008; 2010) looked at how rail accessibility affected population density, urbanisation, and industrial development in 19th century United States. Similar efforts followed in other contexts (Gregory and Martí-Henneberg, 2010; Schwartz *et al,* 2011; Koopmans *et al,* 2012; Barquín *et al,* 2012; Hornung, 2015; Jedwab and Moradi, 2016).

Yet, several compelling issues have been raised. Donaldson and Hornbeck (2016) pointed that, in general, studies estimate relative impacts and proposed a *"market access"* methodology, which captures *"lowest-cost county-to-county freight routes"*[2]. Though appealing, this approach is a challenge in historical settings, as it demands a complete spatial data infrastructure. Similarly, there are different ways to measure rail accessibility which might affect the analysis (Mimeur *et al,* 2018). Likewise, and as regards growth, the literature has predominantly focused on urban growth (Barquín *et al.,* 2012; Hornung, 2015; Jedwab and Moradi, 2016; Berger and Enflo, 2017) thereby overlooking the transformative

[1] For a seminal contribution on railroads and economic growth, see Fogel (1964) and Fishlow (1965). Recent studies have explored other dimensions (Melander, 2020; Medina-Albaladejo *et al*, 2020).

[2] Using changes in the railroad and waterways from 1870 to 1890, Donaldson and Hornbeck (2016) then quantified the aggregate impact of railroads on agriculture in the United States. See also Donaldson (2018); Hornbeck and Rotemberg (2019).

power of railways on rural areas. However, recent Works such as Büchel and Kyburz (2020) or Braun and Franke (2019) (for Switzerland and Wurttemberg respectively) analyse the effects of railways in mostly rural setups.

Our paper also relates with the strand of the literature that, using a New Economic Geography approach, focuses on the spatial distribution of population in historical perspective[3]. These studies search for the factors that can explain population growth and its spatial distribution in the long term, being access to the markets (transport infrastructure development) one of the most important (Ayuda *et al*, 2010; González-Val *et al,* 2017).

This study aims at providing further evidence by analysing the case of Spain at the municipal-level, which is the smallest administrative unit. For this purpose, a novel dataset with 7,641 (urban and rural) municipalities is presented. It includes data on population and rail accessibility, measured as the distance between each municipal centre and its nearest station, among other characteristics, for each of the following census years: 1860, 1877, 1900 and 1930. In doing so, the location of 2,957 access points (stations and stops) in operation between 1848 and 1941 has been digitised. This spatial data infrastructure thereby allows for an in-depth analysis of the relationship between the expansion of railways and growth in a developing economy.

Then, we assess whether accessibility induced population growth in two periods (1877-1900; 1900-1930) and in the long-run, 1877-1930. However, municipal dynamism in the pre-railroads era might have conditioned the network layout. To address this endogeneity issue, we follow Banerjee *et al.,* (2020), and introduced an Instrumental Variable (IV) strategy based on optimal routes or Least Cost Paths (LCPs) between *nodes,* that is, cities and towns that were of special economic or political relevance in the past.

In line with similar studies (Braun and Franke, 2019; Büchel and Kyburz, 2020), our results confirm that railways positively affected municipal population growth. More specifically, municipalities with *direct access* (less than 1-hour walking distance from a station or stop) grew more rapidly. This effect is significant and robust, especially in the early 20th century and the long-run[4]. Also, the magnitude of

[3] Krugman (1991, 1993) was the catalyst for this types of studies, which have had great influence since the edge of the XX century.
[4] Whereas the monumental work of Artola et *al.,* (1978) described in great detail the expansion of the railroads in Spain; Herranz-Loncán (2007), following Gómez-Mendoza (1982), adopted a *social savings* approach in order to quantify its impact on economic growth. Both authors found a modest contribution before 1913.

the coefficients resembles that of previous studies, which is somehow compelling, since Spain was a developing economy with moderate growth and sparsely populated.

That said, this study offers further insight. In Spain, the construction of railroads occurred in parallel with the process of State-building and, at least in the early stages, preceded industrialisation (Nadal, 1975)[5]. Besides, peninsular Spain suffers from a peculiar physical geography. A mountainous country with few waterways and central plateaus which are separated from coastal areas by mountain ranges[6].In this context, the case of Spain differs from previous studies since: steam-powered rail transport was clearly the first-best alternative for internal transport; the challenging physical geography severely restricted the construction of railroads; and this transport infrastructure was *built ahead of demand* in a country with low-population density[7].

The following section describes the historical context. Section 3 presents the data, while sections 4 and 5 present the methodology and results. Then, some concluding remarks are added in section 5.

2. Historical context

During the 19th century, the Kingdom of Spain was essentially an agrarian economy, especially when compared to other Western European countries[8]. Still, some territories, such as Catalonia or the Basque provinces, witnessed profound socioeconomic change, especially in the late decades (Díez-Minguela *et al.,* 2018). It was precisely in these regions where the earliest successful initiatives to construct railroads were launched. Pioneered in England during the 1820s, the first steam-powered railway in peninsular Spain was developed between Barcelona and Mataró in 1848. By then, there were already 3,000 km of tracks in France and almost 10,000 km in the United Kingdom.

[5] In the literature, one of the main research questions has been to examine whether railroads were "*built ahead of demand*" (Fishlow, 1965).
[6] Regarding human geography, Spain presents another anomaly, an extremely low density of settlements, at least when compared to other European countries (Gutierrez et *al.,* 2020). See also Oto-Peralías (2020). Furthermore, the fact that most of the country lies within the peninsula may have condition a coastal nation status, in which coastal shipping was clearly the pre-railway best alternative.
[7] By 1848, there were only two canals operating in Spain: "*Canal Imperial de Aragón*" (88 km.) and "*Canal de Castilla*" (210 km.). Between 1871 and 1875, with the railroads still far from being complete, cabotage trade represented 30% of the merchandise carried by Norte and MZA companies, whose joint lines accounted for around 50% of the network (See Figure A.1 in the appendix) (Tafunell and Carreras 2005, pp. 515). Although we don't have a detailed way to account for the influence of coastal shipping, our data indicates that even though it was important, it was heavily overshadowed by rail transport.
[8] In 1870, the Spanish industrial per capita product was 215, whereas the German, French and English reached 536, 654 and 1305 respectively (Tafunell and Carreras, 2005: Cuadro 3.17).

Notwithstanding this relative backwardness, the construction of a railway network became a national priority, as pointed out in some of the parliamentary sessions that took place throughout the 1830s and 1840s. In 1844, for instance, Bravo Murillo argued[9]:

> *"The roads of iron are called to change the conditions of states, to spread ideas with an extraordinary and hitherto unknown speed, to facilitate all the relationships that make up the life of peoples, and to be powerful agents of administration and government, with whose help it will be easy for the executive power to make its tutelary action or its severe hand immediately felt at all the angles of the monarchy at the same time, establishing among the various parts of the State the cohesion and unity which constitute the centralization and the real and effective force"* (Artola *et al.* 1978, Volume 1, Chapter 1).

As the above text illustrates, steam-powered rail transport was perceived as a central instrument for the articulation of the country[10]. Nevertheless, some authors have pointed to the State for the slow progress in this matter (Tortella, 1974). However, the "Carlists Wars" and political instability hindered the enaction of a General Railway Law[11]. Yet, the construction of railways faced other major issues. First, the lack of a general consensus precluded the creation of a well-defined plan for the trunk network. There were also problems in acquiring the appropriate technical and financial means of construction; and troubles in adopting appropriate forms of network operation, i.e. those that guaranteed the best public interest at the lowest cost.

By mid-19th century, a draft for a General Railway Law was drawn up, in which the concession and regulation of lines between the State and companies was clearly defined. Approved in 1855, this draft became the *"Ley General de Ferrocarriles"* or General Railways Law (GRL1855), which stated that *"among the general service lines, those that depart from Madrid and end up at the coasts or borders of the kingdom will be classified as trunk lines"*. That is to say, and as Map 1 shows, the State roughly established the trunk lines. Still, the construction of the infrastructure was the responsibility of private companies, though

[9] Juan Bravo Murillo (1803-73) was an important political figure during the reign of Isabel II.

[10] The development of a modern territorial administration with the creation of *"Provincias"* (1833) and *"Partidos judiciales"* (1834) were also of great relevance. See Pro (2019) for a recent overview of the State-building process.

[11] For instance, a parliament lockdown did not allow the Railway Law of 1848 to be passed. The "Carlists wars" were a series of civil conflicts between 1833 and 1876. The Carlists controlled parts of the north-east which could hinder the development of the railway network.

in many cases these ventures received sizable subsidies[12]. Moreover, the concession of exclusive rights to operate these lines was granted for 99 years, in counterpart certain aspects were regulated, such as setting maximum price-rates or determining features of the morphology of the network (Ortúñez, 2016)[13]. In short, the GRL1855 was somewhat vague since it did not present a detailed plan for the network development.

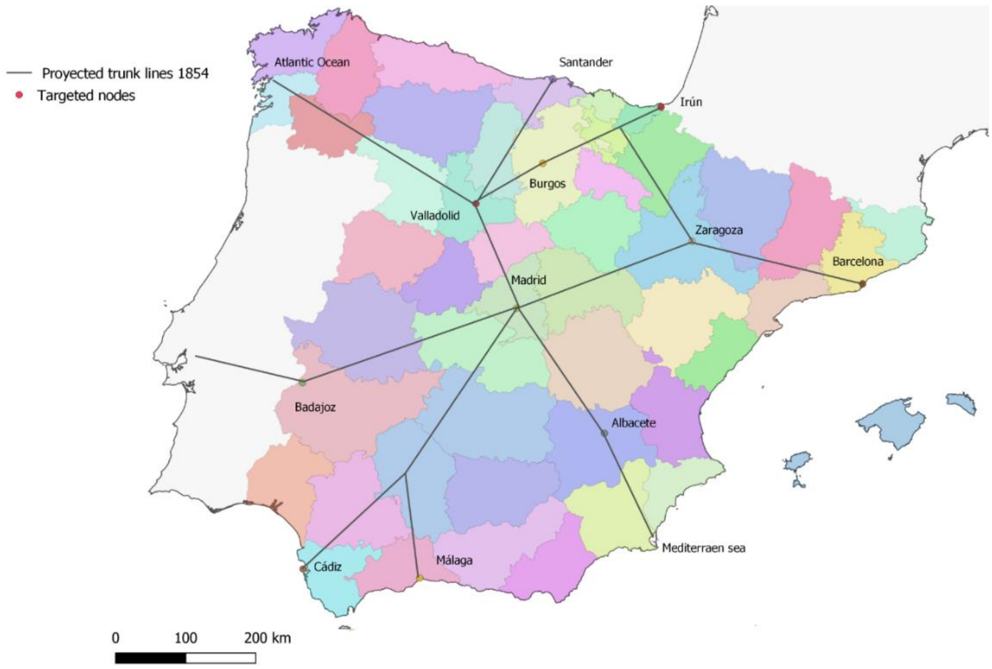
As a result, the proposals and plans that privately owned companies made conditioned the definitive development and morphology of the network (Barquín and Larrinaga, 2020). The GRL1855 created a framework that lasted until 1866; and its resulting outcome was a tree-shaped network, in which the major cities were linked but left aside transversal or regional connections. The *"Ley del Ferrocarril de 1877"* (or GRL1877) aimed at correcting this imbalance, giving priority to isolated territories and the construction of transversal lines (see Map 2). Around 1896 the broad-gauge network was practically concluded. Note that, there are two track gauges (broad and narrow) in Spain. Differences in their territorial distribution, morphology, size and usage force us to distinguish between them. That said, our focus lies on the broad-gauge network, since narrow-gauge was mainly used as auxiliary[14]. In this sense, and as Figure 1 shows, the broad-gauge network developed earlier and was greater in extension and use.

