UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Cambridge Working Paper Economics

Cambridge Working Paper Economics: 1636

TRACING THE EVOLUTION OF AGGLOMERATION ECONOMIES: SPAIN, 1860-1991

Francisco J. Beltrán Tapia
University of Cambridge

Alfonso Díez-Minguela
Universitat de València

Julio Martinez-Galarraga
Universitat de València

3 June 2016

This article attempts to quantify how the effect of agglomeration economies on population growth has evolved over time. Using district population in Spain between 1860 and 1991, recorded approximately every decade, this article examines whether initial population affects subsequent population growth. Our results show that, while the relationship between these two variables hardly existed during the second half of the 19th century, this link increased significantly between 1910 and 1970, although this trend was abruptly interrupted by the Civil War and the autarkic period that followed. The intensity of this relationship debilitated in the 1970s, a process that continued during the 1980s as rural out-migration diminished and de-industrialisation hit traditional manufacturing sectors. Our findings also stress that agglomeration economies were stronger in medium-size districts, especially from 1960 onwards, thus suggesting that congestion costs began to mitigate the benefits arising from agglomeration economies in the largest locations.

Tracing the Evolution of Agglomeration Economies: Spain, 1860-1991

Francisco J. Beltrán Tapia
University of Cambridge
(fjb38@cam.ac.uk)

Alfonso Díez-Minguela
Universitat de València
(alfonso.diez@uv.es)

Julio Martinez-Galarraga
Universitat de València
(julio.martinez-galarraga@uv.es)

May 30, 2016

Abstract

This article attempts to quantify how the effect of agglomeration economies on population growth has evolved over time. Using district population in Spain between 1860 and 1991, recorded approximately every decade, this article examines whether initial population affects subsequent population growth. Our results show that, while the relationship between these two variables hardly existed during the second half of the 19th century, this link increased significantly between 1910 and 1970, although this trend was abruptly interrupted by the Civil War and the autarkic period that followed. The intensity of this relationship debilitated in the 1970s, a process that continued during the 1980s as rural out-migration diminished and de-industrialisation hit traditional manufacturing sectors. Our findings also stress that agglomeration economies were stronger in medium-size districts, especially from 1960 onwards, thus suggesting that congestion costs began to mitigate the benefits arising from agglomeration economies in the largest locations.

Keywords: Agglomeration economies, Regional growth, Spain

JEL classification: N93, N94, O18, R11, R12

Acknowledgements: This paper was presented at the 2016 Economic History Society Annual Conference at Cambridge. We would like to thank the comments received there, as well as those by Javier Silvestre and Rafael González-Val. This research received financial support from the Spanish Ministry of Science and Innovation (projects HAR2012-30732, ECO2012-33286 and ECO2015-65049-C12-1-P).

1. Introduction

Agglomeration economies refer to the benefits that arise when firms and people locate near one another (Glaeser, 2010). These might appear from the scale and/or the diversity of the local economy (Marshall, 1890; Jacobs, 1969; Henderson, 2003). In particular, the spatial concentration of economic activity increases market access, thus resulting in cheaper and more varied inputs, as well as allowing the sharing of risk and indivisible facilities (i.e. airports, universities, hospitals). Besides, denser locations enable a more efficient matching between firms and workers or buyers and sellers, both in terms of quantity and quality of matches; and facilitate knowledge spillovers within and across industries (Duranton and Puga, 2004)[1]. Yet, agglomeration is also associated with expensive housing, long commutes and pollution, among other costs. There appears then to be a trade-off between increasing returns and congestion costs (Fujita and Thisse, 2002)[2].

In a seminal contribution, Krugman (1991) suggests that, through a process of circular causation, the interaction between economies of scale and transport costs might lead to the emergence of an industrialised «core» and agricultural «periphery»[3]. In fact, this economic geography model also entails that initial conditions (i.e. population density) are self-reinforcing, thereby stressing the role of history on the spatial concentration of industry. In this vein, Puga (1999) stresses the relevance of workers mobility to income differentials. If the agglomeration of economic activity raises wages and workers are perfectly mobile, these will tend to relocate near industrial clusters. Consequently, structural change should be one of the major driving forces of urbanisation, as Michaels et al. (2012) show. However, urban agglomeration continues, even in countries where employment shifted away from agriculture long ago, thereby stimulating the need for further research[4].

The relationship between economic density and productivity has been at the core of most debates on agglomeration economies. Ciccone and Hall (1996) pioneered this strand of the literature studying productivity differentials within the United States. On the whole, these authors found that variation in output per worker across states partly reflects differences in the density of economic activity. Since then, several studies have attempted to quantify the effect of economic density on productivity. Although results may vary according to the level of aggregation, period of study and/or estimation method, it is

---

[1] See Combes and Gobillon (2015) for a recent review.

[2] Using a novel methodology, Combes et al. (2012) estimate the elasticity of urban costs with respect to population, and provide evidence in support of the trade-off between agglomeration economies and urban costs.

[3] See also Krugman and Venables (1995), Venables (1996), Fujita et al. (1999) and Combes et al. (2008).

[4] As the *World Urbanisation Prospects: The 2014 Revision* indicates, all regions are expected to urbanise further in the next decades.

somewhat accepted that density increases the productivity of firms and workers (Combes and Gobillon, 2015)[5]. Yet, most empirical studies within this regional economics literature have a static or short-term perspective, thereby ignoring long-run dynamics[6]. Data availability has usually constrained research on how agglomeration economies vary along the process of economic development. In fact, long-run studies have usually replaced employment for population, and use population growth as a proxy for economic dynamism (Beeson et al., 2001; Dobkins and Ioannides, 2001; Michaels et al., 2012; Desmet and Rappaport, 2016).

In addition, the literature in the field of urban economics has focused on the study of one of the most peculiar empirical regularities, the Gibrat's law, which suggests that city growth is independent of its initial size (Clark and Stabler, 1991; Gabaix, 1999; Eeckhout, 2004). In short, if the spatial concentration of population follows a random process, agglomeration economies would be irrelevant. But, if this is not the case, and hence Gibrat's law is violated, agglomeration economies might play a critical role. Similarly, location fundamentals or «first nature» advantages (i.e. climate, geography) are expected to be essential in the location of preindustrial populations (Beeson et al., 2001; Dobkins and Ioannides, 2001; Davis and Weinstein, 2002; Bosker and Buringh, 2015)[7]. Yet, industrialisation brought about a new scenario where agglomeration economies overcame other major forces (Gabaix and Ioannides, 2004; Duranton, 2007; Rossi-Hansberg and Wright, 2007; Michaels et al., 2012). In this regard, it appears that the link between initial population and subsequent population growth only started to be visible in the late 19th century, accelerated from then onwards and slowed down in the late 20th century. Michaels et al. (2012) argue that this pattern is related to reallocation away from agriculture, a process that weakened during the latter period. However, congestion costs, especially in the largest locations, may also contribute to explaining the mitigation of agglomeration economies in recent decades (Puga, 1999; Graham, 2007; Combes et al., 2012)[8].

In this paper, we analyse how the relationship between agglomeration economies and the spatial distribution of population has evolved over time. Our contribution is twofold.

---

5 Ciccone and Hall (1996) find that doubling employment density increases average labour productivity by 6 per cent. Melo et al. (2009) use a meta-analysis, which includes 729 estimates from 34 studies, to analyse how study characteristics might affect results. Combes and Gobillon (2015) survey the existing literature and report that the elasticity of productivity with respect to density usually ranges between 0.04 and 0.07. Furthermore, recent studies have shown that agglomeration economies appear to have a greater impact in developing economies, such as China and India (Combes et al., 2013; Chauvin et al., 2014).

6 Ciccone and Hall (1996); Desmet and Fafchamps (2005).

7 Cronon (1991) introduced the terms «first nature» and «second nature» referring to *"original, prehumen nature"* and *"artificial nature that people erect atop first nature"* respectively.

8 See also Krugman and Venables (1995) and Venables (1996).

First, we introduce a novel dataset that traces the evolution of the Spanish population at the district level during the period 1860-1991. The data, which comprises 464 districts and is recorded on a decadal basis, allow us to capture the transition from a pre-industrial society to a modern economy. Second, our empirical analysis examines whether initial size affects subsequent growth. If agglomeration economies play a role, orthogonal growth would not hold and large districts would grow more rapidly than small ones, thus violating Gibrat's law and increasing spatial concentration. In order to isolate the effect of initial size from other potential determinants of population growth, we have considered climatic and geographic information to capture the «first nature» advantages of each district. We also control for other issues, such as the «capital effect» and the economic dynamism of neighbouring locations. Potential endogeneity is further addressed by instrumenting the size of the local economy using historical urban population.

Our results show that the relationship between district size and population growth hardly existed during the second half of the 19th century. Although interrupted by the Spanish Civil War and the autarkic period that followed, the link between these two variables increased significantly between 1910 and 1970. These findings, in line with previous studies, illustrate the relevance of structural change and agglomeration economies in the shaping of a modern economy. The intensity of this relation slightly weakened in the 1970s, a process that continued during the 1980s as rural out-migration slowed down and de-industrialisation hit traditional manufacturing sectors (i.e. metallurgy, extractive). Lastly, we also find that agglomeration economies appear to have a greater impact on medium-size districts, especially from 1960 onwards, thus suggesting that congestion costs might have started to mitigate the benefits arising from economic density in the largest locations.

The rest of the paper is structured as follows. Section 2 provides a brief historical account of the evolution of the Spanish economy between 1860 and 1991. Data and descriptive analyses are introduced in Section 3, whilst Section 4 presents the empirical analysis. Section 5 discusses our main findings and Section 6 concludes.

2. Historical background: Spain 1860-1991

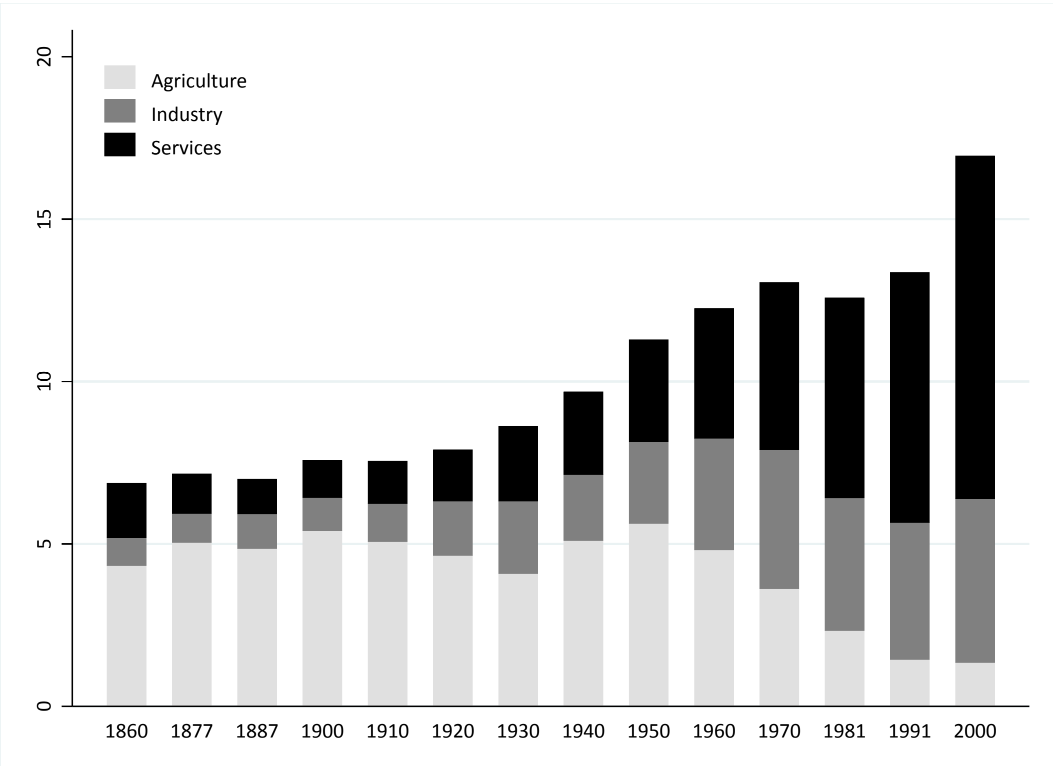
Our period of study, 1860-1991, spans one and a half centuries, thus covering most of the process of economic development. During this time, the Spanish economy undertook a profound structural transformation that turned a predominantly agricultural society into a modern economy by the late 20th century: labour shifted away from agriculture to industry and services, and income per capita increased accordingly (see table 1 and figure 1).

