

WHY DID SPANISH REGIONS NOT CONVERGE BEFORE THE CIVIL WAR? AGGLOMERATION ECONOMIES AND (REGIONAL) GROWTH REVISITED*

ALFONSO DÍEZ-MINGUELA
Universitat de València^a

JULIO MARTÍNEZ-GALARRAGA
Universitat de València^a

DANIEL A. TIRADO-FABREGAT
Universitat de València^a

ABSTRACT

In this paper we explore the relationship between the presence of agglomeration economies and regional economic growth in Spain during the period 1870-1930. The study allows us to revisit the existence of a trade-off between economic growth and territorial cohesion, and also to examine whether the existence of agglomeration economies could explain the upswing in regional income inequality during the early stages of development. In doing so, we present alternative indicators for agglomeration economies and estimate conditional growth regressions at province (NUTS3) level. In line with new economic geography models, agglomeration economies in a context of market integration widened regional inequality in the second half of the 19th century and hindered its reduction during the early decades of the 20th.

Keywords: agglomeration, economic growth, economic history, Spain

JEL Code: N93, O10, O40, R10

* Received 11 November 2014. Accepted 23 July 2015. Authors thank three anonymous referees and participants at Dublin FRESH meeting 2014 and 7th Iberometrics Conference for their comments. Authors acknowledge financial support from the MINECO project ECO2012-39169-CO3-02.

^a Área de Historia e Instituciones Económicas, Departamento de Análisis Económico, Facultad de Economía, Universitat de València, Avda. Tarongers, s/n. 46022 Valencia. alfonso.diez@uv.es, julio.martinez-galarraga@uv.es, daniel.tirado@uv.es

RESUMEN

En este artículo se analiza la existencia de una relación entre la presencia de economías de aglomeración y el crecimiento económico regional en España durante el periodo 1870-1930. El estudio permite revisar la existencia de un *trade-off* entre crecimiento económico y cohesión territorial y, además, examinar si las economías de aglomeración fueron un elemento clave a la hora de explicar el incremento de la desigualdad económica regional en España a lo largo de las primeras fases del desarrollo. Para ello, se presentan diferentes indicadores de aglomeración a nivel provincial (NUTS3) que posteriormente se incluyen en la estimación de regresiones de crecimiento condicionadas. En la línea de los modelos de Nueva Geografía Económica (NEG), sugerimos que la presencia de economías de aglomeración en un contexto de integración de mercado favoreció la aparición de una causación acumulativa que amplió la desigualdad regional en la segunda mitad del siglo XIX y dificultó su reducción durante las primeras décadas del siglo XX.

Palabras clave: aglomeración, crecimiento económico, historia económica, España

1. INTRODUCTION

In this article we explore the existence of a relationship between the presence of agglomeration economies and regional economic growth in Spain during the period 1870-1930. Studying the existence of complementarity between the spatial concentration of economic activity and growth allows us to revisit one of the long-standing debates in economics, that of the existence of a trade-off between equity and efficiency or, to put it in dynamic spatial terms, between territorial cohesion and growth. This has important implications for debates on economic policy today. Also, from the standpoint of economic history, it adds to our knowledge of the determining factors behind the increase in regional economic inequality in Spain during the early stages of the country's development process.

A long tradition in the literature on economic development (Myrdal 1957; Hirschman 1958) has tended to highlight that, from a spatial perspective, economic growth has been characterised by the appearance of agglomerations in production (the formation of clusters, high-density areas and development hubs) or population (the emergence of large urban areas). Studies in economic history have also shown that the earliest stages of economic development processes have a marked local or regional

component (Pollard 1981) and that there is a connection between the start of economic growth processes and the emergence of large-scale inequalities in the distribution of economic activity or population across the territory (Williamson 1965). Contributions on the subject from both fields have pointed to the possible existence of a causal relationship between the two elements. However, economic historians and development economists have based their arguments on the accumulation of new evidence without making it clear what theoretical framework they have used to give structure to their assessments.

In this respect, new developments originating from growth theory and economic geography since the 1990s have supplied an economic foundation for the existence of a positive relationship between the agglomeration of production and growth. Broadly speaking, they have provided arguments in support of the idea that the spatial proximity of producers and/or consumers favours growth insofar as it tends to strengthen it by allowing agglomeration economies to be used in the generation of knowledge through investment in R&D, in the training of human capital or in other economic activities such as industry and services (Martin and Ottaviano 1999; Fujita and Thisse 2002; Baldwin and Martin 2004). With the economic basis established, some papers have supplied the evidence needed to verify the hypothesis. Prominent in this line of research is the work done by Ciccone and Hall (1996) and Ciccone (2002) from a static perspective and Crozet and Koenig (2005), Brülhart and Sbergami (2009) and Gardiner *et al.* (2011) from a dynamic view.

On the one hand, Ciccone and Hall (1996) present a theoretical framework that identifies an agglomeration effect linking the density of economic activity with inter-regional differences in labour productivity. The authors empirically analyse the presence of this kind of relationship in the case of the U.S. counties. In an ulterior analysis Ciccone (2002) proves the presence of a relationship between economies of agglomeration arising from density and industrial labour productivity for the regions of France, Germany, Italy, Spain and the United Kingdom in the 1980s. This literature has established the links between economic density and agglomeration effects¹.

¹ The links are as follows: «Density affects productivity in several ways. If technologies have constant returns themselves, but the transportation of products from one stage of production to the next involves costs that rise with distance, then the technology for the production of all goods within a particular geographical area will have increasing returns – the ratio of output to input will rise with density. If there are externalities associated with the physical proximity of production, then density will contribute to productivity for this reason as well. A third source of density effects is the higher degree of beneficial specialization possible in areas of dense activity. Although the idea that denser economic activity had advantages from agglomeration was implicit in a large earlier literature, there does not appear to be any earlier work on which density was an explicit element of the theory, nor has there been empirical work based on measures of density» (Ciccone and Hall 1996, p. 54).