[12] Railway subsidies financed up to considerable amount of the construction costs. For example, subsidies covered around 70% of the construction costs of the Palencia-Galicia line (Artola *et al.*, 1978; pp. 356).

[13] The most important railway companies of the period were the "Compañía de los caminos del hierro del Norte de España" (Norte) and the "Compañía de los ferrocarriles de Madrid a Zaragoza y Alicante" (MZA). These two companies will absorb minor ones until controlling around 70% of the total broad-gauge network. This quota will remain stable until its nationalization in 1941.

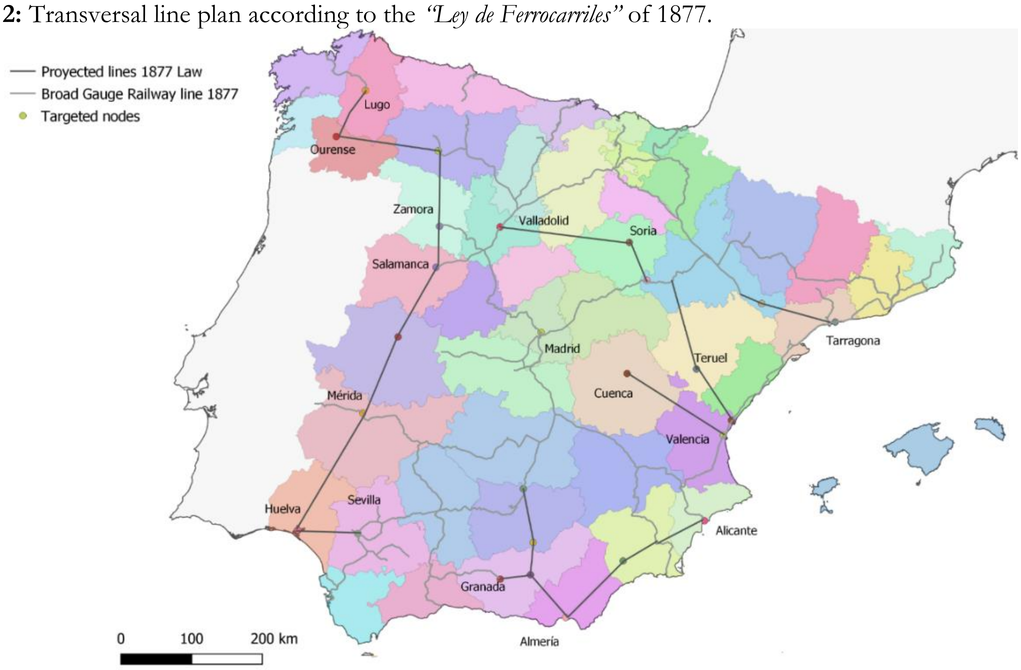
[14] Narrow gauge usage was mostly related to the transport of industrial inputs, mining and (latter on) metropolitan connections.

Map 1: Sketches of trunk Lines in the *"Ley General de Ferrocarriles"* of 1855.



Source: Own elaboration based on the *"Ley General de Ferrocarriles"* of 1855.

Map 2: Transversal line plan according to the *"Ley de Ferrocarriles"* of 1877.

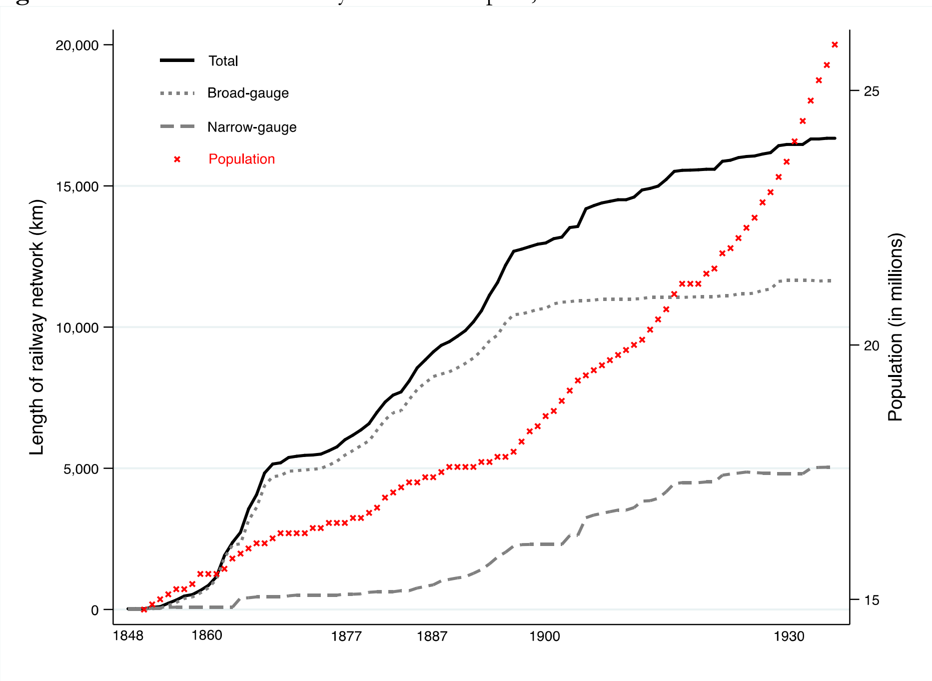


Source: Own elaboration based on the *"Ley General de Ferrocarriles"* of 1877.

The GRL1855 gave a huge boost to the network expansion. Between 1844 and 1855, 1,407 km of tracks were granted but only 477 were built, whereas between 1855 and 1865, 6,919 km were granted and 4,756 were built.

Likewise, the creation of a Statistical Commission in 1856 gave birth to national statistics and the first modern censuses (Muro *et al,* 1996). Figure 1 displays population and the extension of railways (broad-gauge and narrow-gauge) in Spain from 1848 to 1936[15]. The expansion of railroads thus occurred in parallel with the development of national statistics which were two of the pillars of the so-called liberal State (Pro, 2019).

Figure 1: Extension of the railway network in Spain, 1848-1941.



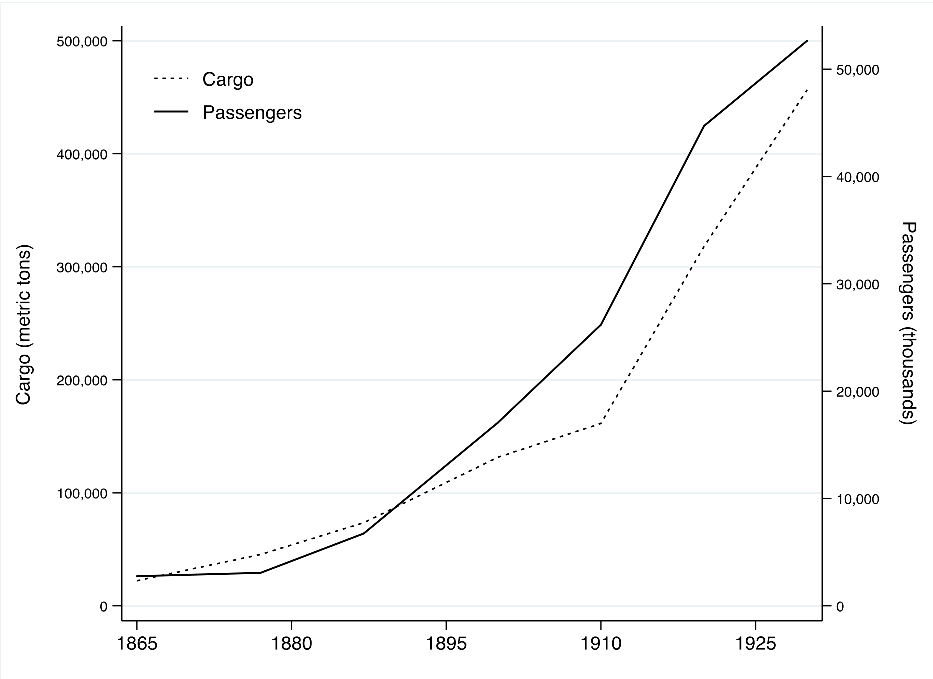
Source: Based on (Muñoz Rubio, 2005; García-Raya 2006). Population (Prados de la Escosura, 2017).

Regarding the construction of railroads, it is worth stressing that Iberia is a mountainous peninsula whose internal communications are *"not easy"* (Cabo and Vigil 1973: 35). In Spain, before the railroads, and with few waterways (navigable rivers or canals), people and goods essentially moved overland and through coastal shipping, a transport mode whose territorial reach was much more limited

[15] In Spain, there are two track gauges, broad and narrow. As figure 1 documents, broad-gauge preceded narrow-gauge railways. The latter one was aimed at mining, industry or regional connections (Muñoz Rubio, 2005). Created in 1941, RENFE (*"Red Nacional de los Ferrocarriles Españoles"*) is the State-owned enterprise that operates the entire broad-gauge network.

(Madrazo, 1984). Ringrose (1972) argued that this *"inadequate transportation system"* was fundamental in understanding the *"economic stagnation"* of the 19th century[16]. More recently, Grafe (2012) has added an institutional perspective to the conventional story of high transportation costs. Anyhow, given these particular geographic conditions, rail transport revolutionised the dynamics of internal transport.

Figure 2: Cargo (metric tons) and passengers (thousands) transported by Norte and MZA 1865-1930.



Source: Own elaboration based on Artola *et al.* (1978: Chapter 2, pp. 485-510).

In agriculture, the effect was truly remarkable. The reduction of transport costs stimulated the trade of both grains, departing from the central plateaus, and cash-crops (wine, citrus fruits, oil or raisins…) destined for export[17]. As a result, a slow yet gradual process of national economic integration began. Although Barquín (1997) found scant evidence of falling wheat price-gaps during the early expansion of railways (until 1875), later studies, such as Herranz-Loncán (2005), have pointed that rail transport provided a cost per unit shipped that was much lower (in terms Tm per Km) than that of the cheaper land transport alternative.

[16] In Spain, railroads *social savings* coefficients are quite large, especially compared to those of the United Kingdom or the United States (Herranz-Loncán 2005: Figure 1).
[17] Wine and cereals were the first and third most transported product by the MZA railways company during the first decades (1900-1930) of the 20th century (Artola *et al.*, 1978, pp. 446).

Railways also influenced mining and manufacturing. The extraction of mineral resources (coal, iron, copper…) critically benefitted from the presence of rail transport (Gómez-Mendoza, 1982; Cuellar 2003). Similarly, falling transport costs enabled the integration of manufacturing producers in the periphery with consumers in the interior (Pascual, 1984), thereby affecting the spatial distribution of economic activity (Tirado *et al,* 2002). In sum, the expansion of railroads was accompanied by a growing concentration of manufacturing, which marked the Spanish industrialisation.

Regarding passenger-traffic, railways allowed to drastically shorten travel times. It has been estimated that in 1867, and for trips from Madrid to a provincial capital, railway transportation shortened the journey by 60% (compared to the stagecoaches). Furthermore, at around the end of the 19th century these travel times would be 80% shorter than by traditional means of land transport (Cabanes Martín and González Sanz, 2009). As Figure 2 shows, the combined passenger traffic of the Norte and MZA networks increased from 3 million passengers in 1877 to 17 in 1900 and 56 in 1930, whereas freight traffic rose from 45 million tons to 130 and 450. Railway thus facilitated trade, factor and information flows in the early stages of modern economic growth, a period characterised by moderate growth and regional disparities (Rosés et *al.,* 2010; Martínez-Galarraga *et al.* 2015)[18].

3. Data

This study presents a novel database which combines *de facto* municipal population from the 1860, 1877, 1900 and 1930 censuses, with the GIS location of railway access points (stations and stops). Therefore, the spatial unit of analysis is the municipality, which is the lowest administrative unit in Spain since the liberal reforms of the 19th century. Still, and since the dependent variable is population growth and the number of municipalities in each census varies, it is important to have homogeneous units. Besides, in order to compute accessibility municipalities need to be georeferenced. Consequently, we convert Beltrán-Tapia *et al.,* (2019) municipal database to GIS information[19]. As

[18] While population increased from 15.6 to 23.5 million inhabitants, real Gross Domestic Product (GDP) rose from 35.8 to 211 million euros (Prados de la Escosura, 2017).