TABLE 1. Real GDP, population and per capita GDP growth, 1850-2000

	GDP	Population	Per capita GDP
1850-1883	1.8	0.4	1.4
1884-1920	1.3	0.6	0.7
1921-1929	3.8	1.0	2.8
1930-1952	0.8	0.9	0.0
1953-1958	4.7	0.8	3.9
1959-1974	6.9	1.1	5.8
1975-1986	2.5	0.7	1.8
1986-2000	3.5	0.2	3.3

Source: Prados de la Escosura (2008, 288). Annual average logarithmic rates.

Figure 1. Labour force by main economic activity in Spain, 1860-2000 (unit: millions)



Source: Population censuses

The Spanish economy initiated the early stages of modern economic growth in mid-19th century. Economic growth was initially fostered by the integration of the national market and the adoption of industrial innovations. However, up to the First World War, economic growth rates progressed at a slow pace, industrialisation advanced with difficulties and unevenly distributed across space, and structural change was limited. By 1910, as figure 1 illustrates, more than two thirds of the labour force were still enrolled in agriculture.

The integration of Spain's domestic market received a strong impulse in the middle of the 19th century[9]. Before that, as a consequence of the persistence of barriers and limitations to internal trade, the national market was fragmented into various local and regional markets that were largely unconnected. Local tariffs and regulations restricting trade were widespread and weights and measures differed across regions. In addition, transport costs were very high due to low public investment in transport infrastructures, the use of traditional means of transport and the particular geography of Spain, which was highly rugged and lacked an extensive water transport system. As a result, regional commodity markets were scarcely integrated and prices markedly differed from one region to another. It is true though that some interdependence in commodity prices had existed since the 18th century (Ringrose, 1998).

The successive political reforms of the 19th century promoted market liberalisation. Laws were unified, legal support was given to property rights, and tariffs and local restrictions on home commerce were eliminated (Tedde, 1994). In addition, the expansion of the rail network brought with it major changes that favoured the progressive development of the domestic market. The first line was finished in 1848, covering the 28 kilometres that separated Barcelona and Mataró. By 1866 the railway linked up Spain's main economic centres and by 1901 all the provincial capitals were connected to the railway (Wais, 1987); at that time, the railway network covered a distance of 10,827 kilometres (Herranz, 2005)[10] and the country's infrastructure stock as a share of GDP rose from 4,27% in 1850 to 27,2% in 1900 (Herranz, 2007). Transport improvements, particularly the completion of Spain's railways network, favoured the fall in transport costs and the creation of a national market for most important commodities during the second half of the 19thcentury[11]. Indeed, regional wheat prices converged in this period (Peña and Sánchez-Albornoz, 1984)

In parallel to the integration of the domestic market, manufacturing industries became increasingly concentrated in space (figure 2). While inland regions experienced a substantial process of deindustrialisation (with the exception of Madrid), Spanish industrialisation was mainly led by Catalonia and the Basque Country (Nadal, 1987)[12]. By 1910, the contribution of these regions to Spanish industrial output was 31.6% and 6.9%,

---

[9] See Rosés et al. (2010) for a detailed description.

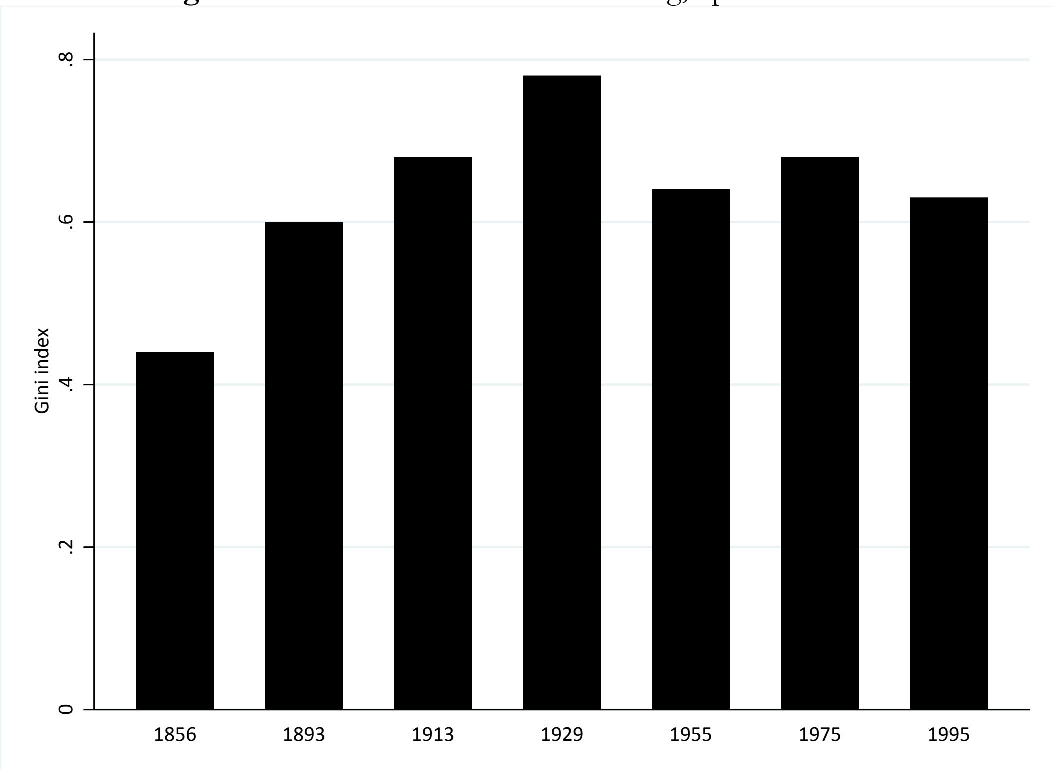
[10] By 1990, the length of the railway network was 12,560 km (Gómez Mendoza and San Román, 2005).

[11] According to Herranz (2005), the introduction of the railway represented a reduction of up to 86% in haulage costs in 1878.

[12] In Catalonia, the cotton industry, with a tradition that stretched back to the 18th century, gradually became mechanised in the 19th century. In the Basque Country, the iron and steel industry underwent rapid growth in the last quarter of the century.

while their population only represented 10.5% and 3.4% respectively (Tirado and Martinez-Galarraga, 2008). In addition, internal migratory flows were relatively low throughout most of the 19th century (Silvestre, 2005). Due to the predominance of agrarian activities and their subsequent seasonality, an important part of these movements was temporary and occurred over short distances (Silvestre, 2007). Indeed, up to the 1920s the number of permanent internal migrations remained small (see figure 3)[13]. International migration, on the other hand, experienced a notable increase in the late 19th century and the first decades of the 20th century, mainly heading to Latin America (Sánchez-Alonso, 2000).

Figure 2. Concentration in manufacturing, Spain 1856-1995



Source: Paluzie et al. (2004)

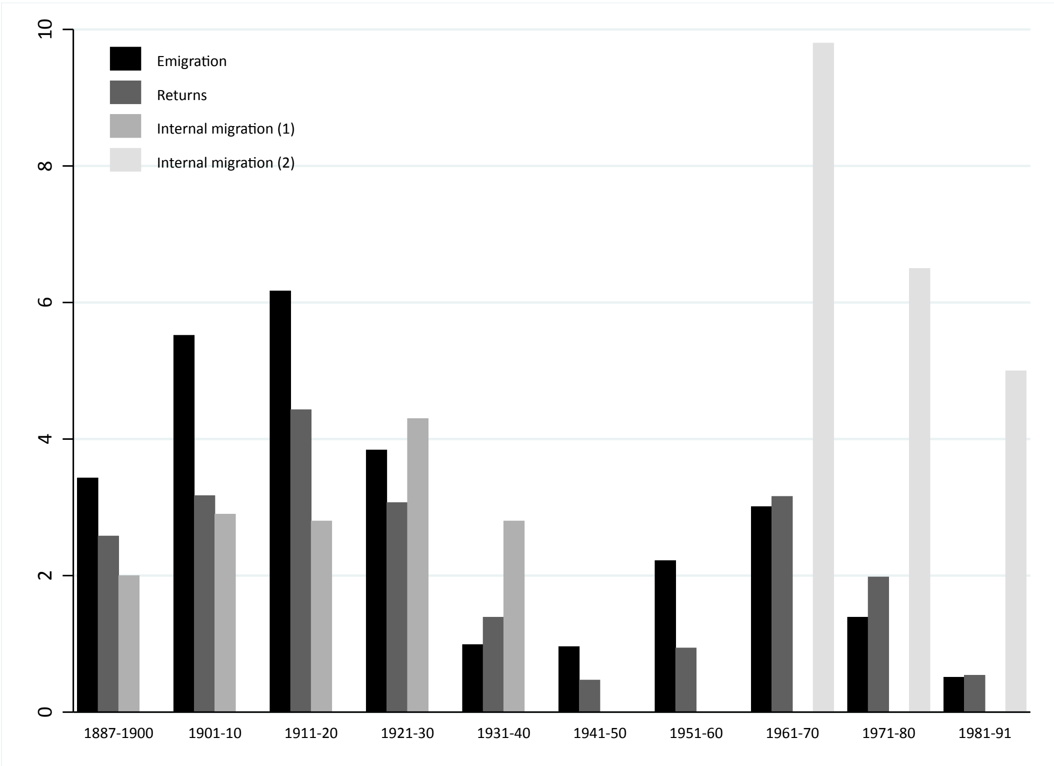
The integration of the Spanish market further progressed throughout the interwar years, especially with a substantial increase of paved roads, which complemented the previous development of the railway network (Herranz, 2005)[14]. In addition, although the notable progress of electrification mitigated previous energy restrictions traditionally faced by Spain's industry and the number of industrial locations expanded, mainly to Madrid and Zaragoza, the spatial concentration of manufacturing continued. The increasing market

[13] The number of Spaniards residing outside their province of origin was relatively small, about 9.3 per cent by 1910 (Silvestre, 2005).