Crozet and Koenig (2005), on the other hand, find a positive relationship between inequality in the distribution of economic activity across the territory and the growth of GDP *per capita* in a sample of European NUTS3 regions for the period 1980-2000. Brülhart and Sbergami (2009) carry out a Barro-style empirical analysis of the determinants of economic growth for a large sample of countries over the period 1960-1996. In their study, alongside the explanatory variables traditionally included in this type of exercise, they introduce different indicators for the spatial agglomeration of production and population. They subsequently repeat the same kind of exercise using a sample of European Union (EU) countries for which different indicators of agglomeration, based on data for population and economic activity by branch of activity and region (NUTS2), are constructed. Their results support the existence of a positive relationship between the presence of agglomeration economies and growth in the early stages of regional development processes. However, their work also indicates that once a certain level of income *per capita* is reached (which the authors put at around \$10,000 in 2006), this relationship disappears or becomes negative. In this respect they point out that their empirical analysis provides evidence in support of the so-called «Williamson hypothesis» (Williamson 1965), according to which agglomeration will accompany growth in the early stages of economic development, characterised by the presence of high transport costs, but this relationship will become negative when these costs are reduced.

Gardiner *et al.* (2011) questioned the results obtained by Brülhart and Sbergami (2009). They explore the relationship between agglomeration and growth in EU countries and obtain inconclusive results. In particular, they note that the existence of this relationship lacks robustness when different agglomeration measures are introduced and when the size of the territorial units considered is changed (NUTS1 or NUTS2). They therefore suggest that the results are those to be expected only when using a small enough territorial scale, such as NUTS3. They also point out that the limited period for which homogeneous information is available reduces the robustness of the results. Hence, the empirical literature still presents partial and basically inconclusive evidence regarding the existence of a relationship between the agglomeration of production and economic growth. Nevertheless, it is possible to identify some of the problems that need to be considered when it comes to suggesting how to move forward in the empirical analysis of this relationship.

In these circumstances, studying the case of Spain for the period 1870-1930 implies a dual contribution to the subject. From the standpoint of the empirical debate, exploring the presence of a relationship between agglomeration and regional growth in the early stages of the economic growth process in Spain allows us to overcome many of the limitations identified in previous empirical work. First of all it involves an economy

that, because it was still in the early stages of economic development, was characterised by the presence of high transport costs (Herranz 2007; Prados de la Escosura and Rosés 2009). Also, all Spain's regions are at income levels clearly below the thresholds that would mark a change of trend in the relationship between agglomeration and growth (Rosés *et al.* 2010). Within the framework of the industrialisation processes typical of the countries of central and southern Europe, economic growth in Spain over this period was led by advances in the industrial sector, in which various papers have shown the presence of agglomeration economies (Betrán 1999; Tirado *et al.* 2002; Martínez-Galarraga *et al.* 2008). In addition, because this is a study based on regional data, it can be carried out using information involving territorial units that are sufficiently limited in size, that is, the Spanish NUTS3 areas (provinces), while minimising the role that might be played by institutional-type elements when it comes to impacting the growth dynamic of the various territories. Finally, the data set on which the study is based allows us to carry out a long-term analysis of the relationship between agglomeration and growth because we have homogeneous information for a period of 60 years that covers the first main stage of economic development in Spain.

On another level, it also contributes to the area of Spanish economic history. It is the first time that a paper has used estimates of provincial GDP for the second half of the 19th century, which complete the data set presented in Rosés *et al.* (2010)². The reconstruction of regional GDP estimates for Spain is in line with the international literature that has recently provided historical estimations for different countries. This would be the case, among others, of Britain (Crafts 2005; Geary and Stark 2015), Italy (Felice 2011), France (Combes *et al.* 2011), Portugal (Badia-Miró *et al.* 2011), Sweden (Henning *et al.* 2011; Enflo and Rosés 2015), Belgium (Buyst 2012) or Mexico (Aguilar-Retureta 2015).

As has been pointed out recently, the early stages of economic development in Spain took place alongside a strong increase in spatial inequality both as regards the location of industry across the territory (Paluzie *et al.* 2004) and levels of GDP *per capita* for the Spanish regions (Rosés *et al.* 2010). In the analysis of the causes of this increase in regional inequalities and its relation with the presence of agglomeration economies in the industrial sector, Martínez-Galarraga (2012) has shown that the marked industrial specialisation of a small number of Spanish provinces came about, among other factors, due to their market size in the presence of economies of scale that acted as an agglomeration force. Besides, Martínez-Galarraga *et al.* (2008),

² Specifically, the new data, provincial nominal GDP at factor cost, is for 1870, 1880 and 1890. The estimates were made following the methodology used by Rosés *et al.* (2010). See Appendix for a detailed description. We therefore have 10-year intervals of time for the whole period between 1860 and 1930.

following Ciccone and Hall (1996), have also shown the presence of a relationship between industrial regional employment density and labour productivity since the origins of industrialisation and modern growth in the second half of the 19th century.

Rosés *et al.* (2010) and Martínez-Galarraga *et al.* (2015) extend the analysis to the determinants of regional income inequality. Rosés *et al.* (2010) shows that the Spanish regions that specialised in industrial production were those that achieved the highest levels of income during these years. Additionally, Martínez-Galarraga *et al.* (2015) find a direct relationship between market potential and regional economic growth in the early decades of the 20th century. Bearing all this in mind, the present study allows us to analyse the hypothesis that seems to derive from these papers. Exploring whether the agglomeration of production, especially in the industrial sector, acted as an element to explain regional growth is key to understanding the upswing in regional economic inequality that characterised the Spanish economy in the period from the mid-19th century to the eve of the Civil War.

Following this introduction, the paper is divided into five sections. Section 2 presents descriptive evidence on the evolution of the concentration of economic activity and regional growth in Spain from 1860 to 1930. Section 3 presents the data set used to carry out the study, along with the descriptive evidence the data provide. Section 4 shows the empirical model on which the analysis is based. Section 5 presents and discusses the main results deriving from the statistical analysis. First of all we look at the relationship between agglomeration and provincial economic growth on an aggregate level. The analysis is then extended on the basis of a sectoral disaggregation, allowing us to explore the origins of the possible relationship between agglomeration and economic growth. Finally, section 6 summarises the main conclusions, links the results to those areas where new evidence contributes to knowledge on the subject and suggests some lines for future research. A provincial map and a detailed explanation of the methodology employed to estimate GDP values for 1870, 1880 and 1890 can be found in Appendices A and B. Appendix C shows the expansion of the railway network in Spain, 1855-1923.