[19] Following (Goerlich Gisbert *et al.*, 2015) and (Franch *et al,*. 2013) who homogenized municipalities between 1900-2011 and 1877-1900, Beltrán-Tapia *et al.,* (2019) imposed the municipality structure (8,108 units) of the 2001 census for the period 1860-1930.

the focus lies within peninsular Spain, the autonomous cities in northern Africa (Ceuta, Melilla) and the archipelagos of the Canary and Balearic islands have been excluded.

Regarding frontiers and surface area, Beltran-Tapia *et al* (2019) dataset uses municipalities corresponding to the 2001 map of the National Geographic Institute (IGN), 2001-IGN; though in our study the information has been updated to the latest available, 2018-IGN. It is worth noting that the IGN defines municipal centres as the point where the main church was located. After adjusting the dataset, we know the coordinates for a total of 7,755 municipal centres in peninsular Spain (1860-1930).

Then, the access points in operation during the period of study, 1860-1930, were georeferenced. To do this we have used as reference the National Topographic Map at a scale of 1:50,000 (MTN50), in which numerous stations and stops can be found. Initiated in 1858, the publication of all these maps took nearly a century, 1875-1968[20]. Additionally, we have used complementary sources to reconstruct the location of some access points[21].

For the opening and closing dates of stations and stops, we have used the opening references of line-sections in the *Cronología básica del ferrocarril de vía ancha en España* (García Raya, 2006). These dates are essential to know the *time* dimension of the municipal access. Therefore, the opening date of each access point corresponds to the opening date of the line-section in which it was located. As regards narrow-gauge, we have followed an identical approach, though the source is *Cronología del ferrocarril de vía estrecha en España* (Muñoz Rubio, 2005). We just used a distinct criterion in provincial capitals and metropolitan areas, for which information on an access-by-access basis was compiled from the *Ferrocarril y Ciudad* project[22]. The dataset thus includes 1,899 stations and 1,058 stops in operation (broad and narrow-gauge) between 1848 and 1941. Based on this information, Table 1 and Map 3 display how accessibility (broad-gauge) evolved in Spain during the period of study.

[20] The MTN50 is the earliest and one of most important works in the history of Spanish cartography. To find out more about its origins and methodology see: https://www.ign.es/web/resources/docs/IGNCnig/CBG-Cartografia-IGN.pdf

[21] The complete dataset and a detailed explanation of the reconstruction process is part of my thesis (forthcoming), Esteban-Oliver (2021).

[22] Project title: *El ferrocarril y la ciudad en la encrucijada: paisaje urbano y patrimonio industrial en el entorno de las estaciones de la península ibérica, 1850-2017*. For more information see: https://www.fbbva.es/equipo/ferrocarril-la-ciudad-la-encrucijada-paisaje-urbano-patrimonio-industrial-entorno-las-estaciones-la-peninsula-iberica-1850-2017-estaciondigital/

Table 1: Expansion of the broad-gauge railway network in Spain 1860-1930

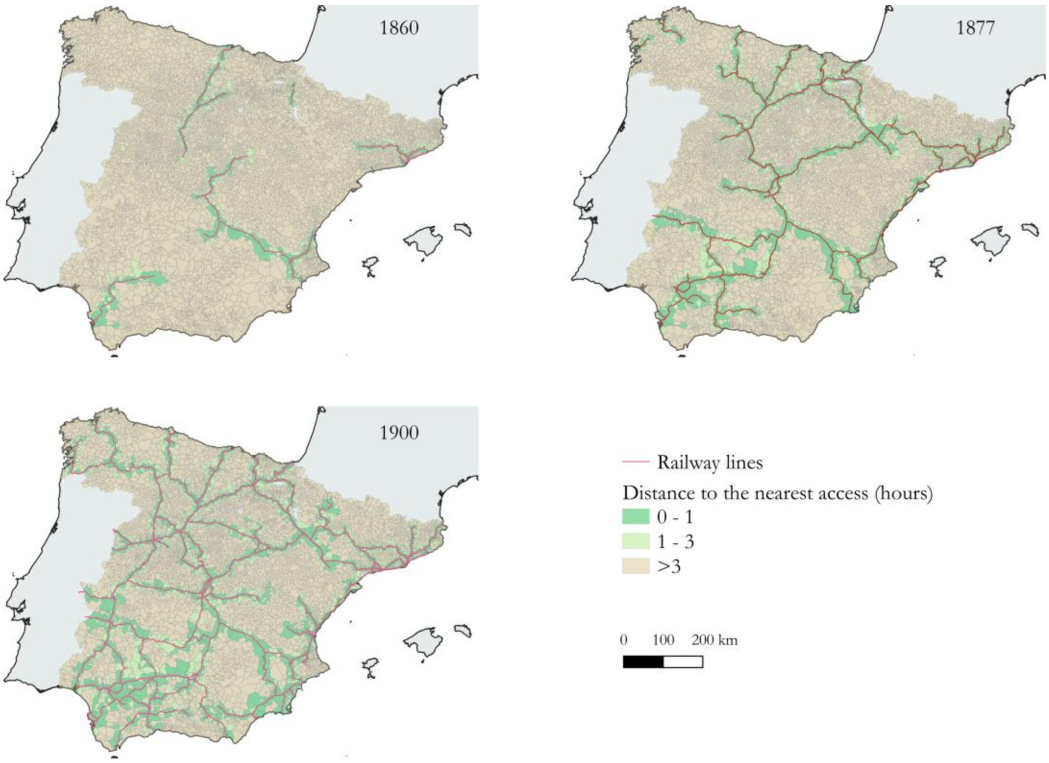
	1860	**1865**	**1877**	**1900**	**1930**
Stations and stops in the broad-gauge network	332	769	1,015	1,671	1,839
Broad Gauge tracks length (Km)	1,880	4,756	6,174	11,040	12,030
Population with *direct access* to the network	1,943,194	3,384,413*	4,747,210	7,640,922	11,497,939
Municipalities with *direct access* to the network	351	814	1,027	1,638	1,788
Percentage of the population with *direct access* to the network	12%	22%*	29%	41%	49%
Percentage of the municipalities with *direct access* to the network	4%	10%	12%	21%	23%

Notes: First row presents the number of stations and stops, while the second row shows the total length of tracks. The third and fourth rows presents the population and municipalities with *direct access*, that is to say, within 1-hour walking distance to a station or stop. The last two rows present percentages.
Sources: Population, 1860, 1877, 1900 and 1930 censuses. Municipal centres, IGN-2018. Points of access to the network, own elaboration. *1865 population data is calculated using the 1860 census.

Rail accessibility can be captured in different ways. In this study we use the stations and stops instead of distance to the track (or other approaches). The rationale is to avoid unnecessary noise, and capture accessibility in a homogeneous manner. As Mimeur *et al.,* (2018) pointed, the way access is measured matters. In this regard, it is worth stressing that both freight and passengers were carried by rail. In Spain, although freight-transport was the main income source for companies, thousands of passengers' boarded trains between 1877 and 1930 (see Figure A.2 in the appendix). Then, our dataset contains access points suited for passengers and cargo, but excludes stops at factories, mines or companies. Nevertheless, in most cases these points were located near municipal stations, and hence rail accessibility is not affected[23].

Map 3: Evolution of the municipal access to the broad-gauge network in peninsular Spain (1860-1900)

[23] Another potential concern is the frequency of use of these stations and stops which depended upon the schedules devised by companies. Sadly, we do not have such information.



Source: Own elaboration

Since we know the geolocation of municipalities and access points; using QGIS computation, we have estimated the Euclidean distance (in km) from the municipal centre to the nearest station or stop. Then, we have adjusted these estimated routes using the orography of the terrain[24]. The results are expressed in *hours of walking distance* from the municipal centre to the nearest access point. This is our main explanatory variable throughout the study. Note that as we already mention in the introduction, there are two track gauges in Spain, and we distinguish between them when calculating rail accessibility. Access to *broad-gauge* network will be the main explanatory variable of the study whereas the information on *narrow-gauge access* will be later used as a control variable.

The rest of the variables come from several sources. Population derives from the Spanish Population Censuses (Beltrán-Tapia *et al.,* 2019). Information on ports is extracted from *Estadísticas del comercio de cabotaje* (1857-1914) and *Estadísticas del comercio exterior de España* (1857-1914), which allow us to identify the ports that had customs and the specific dates on which they were operational[25]. With this, I have

[24] This methodology follows (Álvarez *et al.,* 2013: pp. 6) and is further explained in the appendix (Section A.0) . We believe this method is better to define the access variable in Spain since it takes into the account the complicated geography of the country.

[25] We have located these ports using the MTN50.

calculated the distance from the municipalities to their nearest port (in each census year). Elevation and slope are extracted from the *Copernicus Land Monitoring System*, while the municipal surface area is computed from the IGN-2018 enclosure map after making all the relevant adjustments to adapt these results to the 7,755 municipalities that make up the dataset[26].

4. Empirical analysis.

The empirical strategy follows an Ordinary Least Squares (OLS) approach. However, railway access could have been determined by both observable and non-observable factors, such as municipalities' size or economic structure. Consequently, it is important to address this potential endogeneity issue. As mentioned in Section 2, private companies were responsible for the construction and operation of the rail tracks. Due to this, companies' decision-making regarding the layout and location of stations and stops was crucial. In this context, they could have selected municipalities with above- or below-average growth prospects, thereby leading to selective routing.

IV-strategy

To solve the endogeneity issue we have followed the methodology proposed in Banerjee *et al.,* (2020) and replicated in related studies (Berger, 2019; Büchel and Kyburz, 2020). This approach is built under the assumption that, when developing a transport infrastructure, the connection of large cities or urban areas is prioritized (from now on we will refer to these as *nodes*). Consequently, if tracks are built to connect these *nodes* in the most direct and cheapest way, access is gained ignoring the socio-economic characteristics of the municipalities' located in-between. Still, companies did not always choose the most direct route. These detours or diversions can cause selection bias, which reduces the reliability of the results. To control for this, we use an IV-strategy based on the Least Cost Paths (LCPs) between *nodes*.

These optimal routes (LCPs) reflect how the rail lines should have been if only the construction costs had been considered. That is, if companies had just taken into account the physical geography of the terrain, which is exogenous. Although *nodes* at the endpoint of these lines were not chosen at random,

[26] Copernicus is the European benchmark program for ground monitoring work.

the validity of the instrument is based on controlling for the possibility that the municipalities between these *nodes* might have had a greater probability of being crossed by a broad-gauge line.

Nodes

The first step to build this instrument is to identify the cities of origin, destination and the unavoidable points of transit of the main broad-gauge lines in 19th century Spain. The general criterion to establish these points appears in GRL1855 and GRL1877 (Maps 1 and 2). In the drafts of these laws, the trunk lines origins and destinations were stipulated. Yet, many of these projects have been lost, and in other cases, the municipalities designated as *unavoidable points of transit* were not always revealed. For this, it is critical to find out which municipalities had, prior to railways, great importance as economic, administrative and transport hubs.