[14] The road network increased from 36,300 km in 1900 to 109,176 in 1935. By 1960 the distance covered was 130,644 and reached 162,298 km in 1990 (Gómez Mendoza and San Román, 2005).

integration was accompanied by large inter-regional migrations: Spaniards left declining regions, which were mainly rural and agrarian, to reallocate in the richest regions, which were more urban and specialised in industry and services. Yet, these migratory flows did not originated in the poor areas of the south of the country because these provinces were far away from the industrial centres (Silvestre, 2005)[15]. In parallel to these developments, structural change accelerated and the share of agrarian employment decreased substantially while economic growth rates significantly increased.

Figure 3. Emigration, returns and internal migration in Spain by period (per 1,000 people)



Source: (1) Emigration and returns: Sánchez-Alonso (1995) and Bover and Velilla (1999)[16].
(2) Internal migrations: Paluzie et al. (2009b: Table 1, p. 255)
Note: Internal migrations refer to the annual rate (per 1,000) of born in another province during the inter-census period from 1877 to 1940. Yet, between 1961 and 1991, it is the annual rate (1,000) of residents in another province (at previous census date) during the inter-census period (García-Coll and Stiwell, 1999). These two indicators are dissimilar hence they are not strictly comparable. See Paluzie et al. (2009b) for more information.

The Spanish Civil War (1936-1939) and the first years of Franco's regime supposed a major setback, not only to economic growth, but also to economic integration. The autarkic policy that followed the Civil War came hand in hand with a tight regulation of commodity and input markets, including state control of prices and quantities in most

[15] This process favoured the convergence registered in regional wages that had begun in the mid-19th century (Rosés and Sánchez-Alonso, 2004).
[16] We would like to thank Olympia Bover for kindly sharing the data with us.

goods. Although these policies created a false impression of price convergence, internal trade hardly increased. In addition, due to the lack of investment in infrastructure, transport costs remained unaltered during the 1940s and early 1950s. Economic growth and structural change therefore came to a halt: agrarian employment actually increased during the 1940s and it took twenty years to return to the pre-Civil War per capita income levels.

The economic liberalisation and stabilisation measures introduced at the end of the 1950s, together with foreign assistance, favoured the transition of the Spanish economy toward a new phase of economic development (Prados de la Escosura et al., 2012). This period was characterised by high economic growth rates and by the lead taken by the industrial sector in the country's economic activity. New investments in infrastructures such as roads, railways and communication networks led to further reductions in internal transport costs. Spanish economic growth in the 1960s was also characterised by the growing mobility of the labour force that was becoming increasingly concentrated in the big cities. Rural exodus towards cities, as well as to more developed European countries (figure 3), resulted in a substantial decline in agrarian employment and an increase of the share of manufacturing, construction and services sectors (Ródenas, 1994; Bover and Velilla, 1999; Bentolila, 2001). Contrary to the previous phase of the 1920s, migrants from the southern provinces now played a key role and massively participated on migratory flows. However, although to a lower extent, migrants' destinations were still limited to a relatively small number of large cities, mainly Madrid and Barcelona[17]. Regarding international migration, a new wave of out-migration flows took place in 1960-73, with more than 100,000 workers migrating per year to the core European countries.

However, the crisis of the 1970s, which in the case of Spain stretched well into the 1980s, put a brake on these upward trends and GDP growth rates were lower than in the previous decades. The concentration of manufacturing industries somewhat receded during these years, thus causing the spatial distribution to present a bell-shaped evolution in the long-term (figure 2). Furthermore, traditional industries (i.e. mining, metallurgy) underwent severe reconversion processes in the 1980s. Importantly, inter-regional migration rates fell in the 1970s and early 1980s, arguably as a result of the high unemployment registered during those years (Bentolila and Blanchard, 1990; Bentolila and Dolado, 1991).

[17] In 1930, almost half of the population living in a province different from the birth province was residing in Madrid (22.9%) and Barcelona (22.9%). By 1970, the percentages were similar, 23.8 and 23.7% respectively (Paluzie et al., 2009b, 248).

The new phase in Spanish economic growth, which started after the entry of the country into the European Union in 1986, was no longer linked to the leadership of industrial production, but rather to that of the services and construction sectors. Internal migrations were now characterised by an increase in the dispersion of migratory flows due to the growing importance of services, an economic sector that is much less spatially concentrated than industry. Increasing congestion costs, such as the rise in housing prices, the higher weight of amenities and other aspects related to the quality of life or the effect of redistributive policies would also account for the lower intensity of migratory flows in the last decades (Bover and Velilla, 1999; Bentolila, 2001). Yet, these declining inter-province migrations, shown in figure 3, were partially counterbalanced by growing intra-province migrations (Paluzie et al., 2009b). In addition, a new wave of investment in infrastructure helped to further reduce transport cost across Spanish regions and also across national borders. Huge investments in freeways, high-speed railway and telecommunications were developed during these years and led to major advances in the integration of the internal Spanish market and its connection to international markets. In this respect, the accession of Spain to the EU in 1986 bolstered the Spanish economy, thus further promoting the catching-up process with the most developed countries during the 1990s. The tertiarisation of the economy was completed at the same time that substantial GDP per capita growth rates were reached (Prados de la Escosura and Rosés, 2009).

Different studies have examined the relevance of agglomeration economies and the spatial distribution of population in Spain. On the one hand, using provinces as unit of analysis, Martinez-Galarraga et al. (2008) find that, in line with Ciccone and Hall (1996), since the mid-19th century, doubling employment density increases industrial labour productivity by around 3-5%. It appears though that this relationship declines over time. In addition, Pons and Tirado (2008) analyse the inter-provincial disparities in the distribution of economic activity throughout the 20th century and conclude that the evolution of regional income inequality has eventually been shaped by agglomeration economies, while the impact of the initial geographic conditions has decreased over time[18].

On the other hand, exploring the evolution of the 100 largest cities during the 20th century, Lanaspa et al. (2003) find that a convergent pattern of growth dominated between 1900 and 1970 and divergence followed thereafter. While this study is just concerned with

[18] Other studies have shown the relevance of agglomeration economies in the industrial sector during the early stages of economic growth in Spain (Tirado et al., 2002; Martinez-Galarraga, 2012); examined and verified the existence of the wage equation in industrial wages (Paluzie et al., 2009a; Tirado et al., 2013); and proved the direct relationship between internal migration decisions and market access in the 20th century (Paluzie et al., 2009b).

the upper tail of the distribution, Goerlich and Mas (2008; 2009) employ all municipalities to illustrate a tendency towards spatial concentration over the whole 20th century, especially between 1950 and 1981[19]. Besides, they also stressed the relevance of the "capital status": provincial capitals attracted a significant bulk of the population from neighbouring areas. Analysing provinces, Ayuda et al. (2010) find that location fundamentals or «first nature» explain the spatial distribution of population before industrialisation (up to 1900). Disparities in population density increased from then onwards once «second nature» factors grew in relevance. The initial effect of «first nature» was thus reinforced by «second nature», which is closely related to increasing returns and thus the presence of agglomeration economies. More recently, González-Val et al. (2015) explore the relationship between market potential and city growth during the period 1860-1960. Urban growth was first the result of location fundamentals, whereas market potential appears to be significant over the 20th century. Although these studies suggest the increasing importance of agglomeration economies over time, further analysis is required.

3. Data

In order to better understand the long-run evolution of agglomeration economies in Spain, we have built a panel data set that traces the population of 464 districts from 1860 to 1991[20]. Based on the Population Censuses that were carried out approximately every decade, our dataset thus covers 13 periods. Within the framework of an integrated economy, the use of population rather than income data in measuring economic activity has the advantage of taking into account that migration flows respond to income differences and tend to mitigate them. Regional differences in productivity might then be better reflected in population figures (Glaeser et al., 1995; Beeson et al., 2001)[21].

---

[19] Goerlich and Mas (2008; 2009) are based on a database that covers all Spanish censuses from 1901 to 2001, with municipalities homogenised to their structure in 2001 (Goerlich et al., 2006).

[20] Due to data availability the Canary Islands districts -Arrecife, Guía, La Laguna, Las Palmas, Orotava, Santa Cruz de la Palma, and Santa Cruz de Tenerife-, have been excluded from our sample. The period of study finishes in 1991 because the population category that had been employed since 1860 by the Spanish Statistical Agency (*Población de hecho*) disappears from later censuses.

[21] Our study is somewhat constrained by data availability. Using population, instead of working-age population or employment, might not capture relevant changes in population structure. While migration flows increase the number of workers in the receiving districts, the working-age population decreases in sending regions. Although population growth may initially reflect these flows, it may also result from differential demographic patterns present in younger populations. Employing population data could thus lead to overestimate the effect of agglomeration economies. We acknowledge this issue, though we think it would be more evident from 1991 to 2011.

The unit of analysis is the district or *Partido Judicial*, an administrative category composed of several municipalities[22]. Employing districts as units of analysis presents several advantages (Beeson et al., 2001; Desmet and Fafschamps, 2005; Partridge et al., 2008). On the one hand, districts better capture the potential effects of agglomeration economies than cities because they allow taking into account their hinterland, as well as avoiding the comparability problems generated by the rise of metropolitan areas. On the other hand, given that we cover the whole Spanish territory, district-level data avoids the sample selection bias usually present in the literature focusing on cities. These studies only consider settlements above a certain threshold, thus only focusing on those that have been relatively successful and missing those that did not grow enough to reach that limit or those that declined and fell below that figure. These two features of the data are crucial not only because most population has traditionally lived in rural areas, but also because rural out-migration was an essential dimension of how the spatial distribution of the population evolved. Michaels et al. (2012, 536) show that examining both rural and urban areas significantly enhances our understanding of the urbanisation process. As these authors point out, the unit of analysis should be stable over time (548). Given that during the period under analysis, legislative changes have somewhat affected these administrative boundaries, we have homogenised our panel dataset using the administrative boundaries existing in 1860. We therefore rely on district boundaries that are consistent over the whole sample period. Overall, the average surface area is 1,075 squared kilometres, which allow us to capture metropolitan areas (i.e. Madrid, Barcelona), but at a lower level of aggregation than provinces (NUTS3), thus reducing potential distortions arising from the Modifiable Areal Unit Problem (MAUP) (Briant et al., 2010)[23]. By means of comparison, the average size of a US county is around 1,500 squared kilometres (Michaels et al., 2012, 551).