2. SPATIAL INEQUALITY AND REGIONAL ECONOMIC GROWTH IN SPAIN, 1860-1930: DESCRIPTIVE EVIDENCE

The second half of the 19th century saw the beginnings of modern economic growth in Spain (Kuznets 1955). During the early stages of economic development, the economy as a whole experienced what can be considered modest rates of growth. However, this trend would change after

TABLE 1
GROWTH OF REAL GDP, POPULATION AND GDP *PER CAPITA*, 1850-1929

	GDP	Population	GDP <i>per capita</i>
1850-1883	1.8	0.4	1.4
1884-1920	1.3	0.6	0.7
1921-1929	3.8	1.0	2.8

Note: Annual average logarithmic rates.

Source: Prados de la Escosura (2008, p. 288).

TABLE 2
SPATIAL CONCENTRATION IN INDUSTRY, SPAIN 1856-1929 (NUTS3)

	1856	1893	1913	1929
Gini	0.44	0.60	0.68	0.78
Hirschmann-Herfindal	0.06	0.13	0.15	0.24

Source: Paluzie *et al.* (2004).

the First World War when both GDP and GDP *per capita* registered growth rates substantially higher than in previous periods (Table 1).

The beginning of modern economic growth is associated with structural change and, more specifically, industrialisation. In this respect, various authors including Pollard (1981) emphasise the regional nature of the industrialisation processes, given that industry throughout history has tended to develop in particular regions or specific locations within countries. Spain is a prime example of this historical evolution. Industry showed a higher degree of development mainly in the peripheral regions of Catalonia and the Basque Country. In the former, industrialisation, initially based on textiles, was already under way in the mid-19th century and had roots reaching back to the final decades of the 18th century. In the latter, where the process was driven by iron, steel and mining, there was an unprecedented boom in the last quarter of the 19th century. The arrival of industrialisation in a limited number of regions during the second half of the 19th century (Nadal 1987) and deindustrialisation in others, mainly located in the interior of the Iberian Peninsula (Sánchez-Albornoz 1987), resulted in an increase in the spatial concentration of industry (Paluzie *et al.* 2004). This trend towards higher concentration is characteristic of the early stages of economic development in Spain and would continue until the eve of the Civil War, as can be seen in Table 2³.

³ In the long term, the concentration of manufacturing in Spain shows a bell-shaped evolution reaching a peak in the 1970s (Paluzie *et al.* 2004). A similar evolution in an inverted U-shape has

So what elements would explain this increase in the spatial concentration of industry in Spain between the mid-19th century and the 1930s? Rosés (2003), following Davis and Weinstein (1999, 2003), argued that new modern manufacturing industries in the mid-19th century tended to be concentrated in regions in which home-market effects were greater. Tirado *et al.* (2002), in line with Kim (1995), identified economies of scale and market size as the forces behind Spain's industrial geography in the mid-19th century. By the end of the century, the explanatory power of these new economic geography (NEG) effects had increased in parallel with advances in the economic integration process. More recently, adopting the approach developed by Midelfart-Knarvik *et al.* (2002), Martínez-Galarraga (2012) confirmed and extended the previous findings of Tirado *et al.* (2002). As the domestic market became integrated and industrialisation progressed during the second half of the 19th century, NEG forces became the main determinant of Spain's industrial map. In particular, although comparative advantage factors were a feature of the Spanish case, the scale effects suggested by Krugman (1991), captured by the interaction between economies of scale and market potential, played a decisive role: up to the 1930s industries with increasing returns tended to be concentrated in provinces with better access to demand.

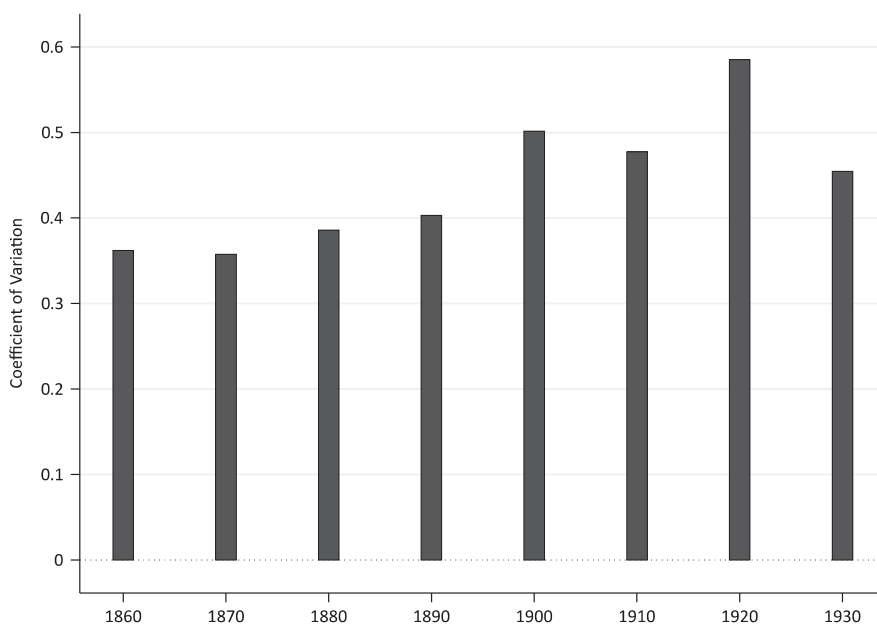
From the standpoint of the evolution of regional inequality in terms of output *per capita*, Rosés *et al.* (2010) have shown that regional inequality grew until 1900. As can be seen in Figure 1, the new evidence supplied for the second half of the 19th century shows that this increase was concentrated especially in the 1890s. From then on the trend was for these territorial disparities to remain, although the aftermath of the First World War was characterised by another increase in inequality.