To do so, we consider the 47 provincial capitals of peninsular Spain as *nodes*[27]. Then, we also examine the transportation system in preindustrial Spain. For this, we look at Roman and 19th century roads (Dufour map of 1860) to single out crossing-points[28]. In the first context, *nodes* are municipalities crossed by at least 2 main roads. Regarding 19th century roads, *nodes* are municipalities that met one of these requirements, containing (1) the crossing of 2 trunk roads; (2) the crossing of a trunk road with a secondary road or (3) the crossing of at least 3 secondary roads[29]. Following this, we found 47 *nodes* in the Roman period and 151 *nodes* in the 19th century. Finally, and out of all these, the ones with less than 5,000 inhabitants in 1860 were dropped[30]. In sum, there are 67 *nodes* that were not provincial capitals and that held a population greater than 5,000 inhabitants in 1860. By adding these *nodes* with the Provincial capitals, we get a total of 114 *nodes*.[31] Due to their special features, these *nodes* were most probably the points of origin and destination of line-sections. Therefore, for this study *nodes* are

[27] The expansion of railways occurred almost in parallel with the process of State-building. Consequently, the connection of first-order administrative entities, such as provincial capitals and towns of particular relevance conditioned the network layout (Artola *et al.,* 1978).
[28] We use the Roman and mid-19th century GIS road location to identify the pre-railway transport nodes (de Soto, 2019). (http://fabricadesites.fcsh.unl.pt/mercator-e/).
[29] The general criterion to appoint a road as "main" or "secondary" is extracted from: de Soto, P. (2019).
[30] We set a 5,000 inhabitants as threshold since it is commonly used in the literature (Álvarez-Palau *et al.,* 2017). Moreover, in the lines for which the original construction plans are available, we find no town with less than 5,000 inhabitants referred to or mentioned in the plans.
[31] For a complete list of the *nodes* see Tables A.2 and A.3 in the appendix.

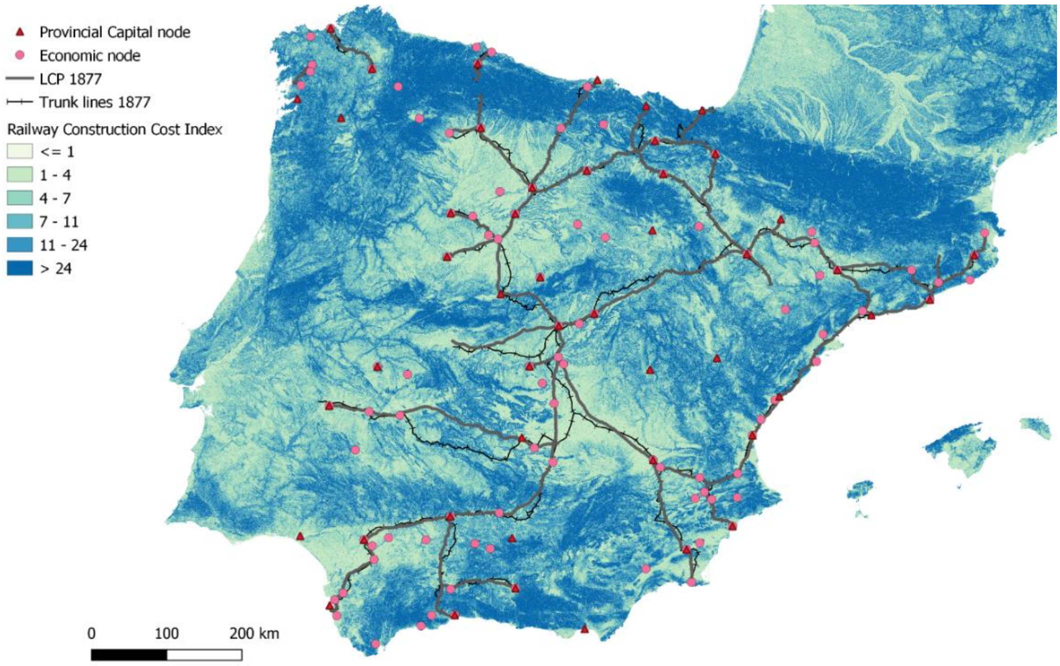
the anchor points in the Least Cost Path layout. This implies that, as points of origin and destination, their inclusion may be a source of endogeneity. As a result, *nodes*-municipalities had been excluded in the empirical analysis, reducing our sample to 7,641[32]. Furthermore, municipalities that received access prior to 1860 are excluded from this sample too (-276) since there is no reliable population data before this date.

Least cost paths (LCPs)

To establish the Least Cost Paths (LCPs), we have used the actual openings of broad-gauge lines that occurred in the period. These lines are selected only if their objective was to link two *nodes*. We exclude the routes that established inter-nodal connections gradually and over a long period of time. For the selected lines, we draw up LCP in 300x300 meter grids using the QGIS Least Cost Path mechanism. This instrument parametrizes construction costs using distance, slope and cost of crossing rivers. The construction costs (track and bridges) come from Álvarez-Palau *et al.,* (2017) and Büchel and Kyburz (2020)[33]. Maps 4 and 5 show the LCPs and the main broad-gauge network in 1877 and 1900.

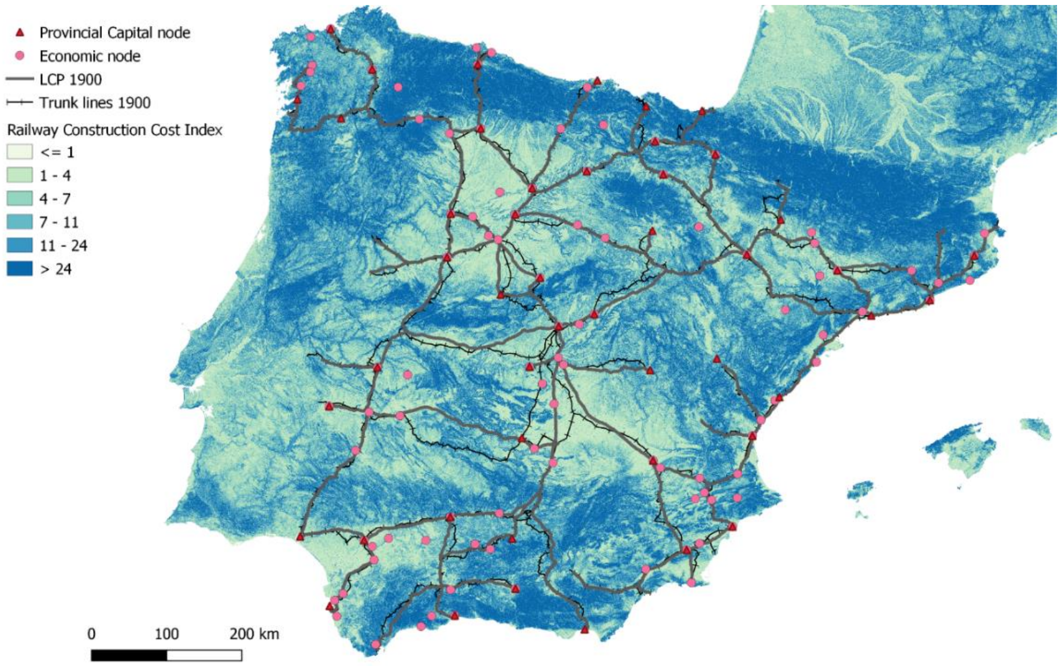
[32] I have contrasted these *nodes* with Madrazo (1984: "*Nudos de transporte*"), which only considers 40. All of these, except for Alcázar de San Juan, Osuna, Talavera de la Reina and Úbeda are *nodes* too.
[33] For more information, see Álvarez-Palau *et al.* (2017: Appendix, Table A.1) and Büchel and Kyburz (2020; Appendix A.2).

Map 4: Least Cost Paths (LCPs), railway lines and *nodes* in 1877.



Source: Own elaboration.

Map 5: Least Cost Paths (LCPs), railway lines and *nodes* in 1900.



Source: Own elaboration.

Balance Test and estimates

Once data and methodology are described, we then proceed to estimate the effect of railway access on population growth as follows: The instrumental variable, LCP_{ip}^{w}, is used to estimate the effects of access, RA_{ip}^{w}, during a wave of construction w (1860-1877; 1877-1900) on annual population growth, APG_{ip}^{t}, of municipality i in province p in the period t (1877-1900; 1900-1930; 1877-1930):

$$RA_{ip}^{w} = \propto_1 + \beta_1 LCP_{ip}^{w} + \emptyset_1 X_{ip}^{1860} + P_{1p} + \epsilon_{ip} \tag{1}$$

$$APG_{ip}^{t} = \propto_2 + \beta_2 \widehat{RA}_{ip}^{w} + \emptyset_2 X_{ip}^{1860} + P_{2p} + \eta_{ip} \tag{2}$$

where P_{1p} indicates the provincial fixed effects, and X_{ip}^{1860} is a vector that includes control variables that will be introduced and explain below.

Table 2: First-Stage: Accessibility and distance to LCPs in the first wave, 1860-1877.

	(1)	(2)	(3)	(4)
ln distance to LCP	0.731*** (0.0251)	0.730*** (0.0249)	0.714*** (0.0260)	0.695*** (0.0266)
Control variables				
Province F.E.	Yes	Yes	Yes	Yes
ln distance to *node*	Yes	Yes	Yes	Yes
ln distance to narrow-gauge	No	Yes	Yes	Yes
Additional controls	No	No	Yes	Yes
Modern roads	No	No	No	Yes
Kleibergen-Paap	846.24	855.06	754.58	679.87
Observations	7,365	7,365	7,365	7,365

Notes: The dependent variable is *ln* distance to broad-gauge in first wave, 1860-1877. Additional controls include: *ln* distance to ports, *ln* of municipal surface, *ln* of municipal elevation, *ln* of municipal ruggedness, *ln* of municipal population in 1877, population growth in previous period (1860-1877) and a binary indicator for capital of judicial district. Statistical significance is denoted by: *** p<0.01, ** p<0.05, *p<0.10. Standard errors are clustered at district level.

For the IV-strategy to be valid two conditions must be met. First, the instrumental and explanatory variable must be correlated (i.e. β is not equal to 0). Table 2 illustrates the First-Stage (or equation 1), where it can be seen that the *ln* distance to LCP is statistically significant. This result is robust when controlling for the distance to the pre-railways era road network and when adding further controls. Besides, the Kleibergen-Paap F-Statistical shown in the table is large enough to reject that there is a bias of more than 10%.

The second condition implies that there must be no correlation between the factors that determine the subsequent municipal growth and proximity to an LCP line. Since the latter is correlated with the distance to *nodes*, this condition could be broken. To avoid this problem, in the IV-strategy we control for the *ln* of distance from each municipality to its nearest *node*. Equally, Table 2 shows the relevance of physical geography in the case of Spain. Railways followed natural corridors, and as a result the LCP routes correlate with the actual broad-gauge lines more strongly than in other contexts, such as Switzerland or Sweden (Büchel and Kyburz, 2020; Berger, 2019).

Besides, it must be taken into account that LCPs choose the routes with the lowest construction costs between two *nodes*, thereby avoiding river crossings or high slopes; preference that they share with the pre-railway commercial routes. This situation determines that the proximity to an LCP route could be correlated with certain municipal economic structures. Specifically, those that are endogenous to municipalities located in the pre-railway commercial routes. To address this problem, we include the *ln* of distance to pre-industrial roads as control (See Map A.1 in appendix).

Equally, it is worth noting that the LCP algorithm prioritizes valleys and plains paths, as lowest slopes characterize these areas. A problem derived from this criterion is that these areas can also be beneficial *per se* for growth. We have thus controlled for the *ln* of municipal elevation and for the *ln* of the municipal area ruggedness. Also, we have added a series of variables to control for municipal dynamism in the pre-railway era. These controls include a binary variable, being 1 if municipality was the capital of a judicial district and 0 otherwise; the *ln* of municipal distance to ports; the *ln* of municipal population size; and population growth rate in the period prior to the construction wave. Lastly, we also control for municipal heterogeneity in surface area. All variables are described in Table A.1 of the appendix.

Table 3: Balancedness of instrument LCP (1877).