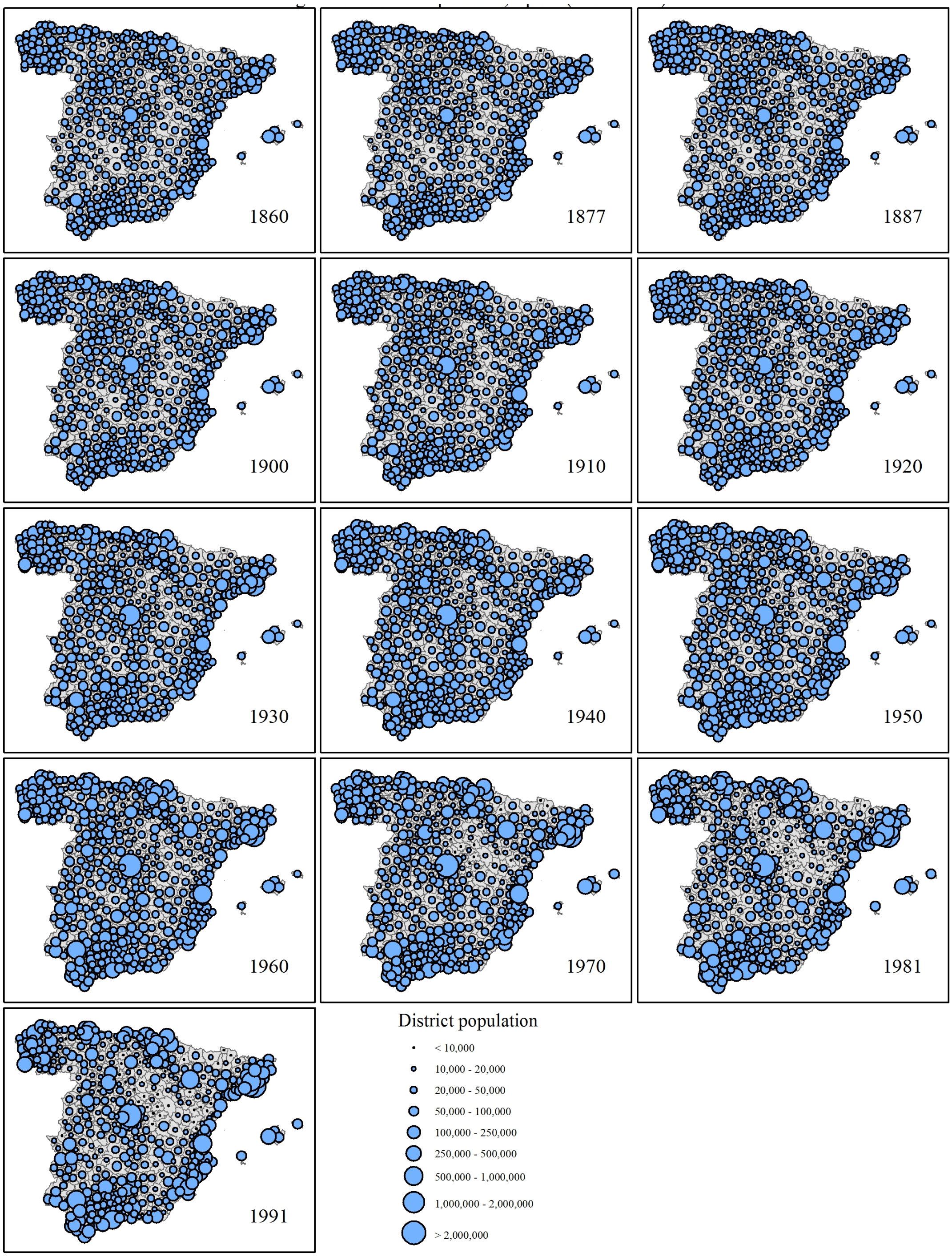
The Spanish population increased significantly during our period of study: from around 15.6 million people in 1860, to 18,5 in 1900, 23.6 in 1930, 30.4 in 1960, reaching forty million in the 1990s. Its spatial concentration also underwent major changes throughout this period. Map 1 depicts how district population evolved between 1860 and 1991. Generally speaking, districts grew in size, yet unevenly, showing a tendency to concentrate along the coast and around Madrid, the capital city.

---

[22] The spatial unit chosen to study agglomeration economies is still an open issue. Theoretically, it depends on the type of agglomeration effect but, in practice, the effect of choosing different spatial units is not that important (Briant et al., 2010; Combes and Gobillon, 2015, 294).

[23] The Modifiable Areal Unit Problem (MAUP) refers to the sensitivity of results to the zoning system employed in the empirical analysis. Figure A.1 in the appendix shows the surface area distribution of districts.

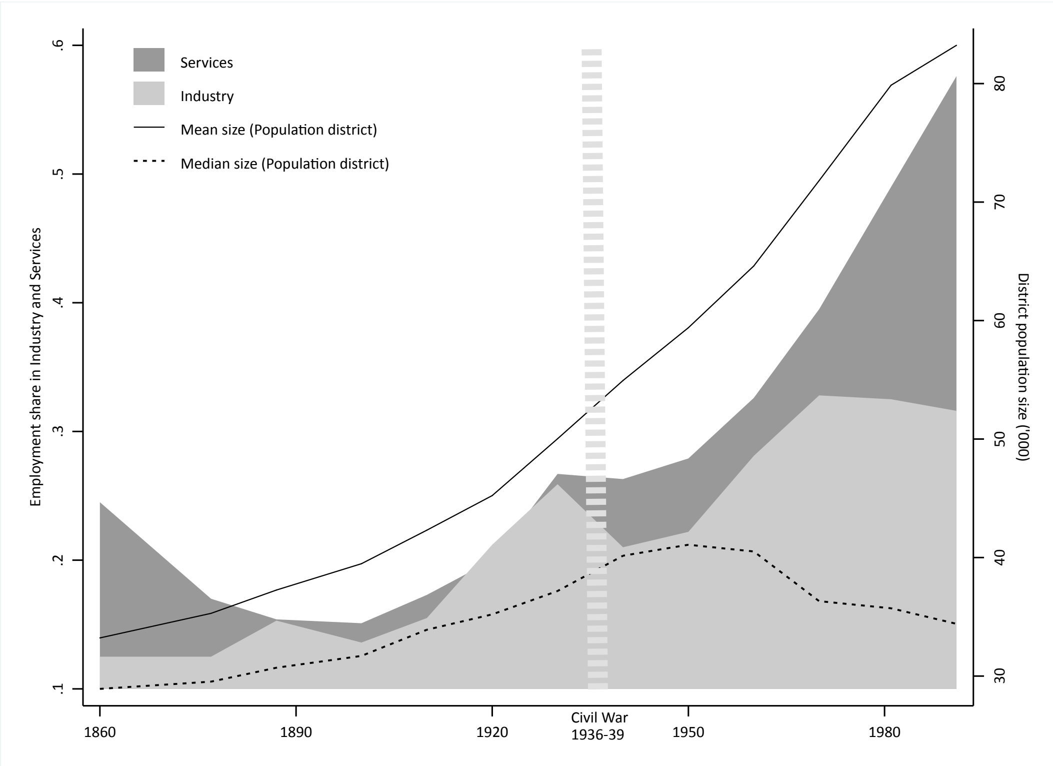
Map 1. Population in Spain by district, 1860-2000



Source: Population censuses

Figure 4 illustrates structural change, proxied with the share of employment in industry and services, and the average and median size of districts by year. In the early stages, there appears to be a steady, though timid, increase in the districts population. This is expected given that structural change was rather modest. Yet, from 1920 to 1970, the Spanish population rapidly concentrated. The mechanisation of agriculture and the spread of industrialisation, especially since the 1950s, fuelled rural-urban migration[24]. Districts, where modern industries located, grew rapidly. Backward and agrarian ones shrunk. This diverging pattern between the average and median sizes continued, but at a slower pace, in the late decades of the twentieth century, thus coinciding with the growth of the service sector and the rise of information and communication technologies (ICT).

Figure 4. Structural change and spatial concentration of population in Spain, 1860-1991



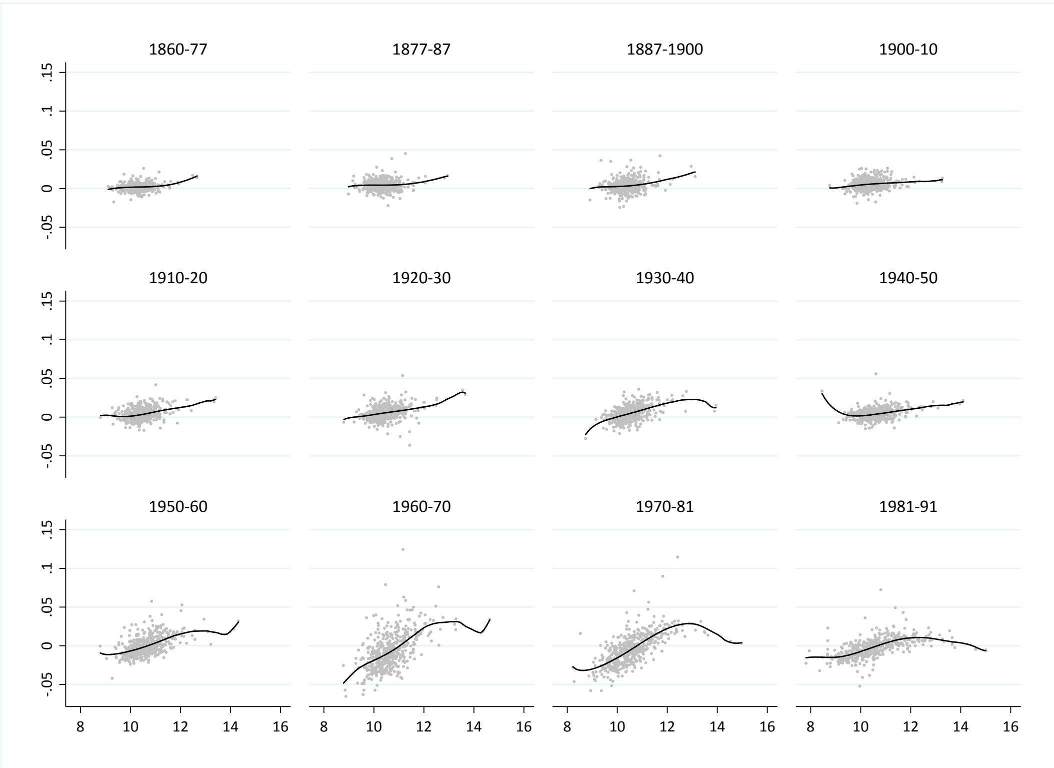
Source: Population censuses

Agglomeration economies take place in densely populated areas. Therefore, the evolution of the spatial concentration of the population can also be examined focusing on the relationship between the initial level of the district population size and subsequent population growth. Figure 5 fits kernel regressions showing the link between these two variables in each period (approximately 10-year intervals). Broadly speaking, while the second half of the 19th century appears to be characterised by quasi-orthogonal growth, the early 20th century witnessed how large districts tended to grow faster than small ones, a feature that intensified during the second half. Interestingly, the positive link between initial population and subsequent growth reverses for the largest locations from 1970 onwards,

[24] As mentioned in the previous section, emigration abroad was also important during the 1950s and 1960s. Although of smaller magnitude, international emigration flows, especially to Latin America, occurred between 1880 and 1920 (Sánchez-Alonso, 2000).

suggesting that congestion costs began to exert a significant effect during this later period. Next section examines these issues in more detail with the aim of quantifying how the effect of agglomeration economies evolved over time.