Rosés *et al.* (2010) have explored the reasons that may lie behind this evolution of inequality. They carried out an analysis that makes it possible to differentiate between the elements most closely linked to regional specialisation, as suggested by traditional international trade theory (Heckscher-Ohlin), and the existence of differences in productivity between provinces, which would more likely be related to explanations typical of NEG. The results suggest that structural change, that is, differences in the timing and intensity of the arrival of industrialisation between regions, was responsible for the increase in inequality that came about in the second half of the 19th century. The growing differences in production structures, however, tended to become smaller in the early decades of the 20th century, when industrialisation spread to a greater number of provinces, especially during the inter-war years (Betrán 1999; Tirado and Martínez-Galarraga 2008).

(footnote continued)

been found in the spatial concentration of the industrial sector in the United States (Kim 1995) and France (Combes *et al.* 2011), where the change of trend came about before the Second World War.

FIGURE 1
REGIONAL INEQUALITY IN GDP *PER CAPITA* IN SPAIN, 1860-1930 (NUTS3 PROVINCES).



Note: Population-weighted coefficient of variation for NUTS3 provinces.

Sources: For 1860 and 1900-1930, see Rosés *et al.* (2010); for 1870-1890, see text.

In this period, differences in productivity were more important in explaining regional inequality. These differences would reflect that the impact of the forces of NEG became stronger over these years, and could have thus slowed down income convergence among the regions of Spain.

As a result of all this, the evidence available for Spain before the Civil War shows that there was a definite trend towards the spatial agglomeration of economic activity during the early stages of modern economic growth insofar as transport costs were falling, the domestic market was becoming integrated and industrial progress meant that the increasing returns associated with economies of scale had a greater presence. In parallel to the process of spatial concentration of industrial production, regional economic inequality was increasing sharply, at least until the start of the 20th century. So, did this presence of agglomeration economies have a positive or a negative effect on regional economic growth? What impact did it have on the evolution of regional disparities during this period? What elements might have hindered the regional convergence process after

the start of the 20th century? Before we explore these questions, we need to quantify the agglomeration.

3. MEASURING SPATIAL AGGLOMERATION IN SPAIN, 1860-1930: MAIN INDICATORS

To find out about agglomeration in the Spanish provinces during the relevant period, we propose two approaches. First, we capture absolute agglomeration, as is common in the existing literature (Brülhart and Sbergami 2009), with provincial urbanisation rates, defined as the percentage of the total population of each province who live in towns of over 5,000 inhabitants. The data used to construct this indicator come from Reher (1994), and the respective Population Censuses. Second, we aim to construct an indicator that proxies for agglomeration economies in production. For this, we provide two provincial indices. Both have been computed using data for gross value added (GVA) at factor cost and Employment (EMP)⁴.

Our first index, ϕ^i , proxies for the presence of agglomeration economies at the aggregate level for each province i , and is constructed as follows:

$$\phi^i = \sum_{r=3} \left[\log \left(\frac{GVA_r^i}{area^i} \right) \cdot \left(\frac{GVA_r^i}{GVA_S^i} \right) \right] \quad [1]$$

where S represents the total GVA for each of the forty-nine provinces i (with I being the total for Spain); r represents each economic sector (agriculture, industry and services); and each province's area is expressed in square kilometres⁵.

Additionally, we introduce an index for the presence of agglomeration economies at the sectoral level for each province i and each economic sector r (agriculture, industry and services). We denote this indicator Ω_r^i , and it can be expressed as

$$\Omega_r^i = \log \left(\frac{GVA_r^i}{area^i} \right) \quad [2]$$

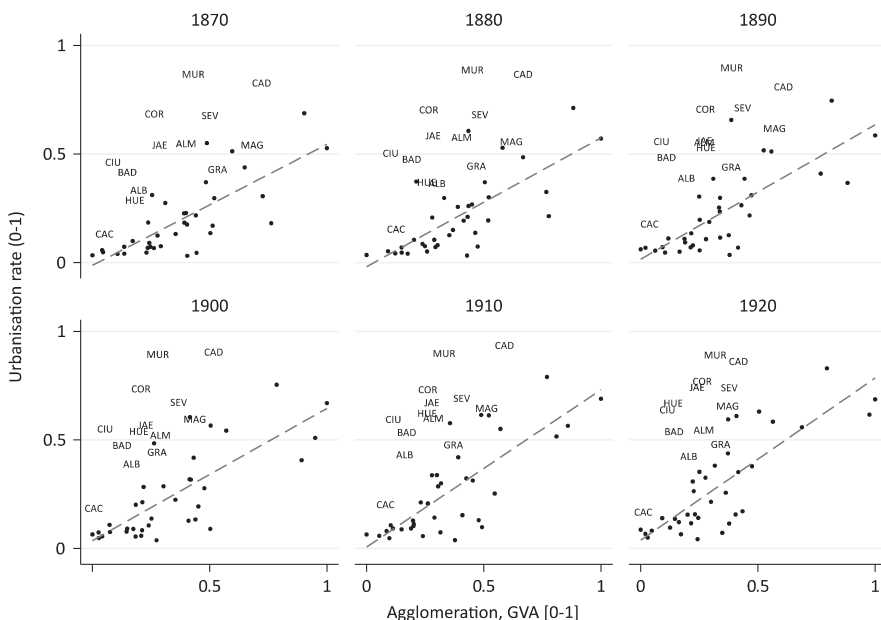
By construction, the second set of indices (ϕ^i ; Ω_r^i) measure relative provincial GVA per square kilometre and can thus be interpreted, in line with Ciccone and Hall (1996), as a measure of relative economic density⁶. These indices have been normalised in the range [0,1].

⁴ In what follows, we present the construction of the indicator using GVA at factor cost. The definition will be equivalent in the case of employment.

⁵ The two Canary Island provinces are counted as one.

⁶ Brülhart and Sbergami (2009) used within-country topographic Theil indices of aggregate employment and own-sector employment to capture the degree of concentration across space. Unfortunately, our historical data only allow us to compute these measures at regional level (NUTS2). Having said that, we found that our proxy for agglomeration economies in production ϕ^i

FIGURE 2
URBANISATION AND AGGLOMERATION ECONOMIES IN AGGREGATE
PRODUCTION (ϕ^i) IN SPAIN BY YEAR, 1870-1930 (NUTS3 PROVINCES, GROSS
VALUE ADDED (GVA)).