	(1)	(2)	(3)	(4)
Outcome	**Pop. 1877**	**Pop Growth 1860-1877**	**Elevation**	**Ruggedness**
ln distance to LCP	0.00426	0.0147***	0.0394***	0.0317***
	(0.00518)	(0.00364)	(0.0061)	(0.00504)
	(5)	**(6)**	**(7)**	**(8)**
Outcome	**Dist. to node**	**Dist. to road**	**Dist. to port**	**Surface**
ln distance to LCP	0.7985***	0.444***	0.175	0.0357***
	(0.03098)	(0.0615)	(0.178)	(0.00437)
Province F.E.	Yes	Yes	Yes	Yes
Observations	7,365	7,365	7,365	7,365

Note: OLS estimates from regressing each municipality outcome on the *ln* distance to the LCP. Standard errors are clustered at the district level. Statistical significance is denoted by: *** p<0.01, ** p<0.05, *p<0.10.

Besides, we analyse whether the observable characteristics of municipalities in proximity to an LCP do differ from the rest. Table 3 presents estimates from regressing explanatory variables on the *ln* distance to LCP. Results indicate that municipalities that lay close to the LCPs mainly differ to those more distant regarding its geographical features. Likewise, there is a positive correlation between the distance from an LCP to a node and to a road, which is determined on how we built the LCP. However, we can observe that the pre-railway population size and distance to ports are not statistically significant relative to the LCP distances. These results seem to indicate that the importance or dynamism of a municipality (excluding *nodes*) prior to the arrival of the railway network did not condition its location along the LCPs.

The cross-section analysis exploits the fact that the GRL1855 and GRL1877 largely influenced the development of the network. Therefore, two main waves naturally emerge: 1860-1877 and 1877-1900. Though debatable, the selection of these is far from being arbitrary. First, as Figure 1 shows, between 1848 and 1860 the expansion of the railroads was minimal. Also, the first modern population censuses were carried out in 1857 and 1860. Then, the broad-gauge network rapidly expanded until 1866, when it came to a halt as a result of a profound financial crisis. After nearly a decade of financial and political uncertainty, the GRL1877 gave way to a second wave. Following the literature, we then assume 1896 (or 1901) as the year of completion of the broad-gauge network (Artola *et al.,* 1978).

Then, for each of the waves, and equations (1) and (2), we have computed annual population growth for the periods (1877-1900; 1900-1930 and 1877-1930). To measure the impact of access in each period, the dependent variable will be the annual population growth observed in the following inter-census periods (t+1). Finally, and as we already mentioned, railroads existed before the period of study, then municipalities with *direct access* (or less than 1-hour walking distance) before 1860 are excluded from the analysis. Also, for the second wave (1877-1900), municipalities that gained *direct access* before 1877 are excluded.

Findings

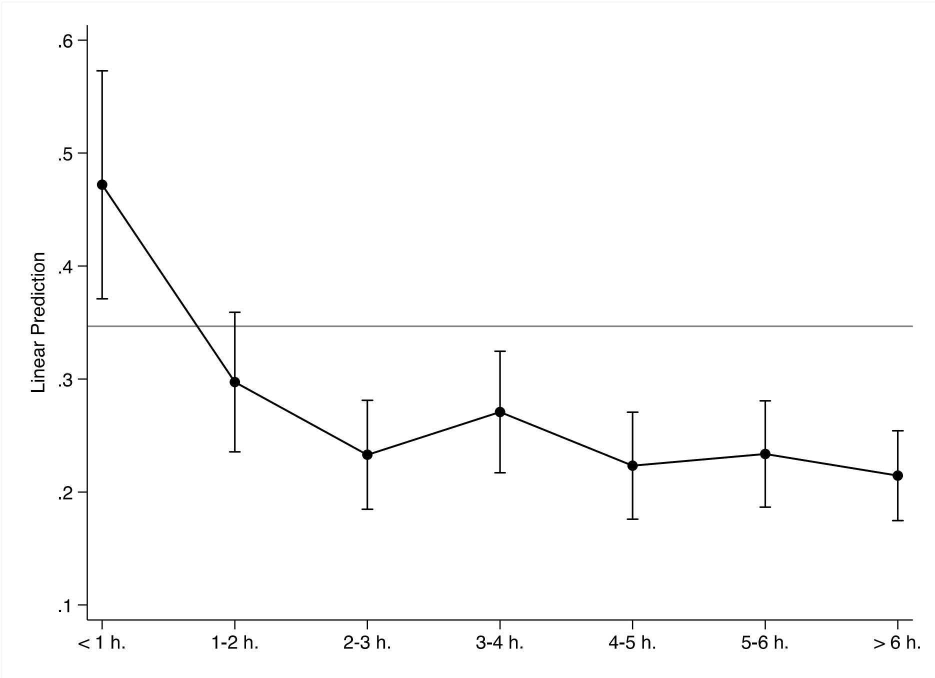
The analysis empirically tests the relationship between railway access, RA_{ip}^{w}, and municipal population growth, APG_{ip}^{t}. For this, accessibility is captured as the walking distance (in hours) from the municipal centre to its nearest access, and hence it is time-varying. To begin with, Figure 3 shows the estimated coefficient, using a simple Ordinary Least Squares (OLS), when municipalities are grouped in 7 bins (<1 hour, 1-2 hours…>7 hours). In this case, we are assessing how rail accessibility in the wave 1860-1877 affected population growth in the long-run, 1877-1930. Then, *nodes* and municipalities with *direct access* (< 1 hour) prior to 1860 are excluded. At first, there is a positive and statistically significant effect for all bins. Still, the effect is much larger for municipalities with *direct access*, while it remains stable for the rest[34].

In line with previous studies (Koopmans *et al.*, 2012; Büchel and Kyburz, 2020), these early results point to the relevance of rail accessibility. Although excluded municipalities (*nodes* and municipalities with *direct access* prior to 1860) can be labelled as "winners", growing at around 1% between 1877 and 1930, municipalities gaining *direct access* in the 1860-1877 wave grew well-above the sample average. Moreover, it does not seem that distance mattered much for the rest, as coefficients remain somewhat similar. Furthermore, the coefficients are positive and statistically significant, thereby implying that displacement effects might not be able to explain this whole story[35].

[34] To check for robustness, we have repeated the exercise with distinct intervals (30 minutes, 60 minutes...), and found 1-hour naturally arises as cut-off point. Additionally, we have carried out a similar exercise for the periods 1877-1900 and 1900-1930. See Figure A.3 in the appendix.

[35] Between 1877 and 1900, and using provinces as a spatial unit, Silvestre (2005) found that internal migrations were moderate, then accelerated at the turn of the 20th century reaching a peak in the 1920s. More

Figure 3. Rail accessibility and municipal population growth: 1877-1930.



Notes: The dependent variable is annual population growth for the period of analysis. The graph shows the marginal effects and a 95% confidence interval from equation (2). As specified, it includes several controls: *ln* distance to narrow-gauge, *ln* distance to *nodes, ln* distance to roads, *ln* distance to ports, *ln* of municipality surface, *ln* of municipal elevation, *ln* of municipal ruggedness, *ln* of municipal population and population growth in previous period. Likewise, a binary indicator for capital of Judicial District and Province fixed effects are included. Notice that *nodes* and municipalities gaining access before 1860 are excluded. Standard errors are clustered at the district level. The grey-line depicts the Spanish average annual population growth between 1877 and 1930 (excluding *nodes* and municipalities gaining access before 1860).

To delve further into the matter and address endogeneity, we estimate equations (1) and (2) using two distinct measures of accessibility, $\mathrm{RA}_{\mathrm{ip}}^{\mathrm{w}}$. First, rail accessibility is captured with a discrete variable. In line with the above findings, a binary variable is created, *direct access*, being 1 if a municipality is less than 1-hour walking distance to the nearest broad-gauge access point and 0 otherwise. Second, we also capture accessibility with a continuous variable, as the distance (*ln* distance to access point) from the municipal centre to the nearest access point. While using a continuous variable is more precise, the discrete variable is easier to interpret; therefore we will comment the results of the discrete variable. In any case both variables show similar results.

Table 4 presents the results for the wave 1860-1877. Results are arranged in Panels A and B according to the way accessibility is captured. Columns (1) and (2) display the coefficients (OLS; IV) for the long-run, 1877-1930, while columns (3, 4, 5 and 6) show the short-run periods, 1877-1900 and 1900-

recently, Caballero, S. (2020) pointed out that from 1850 to 1870, internal migrations were higher than expected. Emigrations followed a similar trend (Sánchez-Alonso, 2000).

1930. As usual, it includes all the control variables introduced above. On the whole, municipalities with *direct access* exhibited greater population growth and their coefficients are statistically significant at the 1% level, except for the discrete variable IV estimates of the period 1877-1900, which stays in the limit. In particular, municipalities with *direct access* experienced, on average, a 0.219% higher annual growth than the rest between 1877 and 1930. If a 500 inhabitant's municipality gained access during this construction wave, it would have around 80 inhabitants more than a non-access one at the end of the period. Note that the long term IV estimates (2), are larger than the OLS ones (1). The greater magnitude suggests that the railway lines built between 1860 and 1877 may have traversed areas with worse than average growth prospects. This result is consistent with the literature, since companies prioritized the construction of direct trunk lines instead of increasing the network accessibility in areas that were experiencing greater dynamism (see section 2).

Table 4: Rail accessibility (broad-gauge) and population growth in Spain [**wave**: 1860-1877]

	1877-1930		**1877-1900**		**1900-1930**	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
			Panel A			
Direct access (=1)	0.219*** (0.0466)	0.346*** (0.0659)	0.119* (0.0679)	0.0932 (0.0757)	0.295*** (0.0502)	0.555*** (0.0875)
Constant	1.306*** (0.208)	1.257*** (0.203)	0.983*** (0.277)	0.993*** (0.272)	1.140*** (0.229)	1.023*** (0.222)
			Panel B			
ln distance to access point	-0.0829*** (0.0169)	-0.0901*** (0.0203)	-0.0610** (0.0241)	-0.0583** (0.0233)	-0.0976*** (0.0184)	-0.115*** (0.0254)
Constant	1.300*** (0.211)	1.292*** (0.209)	0.962*** (0.278)	0.965*** (0.274)	1.180*** (0.235)	1.164*** (0.234)
Province F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,365	7,365	7,365	7,365	7,365	7,365

Notes: The dependent variable is annual population growth for each period. For a list of controls see Figure 3.

On the other hand, Columns (3, 4, 5 and 6) show the impact in the short term periods: 1877-1900 and 1900-1930. As expected, the magnitude is much greater in the latter one which provides evidence in support of a conventional view, railways were especially noteworthy in the early 20th century, once the network was almost complete (Herranz-Loncán, 2007). In this regard, population growth in Spain accelerated (see Figure 1) whereas internal migrations and emigrations intensified. Though we cannot identify the mechanisms, this study provides further empirical evidence of the relevance of transport infrastructure in shaping the spatial distribution of population[36].

Table 5: Rail accessibility (broad-gauge) and population growth in Spain [**wave**: 1877-1900]

	OLS	IV
	(1)	**(2)**
	Panel A	
Direct access (=1)	0.167*** (0.0456)	0.395*** (0.0930)
Constant	1.126*** (0.215)	1.113*** (0.214)
	Panel B	
ln distance to access point	-0.0430** (0.0170)	-0.0438* (0.0244)
Constant	1.229*** (0.216)	1.229*** (0.215)
Province F.E.	Yes	Yes
Additional controls	Yes	Yes
Observations	6,706	6,706

Notes: The dependent variable is annual population growth for 1900-1930. For a list of controls see Figure 3. Notice that *nodes* and municipalities gaining access before 1877 are excluded

Regarding the Second wave, 1877-1900 witnessed the expansion of the railroads to territories that were previously ignored. Moreover, it is the moment when the main transversal lines were built. Table 5 shows the results obtained after estimating equations (1) and (2). In general, the magnitude of the

[36] Pérez (2017), recently found that the expansion of railroads stimulated the transition out of farming in 19th century Argentina. For this, a longitudinal database was used with individuals before and after the railways.

coefficients is lower when compared with the one resulting from estimating the impact of the 1860-1877 wave in this period (Table 4, columns 5 and 6). The reasons can be twofold. It might be that the impact of this transport infrastructure was not fully accomplished by 1930. Likewise, it is also probable that these lines connected areas that had worse intrinsic growth prospects than those already linked (see section 2); thus generating less synergies and spill-over effects, and consequently, less growth.