Figure 5. District population growth vs. initial population



Source: Population censuses; each period roughly corresponds to (approximately) 10-year intervals.

4. Empirical Specification

In order to examine how the size of the local economy affects subsequent growth over time, we first estimate the following model for each period using OLS:

$$\Delta y_i = \alpha + \beta\, y_i + x'_i\, \gamma + \varepsilon_i \qquad (1)$$

where Δy_i is the population growth rate of each district between two censuses ($\Delta y_i = \ln y^{t+1}_i - \ln y^t_i$) while y_i refers to the log of the population level at the beginning of each period. Given that both variables are measured in logs, the estimated parameters can be interpreted as elasticities. In addition, x'_i is a vector of control variables taking into account geographic, climatic and geological features of each district.

On the one hand, given that we attempt to isolate the effect of agglomeration economies from other determinants of population growth, we have included different geographic characteristics that capture the locational fundamentals of each district. Firstly, using the WorldClim 1 kilometre digital data, we have computed average annual temperature and average annual rainfall. Secondly, the SRTM 90-meter resolution digital elevation data allows obtaining a measure of the median altitude of each district, as well as a ruggedness index that measures the standard deviation of altitude. Thirdly, drawing on the European Environment Agency WISE Large rivers dataset, a dummy variable takes the value of 1 if a district has access to a large river. Similarly, we have also computed distance to the coast. Fourthly, to further proxy for the potential agricultural productivity, we have relied on certain soil quality parameters provided by the European Soil Database (ESDB 1-kilometer resolution): top soil available water capacity, base saturation of the top soil, volume of stones and top soil organic content. Following Combes et al. (2010), we have computed which is the most common category in each district and then assigned the corresponding dummy variables to control for that. Given the large heterogeneity of the districts' geographic size, the specification also controls for district area. Moreover, given the central location of Madrid, the country's capital, a dummy variable has also been created to account for this[25].

On the other hand, the growth of a particular district not only depends on its own economic dynamism, but also on that of competing neighbouring population, so our model incorporates the existence of important neighbouring locations using GIS techniques[26]. More specifically, we have computed for each period the population living in towns larger than 10,000 inhabitants within a certain radius from the district centroid: less than 50, 50-100, 100-250 and 250-500 kilometres respectively. Lastly, to further capture potentially unobserved characteristics, we include regional dummies[27]. Tables A1 and A2 in the Appendix explain how the variables employed have been obtained and report summary statistics.

Figure 6 visually reports the results of regressing population growth on initial population, as well as the full set of controls, for each period (see table A3 in the Appendix

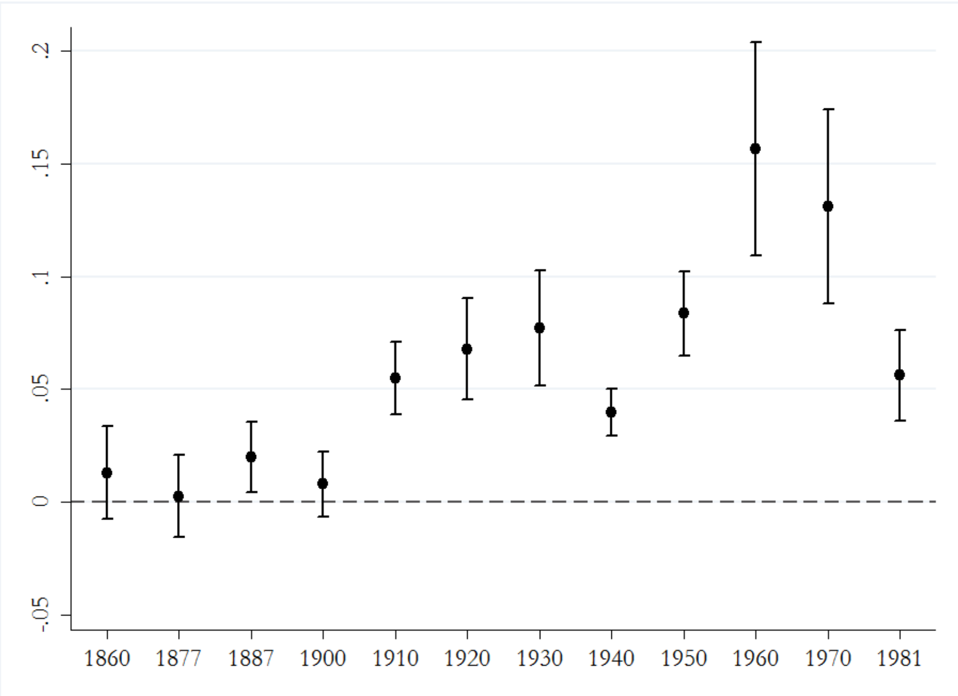
[25] In order to capture the potential capital status effect at the province level, we have also added to our controls a set of dummy variables for the districts where the province capitals are located. Although the results become less precise because initial size and capital status are highly correlated, they remain qualitatively unchanged and are available upon request.

[26] Rappaport and Sachs (2003) and Desmet and Fafchamps (2005) follow a similar approach.

[27] We followed the classification employed by Simpson (1996), which divides the country into ten macro regions: Galicia, Biscay, Castile-Leon, Upper Ebro, Lower Ebro, Levant, Centre, Extremadura, Eastern Andalusia and Western Andalusia.

for the actual results). Standard errors are clustered at the provincial level to take into account that same-province districts may share unobserved characteristics. The estimated coefficients evidence that the effect of size was negligible in the second half of the 19th century and the first decade of the 20th century[28]. Centripetal forces only started to consistently induce population concentration after 1910. The estimated coefficients kept growing during the first third of the 20th century, from 0.055 in the 1910s to 0.077 in the 1930s. The Civil War and its aftermath, however, induced a set back on the intensity of agglomeration economies which were reduced to figures around 0.040. While the 1950s witnessed how the estimated coefficients returned to the levels existent prior to the war, the 1960s experienced a major boost that situated them around 0.156. Although the 1970s still enjoyed relatively high figures (around 0.131), the effect of initial size on subsequent population growth was significantly decreased in the 1980s (0.056).

Figure 6. Evolution of agglomeration effects (OLS)



The previous model, however, may suffer from endogeneity problems. It is plausible that larger locations are the result of some local characteristics, so their growth may not be the result of agglomeration economies but of some underlying feature, such as better agricultural potential or the presence of certain amenities (administrative or transportation

[28] Although quite low, the estimated coefficient is statistically significant for the period 1887-1900. The end-of-the-century crisis, however, temporarily suppressed these incipient developments.

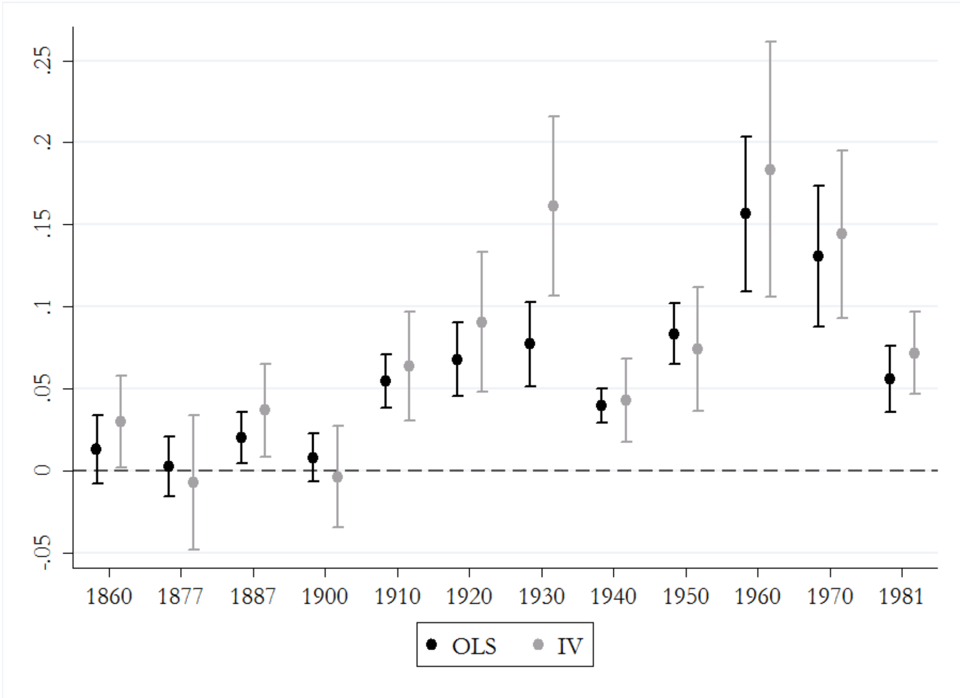
infrastructures), that promotes their future growth. Although we have included a comprehensive set of variables that attempt to control for this issue, we have further attempted to mitigate this concern by instrumenting the size of the local economy using historical urban population. In particular, we employ population living in cities larger than 5,000 individuals in 1500[29]. By doing so, we exploit the long-term persistence of the spatial distribution of population due to the inertias that local population and economic activity generate. Given the long lag employed, this instrument is plausibly exogenous because the modern sources of local productivity differ from those existing in such a distant past.

However, there might be local characteristics that affected population growth in the past and still continue to influence it in more recent times: suitable agro-climatic conditions, the presence of a large river or another geographical feature that increases the location's potential such as the centrality of the location in the country or having access to the sea (Combes and Gobillon, 2015, 287). This is especially important because we start measuring agglomeration economies in the mid-19th century when structural change remained limited and agriculture was still a significant source of local wealth. Crucially, as discussed above, our model directly controls for other geographical, climate and geological variables that may have still influenced each district potential productivity during the period under study.

As the first stage evidences, our instrument is highly correlated with the instrumented variables (see Table A.4 in the Appendix)[30]. Fig. 7 compares the IV results to those obtained using OLS. Broadly speaking, both sets of estimated coefficients depict similar trends, the IV results tend to be higher, especially between 1910 and the Civil War and between 1960 and 1980, thus reinforcing the image of rapid structural change and the increasing spatial concentration of the population that these two periods witnessed. By the 1930s, when initial size was twice as large, population growth was around 16.1 per cent higher. The magnitude of agglomeration economies was even higher during the 1960s. In comparison, these forces paled during both the aftermath of the Civil War and the 1980s.

[29] The historical population is taken from Bosker et al. (2013). Using the population figures for the years 1600, 1700 or 1800 does not alter the results reported here.
[30] Also, the Sanderson-Windmeijer (SW) test always rejects that the endogenous regressor is unidentified.

Figure 7. Evolution of agglomeration effects (OLS vs IV)



5. Discussion

Our results are consistent with Krugman's (1991) model: agglomeration economies strengthen as transport costs are reduced over time. The structural transformation away from agriculture also helps explaining the increasing effect of agglomeration over time, especially for intermediate densities. The increasing relationship between initial size and subsequent population growth at intermediate U.S. locations was stronger from 1880 to 1960 than during the late 20th century (Michaels et al., 2012, 537). This pattern is arguably related to reallocation away from agriculture, a process that was significantly more intense during the former period. Michaels et al. (2012, 536) argue that, in more populated areas, where non-agricultural activities already dominate, further population growth is not necessarily correlated with initial population density. Taking into account the obvious differences between the history of the U.S. and Spain, the evolution of the effect of initial size on subsequent population growth is also similar. In Spain, the relationship between these two variables when the largest districts are excluded increased significantly between 1910 and 1970, although this trend was abruptly interrupted by the Civil War and the autarchic period. Using provinces as unit of analysis, Ayuda et al. (2010) also find that increasing returns only started to play a role in the geographical concentration of the Spanish population from 1900 onwards as the share of increasing-returns sectors in the Spanish economy grew.