Notes: Dots illustrate the Canary Islands, the Balearic Islands and the northern provinces. The southern provinces (Almería, Cádiz, Córdoba, Granada, Huelva, Jaén, Málaga, Sevilla, Badajoz, Cáceres, Murcia, Albacete and Ciudad Real) are shown using a three-digit code.

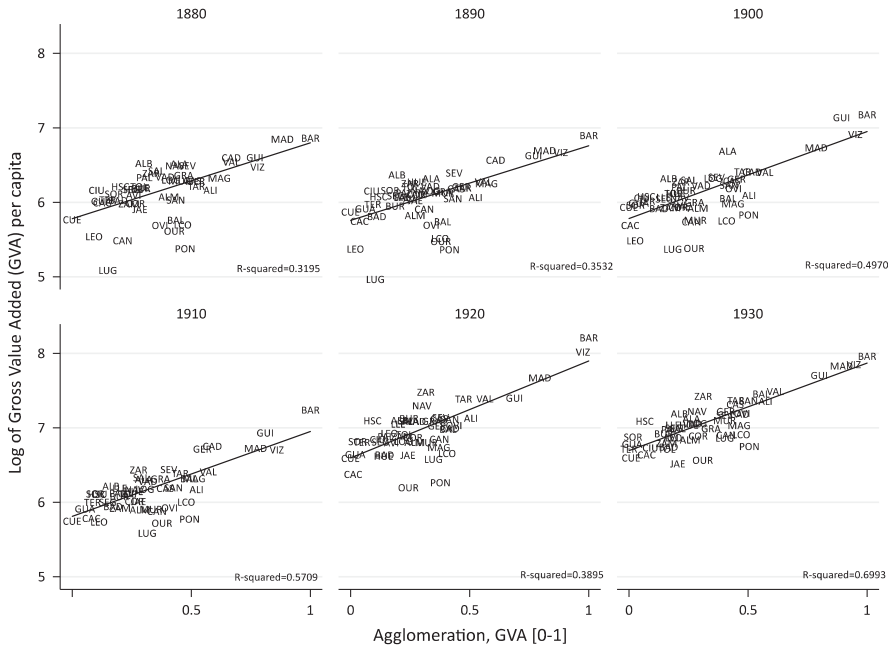
Source: See text.

These measures thus capture different aspects linked to the presence of agglomeration economies in production. The urbanisation rate, on the other hand, is an absolute and population-related measure of agglomeration. Although we employ both in the empirical analysis, our measures (ϕ^i ; Ω_r^i) allow us to consider aspects of agglomeration that are directly connected to the concentration of economic activity. Figure 2 shows the relationship between these two indicators. There is a positive correlation between them that increases and approaches the diagonal when the southern provinces of Spain are excluded. This, in turn, illustrates that urbanisation rates and the

(footnote continued)

is correlated with the Theil index of aggregate GVA for the period of study ranging from 0.5926 to 0.7782.

FIGURE 3
 GDP PER CAPITA AND AGGLOMERATION ECONOMIES IN AGGREGATE
 PRODUCTION (ϕ^i) IN SPAIN BY YEAR, 1870-1930 (NUTS3 PROVINCES, GROSS
 VALUE ADDED (GVA)).



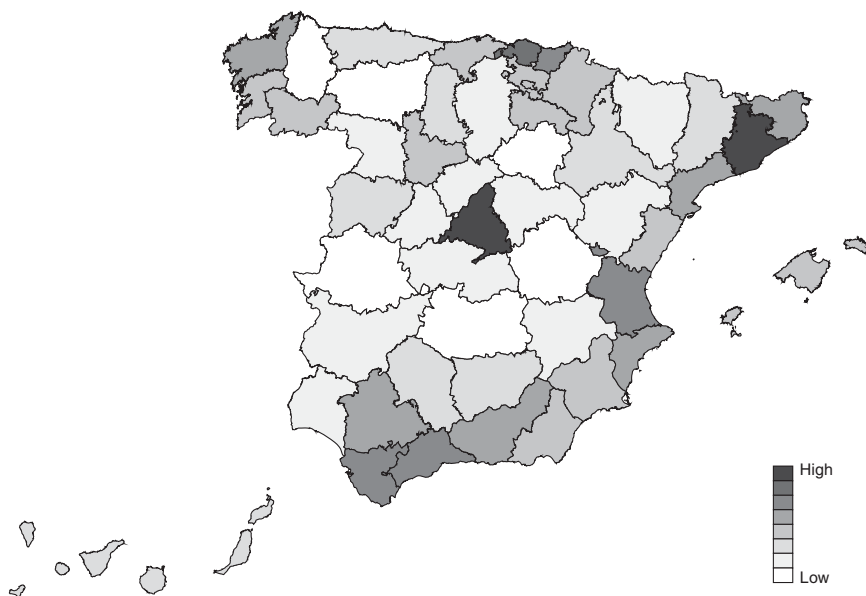
Source: See text.

presence of agglomeration economies in production do not necessarily go hand-in-hand.

Now that our indicators have been presented, an initial approximation can be made of the relationship between agglomeration economies and GDP *per capita* in each of the six time intervals selected. Indeed, Figure 3 shows that since 1870 those provinces in which, in terms of our GVA indicator, agglomeration economies were greater also registered higher levels of GDP *per capita* throughout the period.

As for the analysis of the relationship between the presence of agglomeration economies, or economic density, and the growth of GDP *per capita* at province level during the period studied, this can be done by carrying out a visual examination of Figures 4 to 6. To begin with, our indicator ϕ^i in 1870, the 1st year of the study, can be seen in Figure 4. The group of provinces with the greatest economic density, headed by Madrid and Barcelona, are mainly to be found along the eastern Mediterranean coast, in the Ebro valley, the

FIGURE 4
 AGGLOMERATION ECONOMIES IN AGGREGATE PRODUCTION (ϕ^i) IN SPAIN,
 1870 (GROSS VALUE ADDED).