Finally, it is worth noting that in both waves the control variables appear with the expected sign and that the coefficients are mostly significant and different from zero. Interestingly enough, the variables that relate with the geographical features are of great magnitude; Table A.5 in the appendix briefly summarises this. It should also be noted that although these analyses include control variables that take into account observable characteristics of Spanish municipalities, there might be unobservable features that correlate with both the instrument and their growth potential. In an optimal scenario, a panel data analysis is usually performed to control for this. However, the low periodicity of censuses in Spain prevents us from carrying out this analysis in a reliable manner[37]. In any case, we believe that the control variables, together with the robustness tests that we are presenting below, are sufficient to guarantee the consistency of the model.

Robustness

This section presents robustness checks to reinforce the above findings. First, as previously discussed, access is measured as the distance from the municipal centre to the nearest broad-gauge station or stop. In Table A.4 of the appendix we carry out the same empirical analysis using *distance to the rail track* and the *distance to both broad and narrow gauge stations and stops* instead, being both results very consistent with the ones shown above.

Another concern comes from heterogeneity in the density of settlement. In the north-western territories, municipalities contain several singular entities and population is disperse, whereas in most of the other areas municipalities are basically a singular entity. As a result, railway access might not

[37] While in Switzerland population census were carried out every 5 to 10 years, in Spain they were undertaken in 1857, 1860, 1877, 1887, 1900, 1910, 1920 and 1930. Furthermore, most of the broad-gauge network expansion took place in two well-defined construction waves. Therefore, we are unsure that a panel will add extra value.

be fully capturing this spatial aspect. Then, we remove municipalities in Galicia and Asturias from the analysis. The results, displayed in Table A.6, are no affected by this territorial specificity.

Furthermore, we carried out a robustness test to control for the possibility that the results are affected by municipal size. According to their population and following the literature, municipalities are classified as *small* (< 1,000 inhabitants), *medium* (1,000-5,000) and *large* (>5,000 inhabitants), and then we estimate the model. Again, there is a statistically significant relationship between access and growth, see results in Table A.7. Nevertheless, the effect seems to be of greater magnitude in medium-sized municipalities. Interesting enough, this result is in line the New Economic Geography literature, since it predicts that usually big and middle sized municipalities benefit most from transport infrastructure improvements and access to new markets.

5. Conclusions

In this study, we have examined the impact of rail accessibility on municipal population in peninsular Spain between 1877 and 1930. Our results confirm that municipalities with *direct access* experienced, on average, more rapid growth. Moreover, it is observed that this effect was much weaker from 1877 to 1900, which goes in line with existing knowledge. In sum, we find further empirical evidence of the transformative power of transport infrastructures.

Yet, the case of Spain provides further insight. As literature pointed out, the process of State-building and, above all, a physical geography that had constrained the transportation system in the past, affected the morphology of the national network. In order to link the central plateaus to coastal areas, railways largely followed natural corridors. Under these circumstances, we thus find what seems to be a causal relationship; as municipalities along or near these pathways witnessed greater growth, thereby reinforcing pre-existing spatial dynamics.

Finally, and although the spatial distribution of population in Spain has long been explored (Ayuda *et al,*. 2010; Franch *et al.,* 2013), this study sheds further light. As the municipalities that gained access grew more rapidly, railways could have set in motion a process that, once agglomeration economies became more relevant in the 20th century (Beltrán-Tapia *et al.,* 2018), had long-lasting effects, fuelling the depopulation of certain areas (Collantes and Pinilla, 2011) and the concentration of economic activity.

References

Bogart, D., Álvarez-Palau, E.,Satchell, M., Shaw-Taylor, L., & You, X. (2017). *Railways and growth: evidence from mid-nineteenth century England and Wales*. University of Cambridge. Mimeo.

Alvarez, E., Franch, X., and Martí-Henneberg, J. (2013). Evolution of the territorial coverage of the railway network and its influence on population growth: The case of England and Wales, 1871-1931. *Historical Methods*, *46*(3), 175–191.

Artola, M., Tedde, P., and Anes, R. (1978). *Los Ferrocarriles en España, 1844-1943: El Estado y los ferrocarriles*. Madrid: Servicio de Estudios del Banco de España.

Atack, J., Bateman, F., Haines, M., and Margo, R. A. (2010). Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850-1860. *Social Science History*, *34*(2), 171–197.

Atack, Jeremy. (2013). On the use of geographic information systems in economic history: The American transportation revolution revisited. *Journal of Economic History*, *73*(2), 313–338.

Atack, Jeremy, Haines, M., and Margo, R. A. (2008). Railroads and the Rise of the Factory: Evidence for the United States, 1850-70. *National Bureau of Economic Research*, *w14410*.

Ayuda, M. I., Collantes, F., and Pinilla, V. (2010). From locational fundamentals to increasing returns: the spatial concentration of population in Spain, 1787–2000. *Journal of Geographical Systems*, *12*(1), 22–50.

Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, *145*, 102442.

Barquín, R. (1997). Transporte y precio del trigo en el siglo XIX: creación y reordenación de un mercado nacional. *Revista de Historia Económica / Journal of Iberian and Latin American Economic History*, *15*(1), 17–48.

Barquín, R., and Larrinaga, C. (2020). Los límites de la intervención pública. Ingenieros de caminos y ferrocarriles en España (1840-1877). *Historia y Política*, *43*, 27–56.

Barquín, R., Pérez, P., and Sanz, B. (2012). La influencia del ferrocarril en el desarrollo urbano español (1860-1910). *Revista de Historia Económica / Journal of Iberian and Latin American Economic History*, *30*(3), 391–416.

Beltrán-Tapia, F. J., Díez-Minguela, A., Martínez-Galarraga, J., and Tirado-Fabregat, D. A. (2019). *Capital* *humano y desigualdad territorial. El proceso de alfabetización en los municipios españoles desde la ley Moyano hasta la Guerra Civil.* Madrid: Banco de España, Estudios de Historia Económica 74.

Berger, T., & Enflo, K. (2017). Locomotives of local growth: The short- and long-term impact of railroads in Sweden. *Journal of Urban Economics*, *98*, 124–138.

Braun, S., and Franke, R. (2019). *Railways, Growth, and Industrialisation in a Developing German Economy, 1829-1910*. Munich Personal RePEc archive, Working Paper nr. 93644.

Büchel, K., and Kyburz, S. (2020). Fast track to growth? Railway access, population growth and local displacement in 19th century Switzerland. *Journal of Economic Geography*, (1538).

Santiago-Caballero, C. (2020). Domestic migrations in Spain during its first industrialization, 1840s-1870s. *Cliometrica*.

Cabanes Martín, A., and González Sanz, R. (2009). *El tiempo del tren: Las velocidades comerciales en las líneas españolas desde sus orígenes hasta la actualidad*. V Congreso de Historia Ferroviaria.

Cabo, A., and Vigil, M. (1973). *Condicimientos geográficos; Edad antigua*. Madrid: Alfaguara.

Collantes, F., and Pinilla, V. (2011). *Peaceful surrender: the depopulation of rural Spain in the twentieth century*. Cambridge Scholars Publishing.

Cuellar, D. (2003). Un modelo ferroviario de ciclo minero: auge y declive del ferrocarril en el Sureste andaluz. *Transportes, Servicios y Telecomunicaciones*, *6*, 40–64.

Díez-Minguela, A., Martinez-Galarraga, J., and Tirado-Fabregat, D. A. (2018). Regional Income Inequality in Spain 1860–2015. In *Regional Inequality in Spain* (pp. 81–103).

Donaldson, D., and Hornbeck, R. (2016). Railroads and American Economic Growth: A "Market Access" Approach *. *The Quarterly Journal of Economics*, *131*(2), 799–858.

Fishlow, A. (1965). *American Railroads and the Transformation of the Ante-bellum Economye.* Cambridge: Harvard University Press.

Fogel, R. (1964). *Railroads and American economic growth*. Baltimore: John Hopkins Press.

Franch, X., Morillas-Torné, M. ., & Martí-Henneberg, J. . (2013). Railways as a factor of change in the distribution of population in Spain, 1900-1970. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, *46*(3).

Goerlich Gisbert, F. J., Ruiz González, F., Chorén, P., and Albert, C. (2015). *Cambios en la Estructura y Localización de la Población. Una visión de largo plazo (1842-2011).* Bilbao: Fundación BBVA..

Gómez-Mendoza, A. (1982). *Ferrocarriles y cambio económico en España, 1855-1913: un enfoque de nueva historia económica*. Madrid: Alianza Editorial.

González-Val, R., Tirado-Fabregat, D., & Viladecans-Marsal, E. (2017). Market potential and city growth: Spain 1860–1960. *Cliometrica*, *11*(1), 31-61.

Gregory, I. N., and Henneberg, J. M. (2010). The Railways, Urbanization, and Local Demography in England and Wales, 1825–1911. *Social Science History*, *34*(2), 199–228.

Gutierrez, E., Moral-Benito, E., Oto-Peralías, D., and Ramos, R. (2020). *Spatial Distribution of Population in Spain: An Anomaly in European Perspective.* Madrid: Banco de España, 2028.

Herranz-Loncán, A. (2005). La reducción de los costes de transporte en España (1800-1936). *Cuadernos Económicos de ICE*, (70), 183–203.

Herranz-Loncán, A. (2007). Infrastructure investment and Spanish economic growth, 1850-1935. *Explorations in Economic History*, *44*(3), 452–468.

Hornung, E. (2015). Railroads and growth in Prussia. *Journal of the European Economic Association*, *13*(4), 699–736.

Jedwab, R., and Moradi, A. (2016). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. *Review of Economics and Statistics*, *98*(2), 268–284.

Koopmans, C., Rietveld, P., and Huijg, A. (2012). An accessibility approach to railways and municipal population growth, 1840–1930. *Journal of Transport Geography*, *25*, 98–104.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, *99*(3), 483-499.

Krugman, P. (1993). First nature, second nature, and metropolitan location. *Journal of regional science*, *33*(2), 129-144.

Madrazo, S. (1984). *El sistema de transportes en España*. Madrid: Colegio de Ingenieros de Caminos Canales y Puertos.

Martí-Henneberg, J. (2013). European integration and national models for railway networks (1840–2010). *Journal of Transport Geography*, *26*, 126–138.

Martínez-Galarraga, J., Rosés, J. R., and Tirado, D. (2015). The long-term patterns of regional income inequality in Spain, 1860–2000. *Regional Studies*, *49*(4), 502-517.

Medina-Albaladejoo, F. J., Añón Higón, D., Díez-Minguela, A., and Lana-Berasain, J.-M. (2020). Old wine in new wineskins? Understanding the cooperative movement: Catalonia, 1860–1939. *European Review of Economic History*.

Melander, E. (2020). *Transportation Technology, Individual Mobility and Social Mobilisation*. Competitive Advantage in the Global Economy, Working Paper nr. 471.

Mimeur, C., Queyroi, F., Banos, A., and Thévenin, T. (2018). Revisiting the structuring effect of transportation infrastructure: An empirical approach with the French railway network from 1860 to 1910. *Historical Methods*, *51*(2), 65–81.

Muñoz Rubio, M. (2005). *Historia de los Ferrocarriles de Vía Estrecha en España*. Madrid: Fundación de los Ferrocarriles Españoles.