The relatively high coefficients found around 1960/70s are consistent with the comparatively higher incidence of agglomeration economies found in developing countries nowadays[31]. The intensity of the link between initial population and subsequent population growth debilitated in the 1970s, a process that continued during the 1980s. Desmet and Fafschamps (2005, 262) argue that these recent developments can be attached to de-industrialisation. While services, mostly a non-tradable activity, had traditionally been spread out, declining transportation and communication costs have recently facilitated services to concentrate. In contrast, the same processes have weakened the benefits from agglomeration in manufacturing, thus promoting their geographical dispersion[32]. The depression that followed the 1970s crisis indeed meant that the service and construction sectors replaced industrial production as the main engines of economic growth in Spain (Bentolilla and Blanchard, 1990; Bentolilla and Dolado, 1991). Similarly, as land prices increase, more land intensive activities, such as manufacturing, are replaced by less land intensive activities, such as services (Desmet and Fafschamps, 2005, 262). Consistent with our results, Paluzie et al. (2004) show that the spatial distribution of manufacturing in Spain presented a bell-shaped evolution, with an initial phase characterised by a significant increase in industrial concentration followed by a trend reversal since the 1970s when a growing dispersion of industry is observed[33]. In this regard, Paluzie et al. (2009b) show that the geography and intensity of internal migrations mirrored the patterns of industrial concentration.

Increasing congestion costs, arising from rising housing prices and the higher weight of amenities and other aspects related to the quality of life (Bover and Velilla, 1999; Bentolila, 2001), may also explain why the coefficients on initial population got smaller during the 70s and the 80s. Focusing on the evolution of the largest cities, Lanaspa et al. (2003) find that while differences in city size got amplified between 1900 and 1970, small and intermediate cities grew faster than large ones from that moment onwards[34]. Due to congestion and pollution costs, agglomeration economies are subject to diminishing returns, so we now explore how agglomeration economies depend on district size. The size of the coefficients we have estimated reflect the total net impact of the concentration of

[31] The coefficients are 0.10-0.12 and 0.09-0.12 for China and India, respectively (Combes et al., 2013; Chauvin et al., 2014).

[32] In addition, "the splitting up of the production process into different stages has allowed manufacturing firms to relocate certain activities to less dense areas" (262). On these issues, see also Desmet and Rossi-Hansberg (2009).

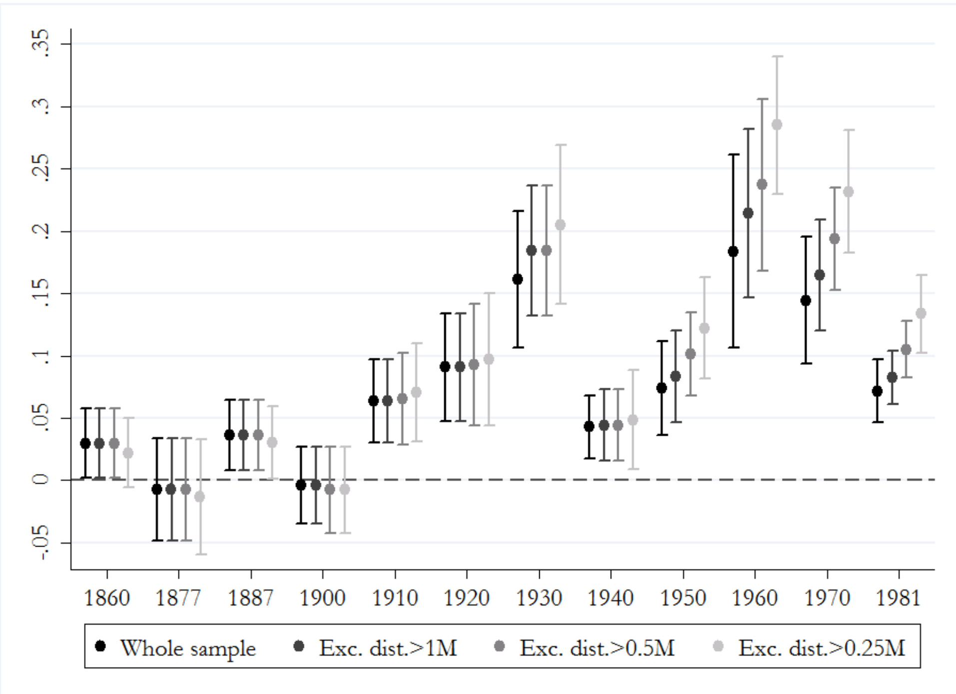
[33] Similarly, Martinez-Galarraga et al. (2008) find that, although employment density is related with inter-regional differences in industrial labour productivity during the early stages of industrialisation, this association becomes less important over time and actually disappears in recent decades (1985-1999).

[34] See also González-Val et al. (2014) where the authors test the validity of Gibrat's Law for the complete set of Spanish municipalities during the 20th century. Additionally, García-López et al. (2015) show that the extensive construction of highways after the late 1960s, and especially from the 1980s onwards, fostered suburbanization in Spain.

economic activity. Given that congestion costs increase as population density grows, it is interesting to examine their importance by estimating the same model but excluding the largest districts.

Figure 8 compares the 2SLS results of estimating equation 1 using the whole sample to those of replicating the same exercise but sequentially excluding the largest locations: those with a population above one million, 500,000 and 250,000 inhabitants respectively (see also Table A5 in the Appendix). This exercise evidences that our results are not just driven by the upper tail of the distribution. The estimated coefficients of initial size on subsequent population growth actually remain qualitatively unchanged up to the 1930s, when they slightly increase, once the largest districts are excluded from the sample. Furthermore, once the disruptions caused by the Civil War and its aftermath waned, this occurs again, though the difference between the coefficients is only statistically significant in the 1980s. These results could be driven by the dynamism of middle-sized districts and the presence of congestion costs in the largest districts, especially during the more recent decades. Yet, it is worth noting that, despite congestion costs, the gains from agglomeration prevailed.

Figure 8. Evolution of agglomeration effects. IVs. Excluding largest districts



6. Conclusion

Agglomeration economies play a fundamental role in the location of economic activity. However, the impact of these forces is likely to have varied over time as economies evolve. We have shown how economic density only started to significantly influence the spatial concentration of the Spanish population in the early 20th century, a process closely related to the structural transformation of the economy. The effect of initial population on subsequent population growth increased between 1910 and 1970, although this trend was briefly interrupted by the Spanish Civil War and the autarkic period that followed. The intensity of this relationship, however, receded in the 1970s, and especially during the 1980s, as rural out-migration slowed down and several traditional industries fell apart. Furthermore, the largest locations did not benefit as much as medium-size ones from the presence of agglomeration economies after the 1960s. This raises several questions, which are not discussed in the paper, but it points to the fundamental trade-off between increasing returns and congestion costs.

References cited:

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): 25-50

Beeson, P.E., Dejong, D.N., and Troesken, W. (2001), 'Population growth in US counties, 1840-1990', *Regional Science and Urban Economics* 31: 669-699

Bentolila, S. (2001), 'Las migraciones interiores en España', *FEDEA Working Papers* 2001-07

Bentolila, S., and Blanchard, O. (1990), 'Spanish unemployment', *Economic Policy* 10: 233-281

Bentolila, S., and Dolado, J.J. (1991), 'Mismatch and internal migration in Spain, 1962-1986', in F. Padoa-Schioppa (ed.), *Mismatch and labour mobility*, (Cambridge: Cambridge University Press): 231-288

Bosker, M., Buringh, E., and Van Zanden, J.L. (2013), 'From Baghdad to London: Unravelling urban development in Europe and the Arab world, 800-1800', *Review of Economics and Statistics* 95 (4): 1418-1437

Bosker, M., and Buringh, E. (2015), 'City seeds: Geography and the origins of the European city system', *Journal of Urban Economics* (In Press).

Bover, O., and Velilla, P. (1999), 'Migrations in Spain: Historical background and current trends', *Banco de España Working Paper* 9909

Briant, A., Combes, P.P., and Lafourcade, M. (2010), 'Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations?', *Journal of Urban Economics* 67 (3): 287-302

Chauvin, J.P., Glaeser, E., and Tobio, K. (2014), 'Urban economics in the US and India', Harvard University

Ciccone, A., and Hall, R.E. (1996), 'Productivity and the density of economic activity', *American Economic Review* 86 (1): 54-70

Clark, J.S., and Stabler, J.C. (1991), 'Gibrat's law and the growth of Canadian cities', *Urban Studies* 28 (4): 635-639

Combes, P.P., Duranton, G., Gobillon, L., and Roux, S. (2010), 'Estimating agglomeration economies with history, geology, and worker effects', in E.L. Glaeser (ed.), *Agglomeration economies* (Chicago: The University of Chicago Press): 15-65

Combes, P.P., and Gobillon, L. (2015), 'The empirics of agglomeration economies', in G. Duranton, J.V. Henderson and W. Strange (eds.), *Handbook of urban and regional economics*, vol. 5 (North-Holland: Elsevier): 247-348

Combes, P.P., De Murger, S., and Li, S. (2013), 'Urbanisation and migration externalities in China', *Discussion Paper 9352, Centre for Economic Policy Research*

Combes, P.P, Mayer, T., and Thisse, J.F. (2008), *Economic geography. The integration of regions and nations* (Princeton: Princeton University Press)

Combes, P.P., Duranton, G., and Gobillon, L. (2012), 'The costs of agglomeration: Land prices in French cities', Discussion Paper 9240, *Centre for Economic Policy Research*

Cronon, W. (1991), *Nature's metropolis: Chicago and the great west* (New York: W.W. Norton & Company)

Davis, D.R., and Weinstein, D.E. (2002), 'Bones, bombs, and break points: The geography of economic activity', *American Economic Review* 92 (5): 1269-1289

Desmet, K., and Fafchamps, M. (2005), 'Changes in the spatial concentration of employment across US counties: A sectoral analysis 1972-2000', *Journal of Economic Geography* 5 (3): 261-284

Desmet, K., and Rappaport, J. (2015), 'The settlement of the United States, 1800-2000: The long transition towards Gibrat's law', *Journal of Urban Economics* (In Press)

Desmet, K., and Rossi-Hansberg, E. (2009), 'Spatial growth and industry age', *Journal of Economic Theory* 144: 2477–2502

Dobkins, L.H., and Ioannides, Y.Y. (2001), 'Spatial interactions between U.S. cities: 1900-1990', *Regional Science and Urban Economics* 31: 701-731

Duranton, G. (2007), 'Urban evolutions: The fast, the slow, and the still', *American Economic Review* 97: 197–221

Duranton, G., and Puga, D. (2004), 'Micro-foundations of urban agglomeration economies', in J.V. Henderson and J.F. Thisse (eds.), *Handbook of Regional and Urban Economics* (Amsterdam: North-Holland): 2063-2117

Eeckhout, J. (2004), 'Gibrat's law for (all) cities', *American Economic Review* 94 (5): 1429-1451

Fujita, M., Krugman, P., and Venables, A.J. (1999), *The spatial economy: Cities, regions and international trade* (Cambridge: MIT Press)

Fujita, M., and Thisse, J.F. (2002), *Economics of agglomeration: Cities, industrial location, and regional growth* (Cambridge: Cambridge University Press)

Gabaix, X. (1999), 'Zipf's law for cities: An explanation', *Quarterly Journal of Economics* 114 (3): 739-767

Gabaix, X., and Ioannides, Y. (2004), 'The evolution of city size distributions', in J.V. Henderson and J.F. Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4 (North-Holland: Elsevier): 2341-2378

García-Coll, A., and Stillwell, J. (1999), 'Inter-provincial migration in Spain: Temporal trends and age-specific patterns', *International Journal of Population Geography* 5: 97-115.