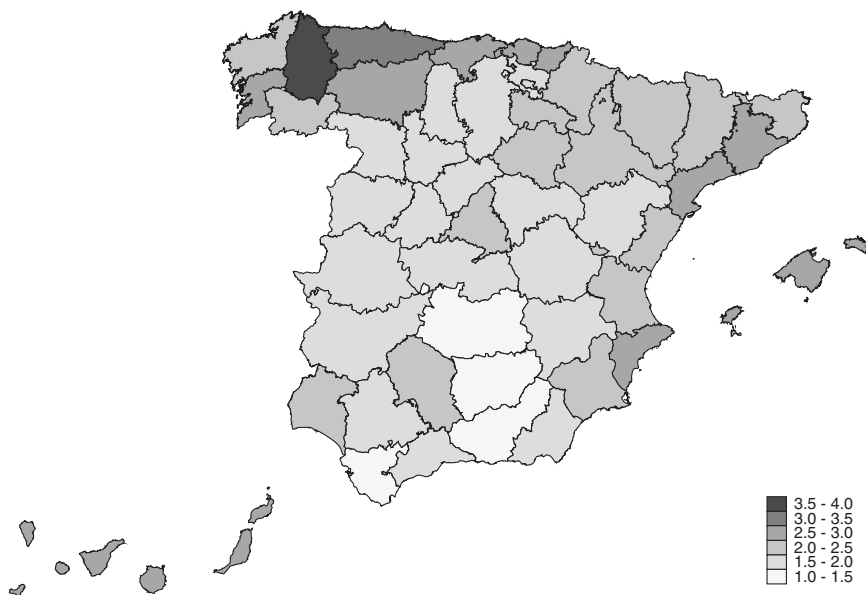


north of the peninsula and in eastern Andalusia. With the exception of eastern Andalusia and some north-western provinces, Figure 5 shows that it is these same areas in which the provinces that registered the highest growth rates between 1870 and 1930 are concentrated. Thus the maps point to the existence of a positive relationship between greater agglomeration of economic activity and subsequent economic growth. A similar relationship is observed when analysing agglomeration economies in industry (Ω_r^i) and provincial GDP *per capita* growth (Figure 6). This would indicate that the evolution of the industrial sector could be behind the observed relationship between agglomeration economies and growth, and thus deserves further examination. However, this initial exploration simply shows us an apparent relationship between the variables studied. The next section explores this relationship in greater detail and with greater rigour within the analytical framework linked to the literature on economic growth.

4. EMPIRICAL MODEL

Following Brühlhart and Sbergami (2009), we set out to empirically evaluate the relationship between the presence of agglomeration economies

FIGURE 5
ANNUAL GROWTH RATE (PER CENT) OF GDP *PER CAPITA* IN SPAIN, 1870-1930.



and economic growth across Spanish provinces (NUTS3) for the period 1870-1930. As Brühlhart and Sbergami (2009) prudently pointed out, choosing an empirical growth model is far from being trivial. We therefore propose a simple regression model based on the existing literature (Barro and Sala-i-Martin 1991; Sala-i-Martin *et al.* 2004):

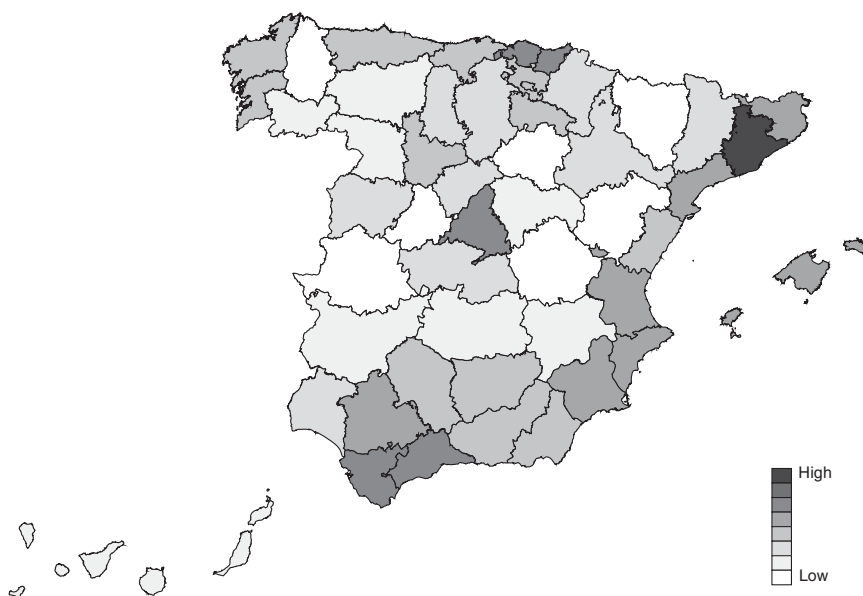
$$g_{i,p} = \alpha y_{i,t-T} + \beta A_{i,t-T} + \gamma Z_{i,t-T} + \mu_i + v_t + \varepsilon_{i,p} \quad [3]$$

The dependent variable encapsulates the average annual growth rate of GDP *per capita* for province *i* over decade *p*, that is $(t, t-T)$, where *t* and *T* stand for the year and length of period ($T = 10$). This has been calculated as

$$g_{i,p} = (y_{i,t} - y_{i,t-T}) / T \quad [4]$$

where $y_{i,t}$ is the log of GDP *per capita* for province *i* in year *t*. Therefore, $y_{i,t-T}$ is the log of the initial value of GDP *per capita* or «catch-up» term. $A_{i,t-T}$ captures the presence of agglomeration economies, our variable of interest, in province *i* at the start of each decade, while *Z* is a vector of control variables, also measured at the start of each decade. As mentioned earlier, we provide distinct proxies for agglomeration economies. First, we use $(\phi^i; \Omega_t^i)$ for agglomeration economies in production (GVA, EMP). Then, urbanisation

FIGURE 6
 AGGLOMERATION ECONOMIES IN INDUSTRIAL PRODUCTION (Ω_i^j) IN SPAIN,
 1870 (GROSS VALUE ADDED).



rates as a proxy for absolute agglomeration (Henderson 2003; Brülhart and Sbergami 2009).