Muro, J., Nadal, F., and Urteaga, L. (1996). *Geografía, estadística y catastro en España: 1856-1870*. Barcelona; Serval.

Nadal, J. (1975). *El fracaso de la revolución industrial en España, 1814-1913*. Barcelona: Ariel.

Ortúñez, P. P. (2016). Gobernanza en la regulación del sector ferroviario español. Un análisis histórico: 1844-1941. *Revista de Historia Industrial*, *25*(61), 79–109.

Oto-Peralías, D. (2020). Frontiers, warfare and economic geography: The case of Spain. *Journal of Development Economics*, *146*, 102511.

Pascual, P. (1984). Ferrocarrils i industrialització a Catalunya. *Recerques: Història, Economia, Cultura*, *17*, 43–72.

Pérez, S. (2017). *Railroads and the Rural to Urban Transition: Evidence from 19th-Century Argentina*. Standford, Mimeo.

Prados de la Escosura, L. (2017). *Spanish Economic Growth, 1850–2015*. London: Palgrave Studies in Economic History.

Pro, J. (2019). *La Construcción del Estado en España: Una historia del siglo XIX*. Madrid: Alianza Editorial.

Ringrose, D. R. (1972). *Los transportes y el estancamiento económico de España*. Madrid: Tecnos.

Rosés, J. R., Martínez-Galarraga, J., and Tirado, D. A. (2010). The upswing of regional income inequality in Spain (1860-1930). *Explorations in Economic History*, *47*(2), 244–257.

Sánchez-Alonso, B. (2000). Those Who Left and Those Who Stayed behind: Explaining Emigration from the Regions of Spain, 1880-1914. *The Journal of Economic History*, *60*(3), 730–755.

Schwartz, R., Gregory, I., and Thévenin, T. (2011). Spatial History: Railways, Uneven Development, and Population Change in France and Great Britain, 1850–1914. *Journal of Interdisciplinary History*, *42*(1), 53–88.

Silvestre, J. (2005). Las migraciones interiores durante la modernización económica de España, 1860-1930. *Cuadernos Económicos de ICE*, *69*, 157–182.

Tafunell, X., and Carreras, A. (2005). *Estadísticas históricas de España: siglo XIX-XX*. Bilbao: Fundación BBVA.

Tirado, D., Paluzie, E., and Pons, J. (2002). Economic integration and industrial location: the case of Spain before World War I. *Journal of Economic Geography*, *2*(3), 343-363.

Tortella, G. (1974). *Los orígenes del capitalismo en España: banca, industria y ferrocarriles en el siglo XIX*. Madrid: Tecnos.

Appendix

A.0 Calculating access from municipal centres to railway stations and stops

In this section, we present the method that we have used to take into account the orography of the terrain in calculating the distance between the municipal centers and the stations or stops on the network.

First, using GIS, we extract the elevation of the municipal centers and stations of the network using the "Copernicus Land Monitoring System" data (See Figure A.0, points *X* and *Y*). Once we have these results, we calculate the average slope of the route (*Z*) by performing a simple triangulation: dividing the difference in elevation (*C*) by the distance between both points (*A*) (expressed in meters). The next step is to calculate the hypotenuse of this triangle (*B*), for this, we perform the square root of the sum of the squares of (*C*) and (*A*): *"Square Root of ((C ^ 2) + (A ^ 2))"*. Subsequently, following the criteria that appear in Table A.0 (extracted from Alvarez-Palau *et al* (2013), pp.6), and using the results of the slope as the reference, we assign an average speed to each route. Finally, we will transform the distance "*B*" from km to hours walking; for which we take into account the average speed that we have previously calculated (based on the slope). We use the result of this analysis as the main explanatory variable through the study, with the name "Distance to nearest access".

Using this variable has the disadvantage of assuming a speed of 4km/h (adjusted) for all traffic, be it ox carts, mules or pedestrians. On the other hand, the advantage of the method is that it allows us to take into account the orography effect on access more consistently; which we believe is essential for our case study.

Likewise, it should be emphasized that we apply this methodology because calculating the Least Cost Paths (see section 4) from each municipality to their nearest station implies an extremely high computational load. However, we believe that since the distances we work with are relatively short, and given the corrections, the result will hardly be affected.

Table A.0. Speed function according to terrain slopes

Slope (%)	Speed function (Km/h)
(0-3)	4
(3-8)	- 40 x Slope + 5.2
(8-20)	- 8.33 x Slope + 2.66
20	1

Figure A.0. Description of variables

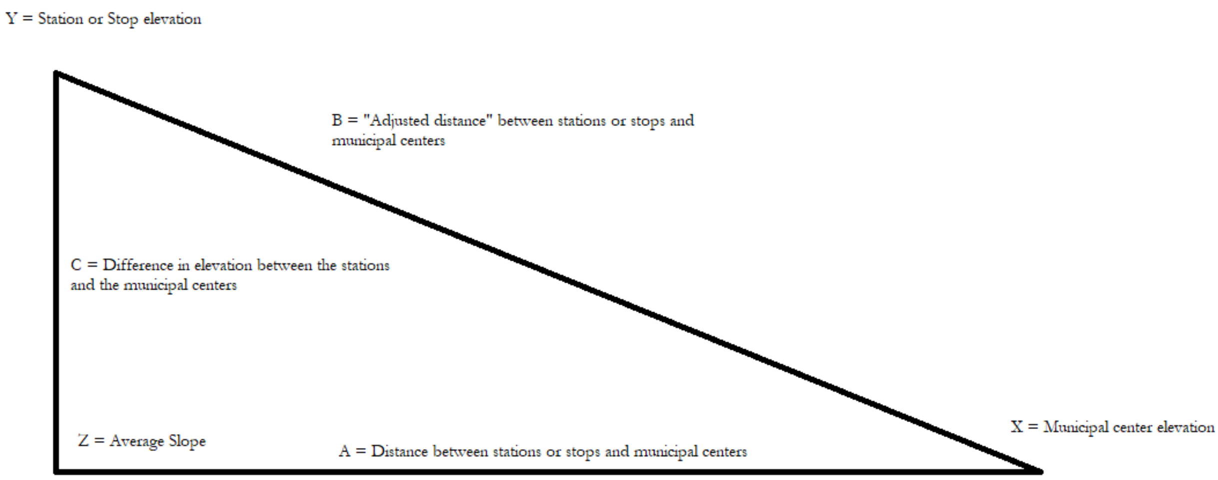


Table A.1 Description of main variables

Variable	Description	Source
Annual Population Growth (APG)	100.[log(POP_{t2}) – log (POP_{t1}))/(t2 – t1)]	Population censuses
Treatment variables		
Direct access	Binary indicator, equals 1 if the distance between a municipal centre and the closest broad-gauge railway access is less than 1 hour, 0 otherwise.	Esteban-Oliver (2020)
ln distance to broad-gauge	Log of distance between a municipal centre and nearest broad-gauge access (in hours)	Esteban-Oliver (2020)
Control variables		
ln distance to narrow-gauge	Log of distance between a municipal centre and nearest narrow-gauge access (in hours)	Esteban-Oliver (2020)
ln distance to road	Log of distance between a municipal centre and nearest main road (in hours)	Mercator-E Project.
ln distance to node	Log of distance between a municipal centre and nearest *node* (in hours)	Esteban-Oliver (2020)
ln distance to port	Log of distance between a municipal centre and nearest port (in hours)	Estadísticas de Comercio Exterior y Estadísticas de Comercio de Cabotaje IGN-2018
ln of surface	Log of the munincipal surface area (in meters)	IGN-2018
ln of elevation	Log of the munincipal elevation (in meters)	IGN-2018
ln of ruggedness	Log of the mean ruggedness (in meters)	IGN-2018
District capital	Binary indicator, equals 1 if capital of the judicial district, 0 otherwise.	Population censuses
ln of population	Log of municipal population.	Population censuses

Table A.2 Population of *nodes* (provincial capitals) in 1877.

Municipio	Provincia	Pop. 1877	Municipio	Provincia	Pop. 1877
Madrid	Madrid	417,424	Castelló de la Plana	Castellón	23,393
Barcelona	Barcelona	353,581	Badajoz	Badajoz	22,965
València	Valencia	165,466	Toledo	Toledo	21,297
Sevilla	Sevilla	134,318	Lleida	Lleida	20,369
Málaga	Málaga	121,987	Albacete	Albacete	18,958
Murcia	Murcia	91,805	Lugo	Lugo	18,909
Zaragoza	Zaragoza	89,222	Salamanca	Salamanca	18,409
Granada	Granada	76,005	Ourense	Ourense	17,270
Cádiz	Cádiz	65,028	Girona	Girona	16,226
Valladolid	Valladolid	52,181	Cáceres	Cáceres	14,816
Córdoba	Córdoba	49,755	Palencia	Palencia	14,493
Bilbao/Bilbo	Vizcaya	41,452	Zamora	Zamora	13,853
Santander	Cantabria	41,021	Ciudad Real	Ciudad Real	13,589
Coruña, A	Coruña	40,987	Segovia	Segovia	13,514
Almería	Almería	40,338	Logroño	La Rioja	13,393
Alacant/Alicante	Alicante	36,312	Huelva	Huelva	13,125
Oviedo	Asturias	35,487	Huesca	Huesca	12,880
Burgos	Burgos	30,794	Teruel	Teruel	12,786
Vitoria/Gasteiz	Álava	26,648	León	León	12,421
Pamplona/Iruña	Navarra	25,630	Ávila	Ávila	11,031
Tarragona	Tarragona	24,709	Cuenca	Cuenca	10,732
Jaén	Jaén	24,395	Guadalajara	Guadalajara	10,418
Pontevedra	Pontevedra	24,308	Soria	Soria	6,928
San Sebastián/Donostia	Guipúzcoa	23,943			

Source: Population census of 1877.

Table A.3 Population of *nodes* (other relevant towns in the past) in 1877.

Municipio	Provincia	Pop. 1877	Municipio	Provincia	Pop. 1877
Cartagena	Murcia	75,908	Almagro	Ciudad Real	8,630
Jerez de la Frontera	Cádiz	64,535	Alcaudete	Jaén	8,498
Lorca	Murcia	52,934	Barbastro	Huesca	8,478
Santiago de Compostela	Coruña	35,043	Alcalá de Guadaira	Sevilla	8,227
Alcoy/Alcoi	Alicante	32,497	Aranjuez	Madrid	8,154
Gijón	Asturias	30,591	Almansa	Albacete	7,964
Reus	Tarragona	27,595	Marbella	Málaga	7,947
Antequera	Málaga	25,664	Mérida	Badajoz	7,390
Écija	Sevilla	25,237	Caldas de Reis	Pontevedra	7,361
Orihuela	Alicante	24,629	Alcañiz	Teruel	7,327
Tortosa	Tarragona	24,057	Mora	Toledo	7,219
Puerto de Santa María	Cádiz	22,122	San Esteban de Gormaz	Soria	7,194
Manresa	Barcelona	18,537	Teo	Coruña	6,979
Carmona	Sevilla	17,349	Fraga	Huesca	6,761
Yecla	Murcia	15,276	Medina del Campo	Valladolid	6,548
Utrera	Sevilla	15,103	Granollers	Barcelona	6,369
Xàtiva	Valencia	14,701	Sagunto/Sagunt	Valencia	6,287
Don Benito	Badajoz	14,692	Madridejos	Toledo	6,263
Valdepeñas	Ciudad Real	13,867	Chinchilla de Monte-Aragón	Albacete	6,080
Baena	Córdoba	13,328	Barrios, Los	Cádiz	6,075
Villareal	Castellón	12,887	Aranda de Duero	Burgos	6,072
Alcalá de Henares	Madrid	12,317	Nava del Rey	Valladolid	6,035
Andújar	Jaén	11,976	Villarcayo de M. de Cast. la Vieja	Burgos	5,918
Figueres	Girona	11,956	Zafra	Badajoz	5,595
Chiclana de la Frontera	Cádiz	11,713	Astorga	León	5,585
Carballo	Coruña	11,445	Piélagos	Cantabria	5,500
Villena	Alicante	11,424	Caudete	Albacete	5,435
Ponferrada	León	11,252	Medina de Rioseco	Valladolid	5,381
Coín	Málaga	10,065	Cervantes	Lugo	5,380
Vinaròs	Castellón	9,528	Aguilar de Campoo	Palencia	5,370
Trujillo	Cáceres	9,428	Blanes	Girona	5,323
Avilés	Asturias	8,979	Ocaña	Toledo	5,123
Toro	Zamora	8,759	Monzón	Huesca	4,921
Tarazona	Zaragoza	8,632			

Source: Population census of 1877.