García-López, M.A., Holl, A., and Viladecans-Marsal, E. (2015), 'Suburbanization and highways in Spain when the Romans and the Bourbons still shape its cities', *Journal of Urban Economics* 85: 52-67

Glaeser, E.L., Scheinkman, J., and Shleifer, A. (1995), 'Economic growth in a cross-section of cities', *Journal of Monetary Economics* 36: 117-143

Glaeser, E.L. (2010), 'Introduction', in E.L. Glaeser (ed.), *Agglomeration economics* (Chicago: University of Chicago): 1-14

Goerlich, F.J., Mas, M., Azagra, J., and Chorén, P. (2006), *La localización de la población sobre el territorio. Un siglo de cambios: un estudio basado en series homogéneas (1900-2001)* (Bilbao: Fundación BBVA)

Goerlich, F.J., and Mas, M. (2008), 'Algunas pautas de localización de la población española a lo largo del siglo XX', *Investigaciones Regionales* 12: 5-33

Goerlich, F.J., and Mas, M. (2009), 'Drivers of agglomeration: Geography vs history', *The Open Urban Studies Journal* 2: 28-42

Gómez Mendoza, A., and San Román, E. (2005), 'Transportes y comunicaciones', in A. Carreras and X. Tafunell (eds.), *Estadísticas históricas de España, siglos XIX-XX* (Bilbao: Fundación BBVA): 509-572

González-Val, R., Lanaspa, L., and Sanz-Gracia, F. (2014), 'New evidence on Gibrat's Law for cities', *Urban Studies* 51 (1): 93-115

González-Val, R., Tirado, D.A., and Viladecans-Marsal, E. (2015), 'Market potential and city growth: Spain 1860–1960', *Cliometrica* (online first)

Graham, D.J. (2007), 'Variable returns to agglomeration and the effect of road traffic congestion', *Journal of Urban Economics* 62: 103–120

Henderson, J.V. (2003), 'Marshalls' scale economies', *Journal of Urban Economics* 53 (1): 1-28

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

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

Jacobs, J. (1969), *The economy of cities* (New York: Vintage)

Kim, S. (1995), 'Expansion of markets and the geographic distribution of economic activities: The trends in US regional manufacturing structure, 1860-1987', *Quarterly Journal of Economics* 110 (4): 881-908.

Krugman, P. (1991), 'Increasing returns and economic geography', *Journal of Political Economy* 99 (3): 483-499

Krugman, P., and Venables, A.J. (1995), 'Globalization and the inequality of nations', *Quarterly Journal of Economics* 110 (4): 857-880

Lanaspa, L., Pueyo, F., and Sanz, F. (2003), 'The evolution of Spanish urban structure during the twentieth century', *Urban Studies* 40 (3): 567-580

Marshall, A. (1890), *Principles of economics* (London: Macmillan)

Martinez-Galarraga, J. (2012), 'The determinants of industrial location in Spain, 1856-1929', *Explorations in Economics History* 49: 255-275

Martinez-Galarraga, J., Paluzie, E., Pons, J., and Tirado, D.A. (2008), 'Agglomeration and labour productivity in Spain over the long term', *Cliometrica* 2 (3): 195-212

Melo, P.C., Graham, D.J., and Nolan, R.B. (2009), 'A meta-analysis of estimates or urban agglomeration economies', *Regional Science and Urban Economics* 39: 332-342

Michaels, G., Rauch, F., and Redding, S.J. (2012), 'Urbanization and structural transformation', *Quarterly Journal of Economics* 127 (2): 535-586

Nadal, J. (1987). 'La industria fabril Española en 1900. Una aproximación', in J. Nadal, A. Carreras, and C. Sudrià (eds), *La economía española en el siglo XX: Una perspectiva histórica* (Barcelona: Ariel): 23-61

Paluzie, E., Pons, J., and Tirado, D.A. (2004), 'The geographical concentration of industry across Spanish regions, 1856-1995', *Jahrbuch für Regionalwissenschaft* 24: 143-160.

Paluzie, E., Pons, J., and Tirado, D.A. (2009a), 'A test of the market potential equation in Spain', *Applied Economics* 41 (12): 1487-1493

Paluzie, E., Pons, J., Silvestre, J., and Tirado, D.A. (2009b), 'Migration and market potential in Spain over the twentieth century: A test of the new economic geography', *Spanish Economic Review* 11: 243-265

Partridge, M.D., Rickman, D.S., Ali, K., and Olfert, M.R. (2008), 'Lost in space: Population growth in the American hinterlands and small cities', *Journal of Economic Geography* 8: 727-757

Peña, D., and Sánchez-Albornoz, N. (1984), 'Wheat prices in Spain, 1857-1890: An application of the Box-Jenkins methodology', *Journal of European Economic History* 2: 353-373

Pons, J., and Tirado, D.A. (2008), 'Los determinantes de la desigualdad económica regional en España', *Información Comercial Española, ICE: Revista de Economía* 842: 195-216

Prados de la Escosura, L. (2008), 'Inequality, poverty and the Kuznets curve in Spain, 1850-2000', *European Review of Economic History* 12 (3): 287-324

Prados de la Escosura, L., and Rosés, J.R. (2009), 'The sources of long-run growth in Spain, 1850-2000', *The Journal of Economic History* 69 (4): 1063-1091

Prados de la Escosura, L., Rosés, J.R., and Sanz-Villarroya, I. (2012), 'Economic reforms and growth in Franco's Spain', *Revista de Historia Económica / Journal of Iberian and Latin American Economic History* 30 (1): 45-89

Puga, D. (1999), 'The rise and fall of regional inequalities', *European Economic Review* 43: 303-334

Rappaport, J., and Sachs, J.D. (2003), 'The United States as a coastal nation', *Journal of Economic Growth* 8 (1): 5-46

Ringrose, D.R. (1998), *Spain and Europe, 1700-1900: The myth of the failure* (Madrid: Instituto Universitario Ortega y Gasset)

Ródenas, C. (1994), 'Migraciones interregionales en España, 1960-1989', *Revista de Economía Aplicada* 2 (4): 5-36

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

Rosés, J.R., and Sánchez-Alonso, B. (2004), 'Regional wage convergence in Spain 1850-1930', *Explorations in Economic History* 41: 404-425

Rossi-Hansberg, E., and Wright, M.L. (2007), 'Urban structure and growth', *Review of Economic Studies* 74: 597–624

Sánchez-Alonso, B. (1995), *Las causas de la emigración española, 1880-1930* (Madrid: Alianza Editorial)

Sánchez-Alonso, B. (2000), 'Those who left and those who stayed behind: Explaining emigration from the regions of Spain, 1880-1914', *Journal of Economic History* 60 (3): 730-755

Silvestre, J. (2002), 'Las emigraciones interiores en España durante los siglos XIX y XX: Una revisión bibliográfica', *AGER-Revista de Estudios sobre Despoblación y Desarrollo Rural* 2: 227-248

Silvestre, J. (2005), 'Internal migrations in Spain, 1877-1930', *European Review of Economic History* 9 (2): 233-265

Silvestre, J. (2007), 'Temporary internal migrations in Spain, 1860-1930', *Social Science History* 31: 539-574

Simpson, J. (1996), *Spanish agriculture: The long siesta: 1765-1965* (Cambridge: Cambridge University Press)

Tedde, P. (1994), 'Cambio institucional y cambio económico en la España del siglo XX', *Revista de Historia Económica* 3: 525-538

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

Tirado, D.A., and Martinez-Galarraga, J. (2008), 'Una nova estimació retrospectiva del VAB regional industrial. Espanya (1860-1930)', *Documents de Treball de la Facultat de Ciències Econòmiques i Empresarials* E08/192

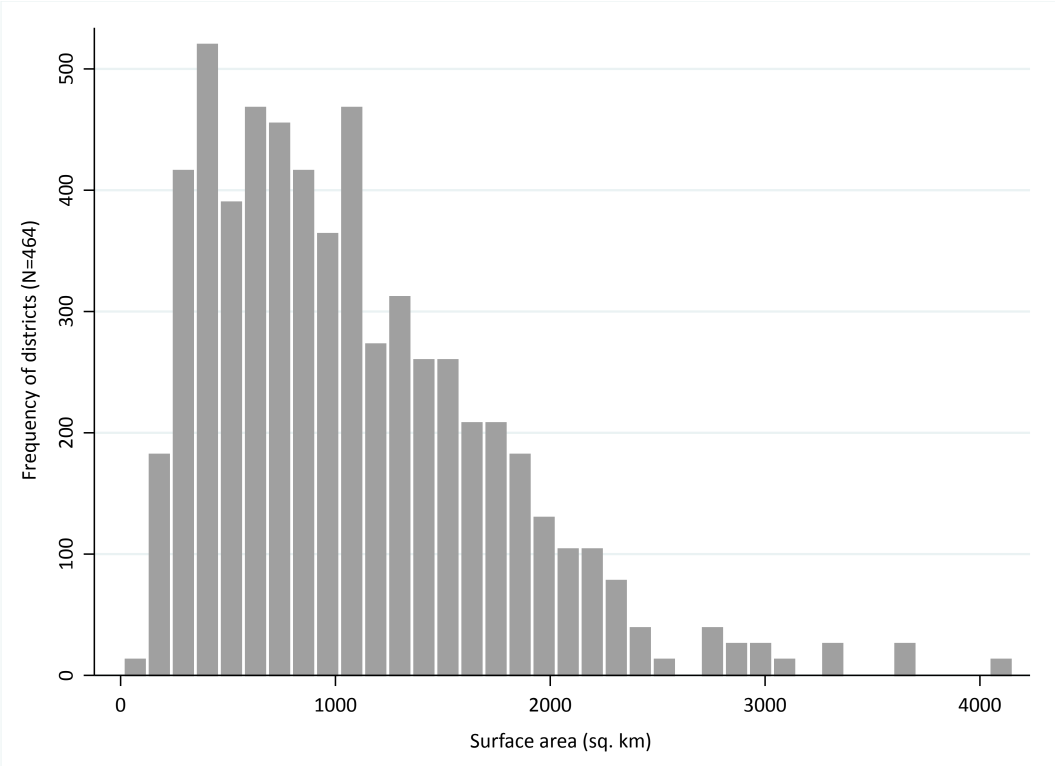
Tirado, D.A., Pons, J., Paluzie, E., and Martinez-Galarraga, J. (2013), 'Trade policy and wage gradients: Evidence from a protectionist turn', *Cliometrica* 7: 295-318

Venables, A.J. (1996), 'Equilibrium locations of vertically linked industries', *International Economic Review* 37: 341-359

Wais, F. (1987), *Historia de los ferrocarriles españoles* (Madrid: Fundación de los Ferrocarriles Españoles)

Appendix

Figure A.1. Surface area distribution of districts



Source: European Soil Data base (ESDB). District area computed using ArcGIS.