The set of control variables includes the literacy rate as a proxy for human capital stock and the log of the stock of infrastructures as a proxy for the regional stock of public capital⁷. The spatial distribution of literacy in Spain showed an increasing polarisation over the period considered (Núñez 1992). In 1860, literacy rates were higher in the area going from Madrid, through Castile-Leon and northern Spain. By 1900, there was a marked division of the country between the north and the south in terms of education. The only exceptions were Galicia in the north-west and south-western Andalusia. In 1930, the transition to universal literacy was almost completed in the north, and the south had started to close the gap with the northern provinces although illiteracy was still an important issue on the eve of the Spanish Civil War (half of the provinces had literacy rates between 50 per cent and 70 per cent).

In addition, the evolution of infrastructures is marked by the construction of the railway network, which was especially intense in the second half of the

⁷ Literacy rates come from Núñez (1992) and the stock of infrastructures is obtained from Herranz (2008).

19th century. In the first stage, between 1855 and 1866, the railway connected the main economic centres. In this period, the length of the network grew from 440 to 5,076 km. In the second stage, spanning from 1873 to 1896, the railway reached other areas of the country (Herranz 2005). By 1901, all provincial capitals had been connected to the network⁸.

In addition, the share of mining (as a percentage of GDP) aims to control for regional differences in natural resources (Rosés *et al.* 2010 and Appendix B). Spain has traditionally been well endowed with mineral resources. Mining experienced a boost in the late 1900s as a result of greater international demand. Major reserves of iron ore were located in the provinces of Vizcaya and Santander in the north, and in Málaga in the south; lead ore mines existed in southern Spain (Murcia, Jaén, Almería, Córdoba, Granada, Badajoz and Ciudad Real); copper was abundant in the south (Huelva); mercury was extracted in Almadén (Ciudad Real); and finally, coal was mainly concentrated in the north (Asturias and León) and in some southern provinces (Ciudad Real and Córdoba)⁹.

These variables are included to control for other relevant factors that positively affect the process of regional economic growth. Although our selection is limited, we have attempted to reduce the potential omission of variables with our set of controls. This limitation to our selection results from a potential problem of instrument proliferation when using the generalised methods of moments (GMM) estimator, as we explain below. Table 3 shows the descriptive statistics of our main variables. Finally, μ_i , ν_t and $\varepsilon_{i,p}$ represent a province-specific effect, a time-specific effect and a well-behaved error term, respectively. Generally speaking, our chosen specification states that economic growth depends on the initial level of income, agglomeration and a set of control variables.

To estimate the relationship between agglomeration and economic growth, we begin with a pooled ordinary least squares (OLS) regression. However, estimating a dynamic panel data model with a pooled OLS regression ignores the province-specific effects ($\mu_i = 0$) and the potential endogeneity problem related to our main explanatory variables. The panel regression allows us to control for omitted province-specific time-invariant effects, and thus $\mu_i \neq 0$. Furthermore, these omitted variables might possibly be correlated with our main variable of interest, agglomeration, and cause economic growth. If this were the case, the strict exogeneity assumption would be violated and therefore our estimate β could be biased. To solve the potential endogeneity problem, we adopt a panel regression using decades or 10-year intervals. The difficulty of finding appropriate valid external

⁸ The expansion of the railway network in Spain during the period of analysis can be consulted in Appendix C.

⁹ Coal reserves were small. Moreover, the coal was of poor quality and difficult to extract. Spanish production, however, did benefit from protectionist measures since 1891, for example, tariffs.

TABLE 3
DESCRIPTIVE STATISTICS

Variables	Mean	SD	Maximum	Minimum	Observations
GDP <i>per capita</i>					
Annual growth rate	0.021	0.034	0.131	-0.038	294
Level (pesetas)	688	474	3,763	95	343
Agglomeration, normalised [0-1]					
Aggregated, GVA	0.342	0.222	1.000	0.000	343
Agriculture, GVA	0.473	0.227	1.000	0.000	343
Industry, GVA	0.359	0.223	1.000	0.000	343
Services, GVA	0.367	0.244	1.000	0.000	343
Total, Employment	0.309	0.231	1.000	0.000	343
Urbanisation rate (0-1)	0.342	0.252	0.944	0.031	343
Literacy rate (0-1)	0.490	0.214	1.000	0.160	343
Stock of infrastructure (millions pesetas)	61.8	42.8	384.4	2.2	343
Share of mining (GDP)	0.013	0.026	0.154	0.000	343

Note: GVA: gross value added.

instruments for our panel regression recommends the use of the system GMM estimator, proposed by Arellano and Bover (1995) and further developed by Blundell and Bond (1998, 2000)¹⁰.

The system GMM estimator combines equations in first differences and levels, both of which are estimated simultaneously. Given our specification, all time-dependent variables are assumed to be potentially endogenous. Hence, first differences are instrumented with lagged levels and levels with lagged first differences. There is therefore no need for us to find external instruments for our explanatory variables. Nevertheless, the generation of numerous instruments in system GMM, as Roodman (2009a) claims, could become a major concern by overfitting the endogenous variables. Given our panel, a maximum of two lags has been imposed to lessen the potential problem of instrument proliferation. We report the Hansen *J* tests for the joint validity of the instruments and have estimated the system GMM

¹⁰ Arellano and Bond (1991) introduced the GMM estimator for dynamic panel data models and proved that it offered significant gains in efficiency when compared with conventional instrumental variable approaches. With the GMM estimator, the endogenous variables in first differences were instrumented with lagged levels. However, Blundell and Bond (1998, 2000) found that lagged levels were weak instruments, especially when time series are persistent. As a result, the system GMM estimator was developed.

estimator with the `xtabond2` package for Stata 13 developed by Roodman (2009b)¹¹.