Table A.4 Accessibility and population growth using different measures of accessibility.

	Distance to broad-gauge stations[a]		Distance to broad and Narrow-gauge stations		Distance to track[a]	
	OLS	IV	OLS	IV	OLS	IV
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
ln distance to broad-gauge	-0.0829***	-0.0901***	-0.101***	-0.116***	-0.0827***	-0.128***
	(0.0169)	(0.0203)	(0.0180)	(0.0250)	(0.0173)	(0.0241)
Constant	1.300***	1.292***	0.962***	0.965***	1.277***	1.214***
	(0.211)	(0.209)	(0.278)	(0.274)	(0.213)	(0.211)
Province fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,365	7,365	7,365	7,365	7,365	7,365

Notes: The dependent variable is annual population growth (%). Additional controls include *ln* distance to narrow gauge railway lines (equations 1, 2, 5 and 6), *ln* distance to ports, *ln* of the municipality surface, *ln* of the municipality elevation, *ln* of the municipality area ruggedness mean, *ln* of the municipal population in 1877, population growth in the previous period (1860-1877) and a binary indicator for capital of judicial district. Transport Nodes and Provincial. [a] Distance to narrow gauge stations appears as a control variable. Nodes excluded from the sample. Standard errors are clustered at the district level. Statistical significance is denoted by: *** p<0.01, ** p<0.05, *p<0.10

Table A.5 Main cross-section results: Determinants of annual population growth rates (1877-1930)

Variables	OLS	IV
	(1)	**(2)**
ln distance to broad-gauge	-0.0829***	-0.0901***
	(0.0169)	(0.0203)
Control variables		
ln distance to narrow-gauge	-0.0737**	-0.0736**
	(0.0341)	(0.0339)
ln distance to node	0.0811***	0.0863***
	(0.0271)	(0.0280)
ln distance to road	-0.00752	-0.00606
	(0.00950)	(0.00974)
ln distance to port	0.122***	0.121***
	(0.0304)	(0.0301)
ln of surface	0.155***	0.155***
	(0.0156)	(0.0155)
ln of elevation	-0.214***	-0.213***
	(0.0241)	(0.0239)
ln of ruggedness	-0.136***	-0.135***
	(0.0195)	(0.0192)
Judicial District	0.0683***	0.0679***
	(0.0262)	(0.0261)
ln of Population in 1877	-0.172***	-0.173***
	(0.0180)	(0.0179)
Population Growth 1860-1877	0.0724***	0.0724***
	(0.0147)	(0.0146)
Constant	1.300***	1.292***
	(0.211)	(0.209)
Province F.E.	Yes	Yes
Observations	7,365	7,365
R-squared	0.391	0.391

Notes: The dependent variable is annual population growth (%). Transport Nodes and Provincial Capitals excluded from the sample. Standard errors are clustered at the Judicial District level. Statistical significance is denoted by: *** p<0.01, ** p<0.05, *p<0.1

Table A.6 Robustness analysis: Cross-section estimates excluding Asturias and Galicia

	OLS	IV
	(1)	**(2)**
ln distance to broad-gauge	-0.0836***	-0.0907***
	(0.0176)	(0.0212)
Constant	1.331***	1.324***
	(0.226)	(0.224)
Province F.E.	Yes	Yes
Additional controls	Yes	Yes
Observations	6,992	6,992

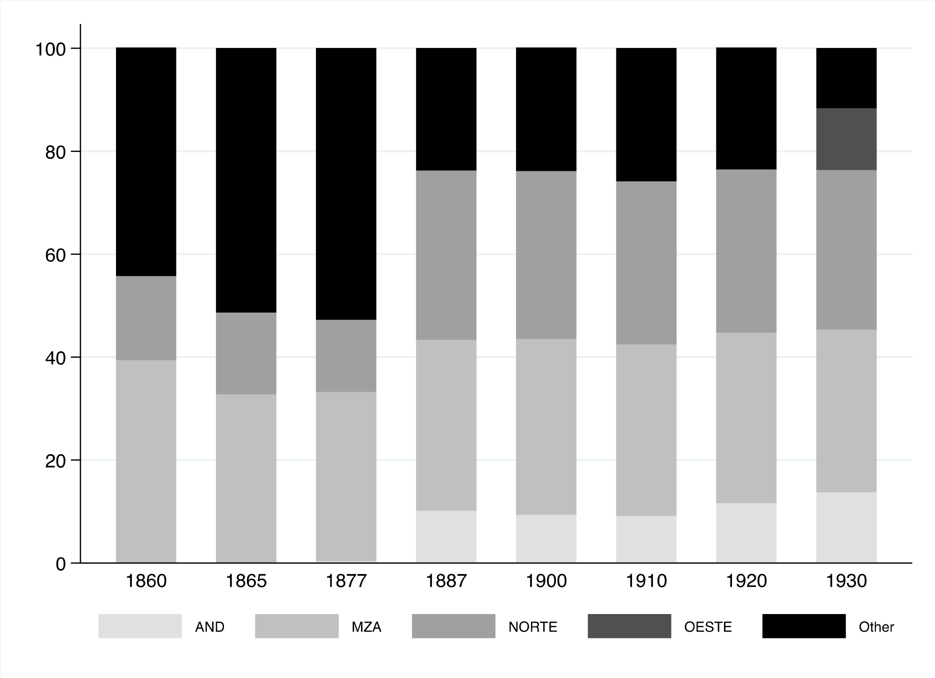
Notes: The dependent variable is annual population growth (%) between 1877 and 1930. For a list of controls see Table A.4.

Table A.7 Robustness analysis: Cross-section estimates based on municipal population sizes

	Small		**Medium**		**Large**	
	OLS	**IV**	**OLS**	**IV**	**OLS**	**IV**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
ln distance to broad-gauge	-0.0716***	-0.0725***	-0.0965***	-0.103***	-0.0634**	-0.0849**
	(0.0209)	(0.0271)	(0.0216)	(0.0226)	(0.0297)	(0.0357)
Constant	1.819***	1.817***	1.284***	1.280***	0.493	0.472
	(0.310)	(0.313)	(0.284)	(0.281)	(0.853)	(0.810)
Province F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,987	3,987	2,868	2,868	510	510

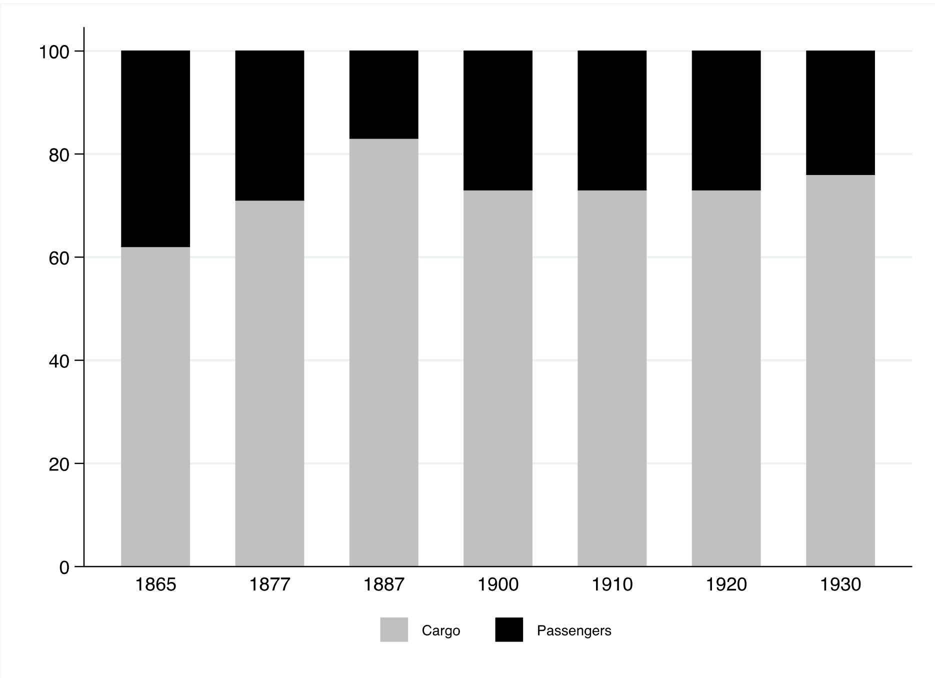
Notes: The dependent variable is annual population growth (%) between 1877 and 1930. Municipalities have been organised according to their size (Small<1000, 1000<Medium<5000, Big>5000). For a list of controls see Table A.4

Figure A.1. Ownership of broad-gauge lines in Spain by company (in % of total track)



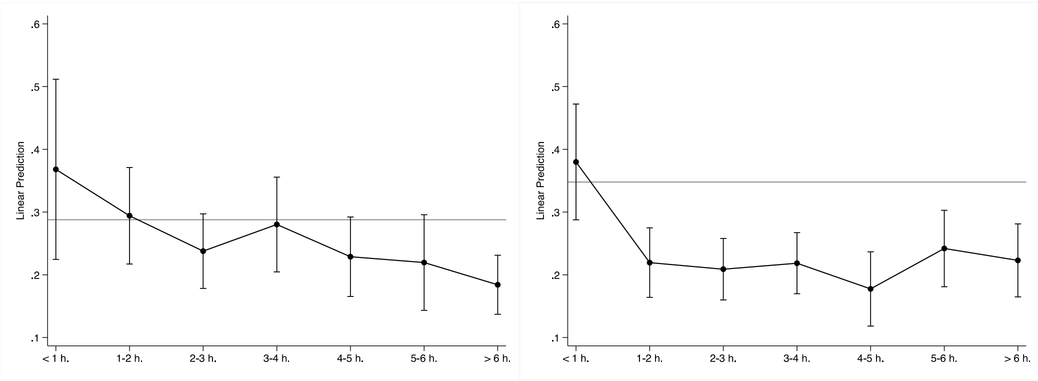
Source: Own elaboration, based on Cuellar (2020) and unpublished material.
Notes: AND, MZA, NORTE and OESTE stand for *Compañía de los Ferrocarriles Andaluces*, *Compañía de los ferrocarriles de Madrid a Zaragoza y Alicante*, *Compañía de los Ferrocarriles del Norte* and *Compañía nacional de los ferrocarriles del Oeste* respectively.

Figure A.2. Percentage of Spanish railway companies' income due to cargo or passengers, 1865-1930.



Source: Own elaboration based Artola *et al.* (1978: Chapter 2, pp. 485-510).

Figure A.3. Rail accessibility and municipal population growth: 1877-1900 (left) and 1900-1930 (right).



Notes: The dependent variable is annual population growth for the period of analysis. The graphs show the marginal effects and a 95% confidence interval from equation (2). The grey-lines depict the Spanish average annual population growth in each period. (excluding *nodes* and municipalities gaining access before 1860 and 1877). For a list of controls see Table A.4.

Map A.1: Main Roads (1860) and Least Cost Paths (1877).

Source: Own elaboration based on the Pau de Soto road network http://fabricadesites.fcsh.unl.pt/mercator-e/results-2/modern-roads/. The road network is based on the Dufour map of Spain and Portugal, 1860.