Table A.1. Description of the variables employed

Population	District population. Municipalities have been grouped into *Partidos Judiciales* using the 1860 boundaries. Data taken from the *Population Censuses* (available at http://www.ine.es/intercensal/).
Population living in neighbouring cities	Sum of population living in cities larger than 10,000 inhabitants within 50, 50-100, 100-250 and 250-500 kilometers from the district geographical centre.
Temperature	Annual average temperature taken from WorldClim 1 kilometre digital data (Hijmans et al., 2005) which can be found at http://www.worldclim.org/. The climate information refers to the average during the period 1950-2000.
Rainfall	Average annual rainfall. As with temperature, this is taken from WorldClim 1 kilometre digital data (Hijmans et al. 2005) which averages pluviosity over the period 1950-2000.
Altitude	Median altitude in each district using the SRTM 90-meter resolution digital elevation data (http://srtm.csi.cgiar.org).
Ruggedness	Standard deviation of altitude.
Large rivers	Dummy variable that takes the value of 1 if a district has access to a large river, namely Rivers Ebro, Tagus, Duero, Guadiana and Guadalquivir. Georeferenced data can be found at http://www.eea.europa.eu/data-and-maps/data/wise-large-rivers-and-large-lakes
Capital city	Dummy variable that takes the value of 1 for the Madrid district, which hosts the country's capital.
Distance to coast	Distance from the district centroid to the nearest coastline.
Soil quality	A number of dummy variables have been generated relying on the following features provided by the European Soil Database (ESDB 1-kilometer resolution): Top soil available water capacity, Base saturation of the top soil, Topsoil organic content, Volume of stones and Distance to rock. After computing which is the most common category in each district, dummy variables have been assigned to the corresponding categories. Raw raster data can be found at http://esdac.jrc.ec.europa.eu/resource-type/european-soil-database-soil-properties
District area	District area (km2).

Note: The geographical data have been computed using ArcGIS.

Table A2. Summary statistics

	Obs.	Mean	St. dev.	Min.	Max.
Population (in thousands)	6,032	52.9	118.8	2.0	3,291.7
Annual population growth (%)	5,568	0.18	1.22	-6.34	13.24
Pre-modern population, 1500 (in thousands)	6,032	1.8	6.6	0	70
Urban pop. within 50 km radius (000)	6,032	192	387	0	4,553
Urban pop. within 50-100 km radius (000)	6,032	453	613	0	5,125
Urban pop. within 100-250 km radius (000)	6,032	1,981	1,814	51	10,925
Urban pop. within 250-500 km radius (000)	6,032	4,936	3,814	239	18,796
Distance to coast (kms)	6,032	103.2	89.8	0.3	356.2
Temperature (°C)	6,032	13.7	2.5	4.1	18.2
Rainfall (mm)	6,032	642.1	263.2	269	1,517
Altitude (m)	6,032	571.1	342.9	2	1,915
Ruggedness	6,032	178.4	115.0	7	707
Area (km2)	6,032	1,075.2	664.5	12.2	4,154.8
Capital	6,032	0.00216	0.0464	0	1
Large rivers	6,032	0.235	0.424	0	1
Top soil average water capacity, d2	6,032	0.907	0.290	0	1
--, d3	6,032	0.0216	0.145	0	1
Base saturation of the top soil, d2	6,032	0.0194	0.138	0	1
--, d3	6,032	0.741	0.438	0	1
Top soil organic content, d2	6,032	0.373	0.484	0	1
--, d3	6,032	0.218	0.413	0	1
--, d4	6,032	0	0	0	0
Volume of stones, d2	6,032	0.0603	0.238	0	1
--, d3	6,032	0.213	0.410	0	1
--, d4	6,032	0.205	0.404	0	1
Distance to rock, d2	6,032	0.224	0.417	0	1
--, d3	6,032	0.330	0.470	0	1
--, d4	6,032	0.0280	0.165	0	1

Source: See table A1.

Table A3. Evolution of agglomeration economies. OLS results

	Dependent variable: Population growth (ln yt+1 - ln yt)											
	1860	1877	1887	1900	1910	1920	1930	1940	1950	1960	1970	1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Initial pop.	0.013 (0.010)	0.003 (0.009)	0.020* (0.008)	0.008 (0.007)	0.055** (0.008)	0.068** (0.011)	0.077** (0.013)	0.040** (0.005)	0.083** (0.009)	0.156** (0.023)	0.131** (0.021)	0.056** (0.010)
Obs.	464	464	464	464	464	464	464	464	464	464	464	464
R2	0.359	0.309	0.309	0.343	0.440	0.355	0.486	0.256	0.499	0.636	0.631	0.510

Note: Robust standard errors clustered at the provincial level in parentheses; ** p<0.01, * p<0.05. All specifications include the full set of controls discussed in the text. Both the dependent and the independent variables are expressed in natural logs, so the coefficients can be interpreted as elasticities.

Table A4. Evolution of agglomeration economies. IV results

Panel A: Second Stage	Dependent variable: Population growth (ln yt+1 - ln yt)											
	1860	1877	1887	1900	1910	1920	1930	1940	1950	1960	1970	1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Initial pop.	0.030*	-0.007	0.037*	-0.004	0.064**	0.091**	0.161**	0.043**	0.074**	0.184**	0.144**	0.072**
	(0.014)	(0.021)	(0.014)	(0.016)	(0.017)	(0.022)	(0.028)	(0.013)	(0.019)	(0.040)	(0.026)	(0.013)
R2	0.349	0.306	0.299	0.339	0.438	0.344	0.342	0.256	0.497	0.633	0.629	0.502
S-W F test	66.94	65.49	60.22	57.05	53.48	53.08	55.82	70.49	72.39	85.34	112.2	129.9
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Panel B: First Stage	Dependent variable: Initial population (ln)											
	1860	1877	1887	1900	1910	1920	1930	1940	1950	1960	1970	1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Pre-modern pop., 1500	0.192**	0.199**	0.195**	0.208**	0.207**	0.221**	0.240**	0.280**	0.292**	0.310**	0.369**	0.428**
	(0.023)	(0.025)	(0.025)	(0.028)	(0.028)	(0.030)	(0.032)	(0.033)	(0.034)	(0.034)	(0.035)	(0.038)
Obs.	464	464	464	464	464	464	464	464	464	464	464	464

Note: Robust standard errors clustered at the provincial level in parentheses; ** p<0.01, * p<0.05. All specifications include the full set of controls discussed in the text. Both the dependent and the independent variables are expressed in natural logs, so the coefficients can be interpreted as elasticities.

Table A5. Evolution of agglomeration economies. IV results. Different samples

Panel A: Whole Sample

	Dependent variable: Population growth (ln yt+1 - ln yt)											
	1860	1877	1887	1900	1910	1920	1930	1940	1950	1960	1970	1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Initial pop.	0.030**	-0.007	0.037**	-0.004	0.064***	0.091***	0.161***	0.043***	0.074***	0.184***	0.144***	0.072***
	(0.014)	(0.021)	(0.014)	(0.016)	(0.017)	(0.022)	(0.028)	(0.013)	(0.019)	(0.040)	(0.026)	(0.013)
Observations	464	464	464	464	464	464	464	464	464	464	464	464
R-squared	0.349	0.306	0.299	0.339	0.438	0.344	0.342	0.256	0.497	0.633	0.629	0.502
S-W F test	66.94	65.49	60.22	57.05	53.48	53.08	55.82	70.49	72.39	85.34	112.2	129.9
p-value	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Panel B: Excluding districts with population over 1,000,000 inhabitants

	Dependent variable: Population growth (ln yt+1 - ln yt)											
	1860	1877	1887	1900	1910	1920	1930	1940	1950	1960	1970	1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Initial pop.	0.030**	-0.007	0.037**	-0.004	0.064***	0.091***	0.184***	0.044***	0.084***	0.214***	0.164***	0.083***
	(0.014)	(0.021)	(0.014)	(0.016)	(0.017)	(0.022)	(0.027)	(0.015)	(0.019)	(0.034)	(0.023)	(0.011)
Observations	464	464	464	464	464	464	462	462	462	462	462	462
R-squared	0.349	0.306	0.299	0.339	0.437	0.344	0.326	0.240	0.499	0.652	0.653	0.516
S-W F test	67.14	65.64	60.28	57.14	53.54	53.16	67.50	82.10	86.02	100.2	117.5	123.6
p-value	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Panel C: Excluding districts with population over 500,000 inhabitants

	Dependent variable: Population growth (ln yt+1 - ln yt)											
	1860	1877	1887	1900	1910	1920	1930	1940	1950	1960	1970	1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Initial pop.	0.030**	-0.007	0.037**	-0.008	0.065***	0.093***	0.184***	0.044***	0.101***	0.237***	0.194***	0.105***
	(0.014)	(0.021)	(0.014)	(0.018)	(0.019)	(0.025)	(0.027)	(0.015)	(0.017)	(0.035)	(0.021)	(0.012)
Observations	464	464	464	462	462	462	462	462	461	459	457	454
R-squared	0.349	0.306	0.299	0.337	0.422	0.316	0.326	0.240	0.522	0.647	0.658	0.518
S-W F test	67.14	65.64	60.28	79.67	73.53	67.87	67.50	82.10	86.42	73.25	95.15	78.69
p-value	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Panel D: Excluding districts with population over 250,000 inhabitants

	Dependent variable: Population growth (ln yt+1 - ln yt)											
	1860	1877	1887	1900	1910	1920	1930	1940	1950	1960	1970	1981
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Initial pop.	0.022	-0.013	0.030**	-0.008	0.070***	0.097***	0.205***	0.049**	0.122***	0.285***	0.232***	0.133***
	(0.014)	(0.023)	(0.015)	(0.018)	(0.020)	(0.027)	(0.032)	(0.020)	(0.021)	(0.028)	(0.025)	(0.016)
Observations	462	462	462	462	461	459	459	457	456	454	450	441
R-squared	0.347	0.301	0.295	0.337	0.421	0.299	0.269	0.222	0.509	0.612	0.630	0.470
S-W F test	77.83	82	69.08	79.67	79.53	47.80	47.30	57.18	57.49	63.81	48.70	42.69
p-value	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000

Robust standard errors clustered at the provincial level in parentheses; *** p<0.01, ** p<0.05, * p<0.1; All specifications include the full set of controls discussed in the text.