5. RESULTS: AGGLOMERATION AND ECONOMIC GROWTH IN SPAIN, 1870-1930

Table 4 reports the pooled OLS results. Our sample contains 294 observations corresponding to the forty-nine Spanish provinces (NUTS3) over six decades. Columns (2) and (4) show our chosen specification with two proxies for agglomeration: GVA and EMP. The pooled OLS regression performs well ($R^2 > 0.77$). The coefficients of our main explanatory variables are highly significant and reasonably stable. The prior hypotheses are empirically supported. Conditional convergence ($\alpha < 0$) is also supported, while agglomeration ($\beta > 0$) appears to have a positive effect on economic growth. The coefficient of our proxy for agglomeration of GVA (0.028) is greater than that for EMP (0.017), and so is the R^2 . This will be a recurring feature. Our proxy for agglomeration of GVA will therefore be our preferred one. Finally, the set of control variables is statistically significant and positively related to economic growth, except for share of mining. Although these preliminary results are encouraging, we need to exercise caution because a pooled OLS regression ignores province-specific effects ($\mu_i = 0$) and the potential endogeneity related to agglomeration, our main variable of interest.

Our two proxies (ϕ^i) do not fully capture the presence of agglomeration economies in province i and year t . We thus need to test our preliminary results. To this end we estimate our preferred specification with urbanisation rates as a proxy for absolute agglomeration. Columns (5) and (6) in Table 4 show the pooled OLS regression with urbanisation rates. Once again, the presence of agglomeration economies, measured with urbanisation, has a positive and statistically significant effect on regional economic growth. However, our indices (ϕ^i) for GVA and employment seem to perform slightly better. In this regard, proxies reflecting economic activity appear to be a more appropriate measure of agglomeration than population-related ones.

To overcome the potential endogeneity problem associated with agglomeration we use the system GMM estimator. This approach takes into account the omission of province-specific time-invariant effects ($\mu_i \neq 0$) that could be correlated with agglomeration, thereby biasing our estimate β . The panel includes forty-nine Spanish provinces (NUTS3) over six decades, 1870-1930. As stated earlier, all time-dependent variables will be treated as potentially endogenous. To mitigate the potential problem of instrument

¹¹ Numerous instruments can overfit endogenous variables and bias the estimators (Roodman 2009a). As a result, a large instrument count can weaken tests for the validity of instruments. Symptomatic of this problem would be P -values (≈ 1.00) for the Hansen J tests.

TABLE 4
 AGGLOMERATION AND GDP *PER CAPITA* GROWTH IN SPAIN, 1870-1930 (ORDINARY LEAST SQUARES (OLS))

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	OLS	OLS	OLS
GDP <i>per capita</i>	-0.017 (0.002)***	-0.029 (0.004)***	-0.015 (0.002)***	-0.020 (0.003)***	-0.016 (0.002)***	-0.022 (0.004)***
Agglomeration, GVA	0.017 (0.005)***	0.028 (0.006)***				
Agglomeration, EMP			0.019 (0.006)***	0.017 (0.006)***		
Urbanisation rate					0.011 (0.004)**	0.012 (0.004)***
Literacy rate		0.016***		0.014***		0.023***
Stock of infrastructure		0.005***		0.005***		0.003**
Share of mining		-0.022		-0.022		-0.037
Time dummies	NO	YES	NO	YES	NO	YES
Constant	YES	YES	YES	YES	YES	YES
R ²	0.053	0.784	0.060	0.779	0.049	0.770
N	294	294	294	294	294	294

Notes: GVA: gross value added, EMP: employment.

Dependent variable: annual growth rate of GDP *per capita*; independent variables are lagged; all variables are in logarithmic scale except for rates/shares (0-1).

Statistically significant at *10%, **5%, ***1% levels.

proliferation, and given the small size of our panel, we have instrumented GDP *per capita* and agglomeration economies with a maximum of two lags. The remainder of the time-dependent variables are instrumented with just one lag. Table 5 reports our main results. Columns (1), (3) and (5) show the one-step estimation, while columns (2), (4) and (6) illustrate the two-step estimation. We report both estimation procedures to shed further light on our results¹².

The results are in line with those reported above. The relationship between agglomeration economies and regional economic growth is empirically supported. The β coefficients are positive and statistically significant when our measure of agglomeration (GVA) is used. Moreover, these estimate values are similar to the ones presented in Table 4. Second, we also find strong support for the prior probability of conditional convergence ($\alpha < 0$). The estimated values fall between 2 per cent and 3 per cent, which is in line with the existing literature on regional convergence (Barro and Sala-i-Martin 1991). Third, the set of control variables performs well. Literacy is highly significant and has a positive effect on economic growth, while the stock of infrastructures also has a positive impact, although now it does not appear to be statistically significant. Finally, the share of mining remains statistically insignificant.

As Table 5 illustrates, the system GMM estimation performs reasonably well given our small panel and strong restrictions. To test for the joint validity of the instruments, we use the Hansen *J* test (Hansen 1982)¹³. We also report the second-order autocorrelation tests. The *P*-values of the Hansen *J* test could certainly be disheartening. However, it is important to remember that although our panel contains just 294 observations, even if we restrict ourselves to a maximum of two lags we generate up to sixty-four instruments. This would explain the unsatisfactory *P*-values in the Hansen *J* test. The weakness of the Hansen test should not, therefore, discourage this empirical approach. The difficulty in finding appropriate external instruments for the potentially endogenous variables such as agglomeration led us to adopt the system GMM estimator. Hence, in spite of the small panel, the system GMM estimation strengthens our previous findings, which found a strong relationship between agglomeration economies and regional economic growth in Spain during the early stages of industrialisation.

Finally, we disaggregate our measure of agglomeration (GVA) by activity (agriculture, industry and services). Table 6 reports the main results using

¹² Although the two-step estimation is more efficient for system GMM, the asymptotic standard errors tend to be downward biased in small finite samples. To correct this bias we follow Windmeijer (2005) and report the («Windmeijer») corrected standard errors in the two-step estimation. The two-step system GMM estimator appears to be the preferred one among researchers.

¹³ We prefer to report Hansen *J* statistics instead of Sargan tests because the latter assume homoscedasticity. Nevertheless, the Sargan tests are available on request.

TABLE 5
AGGLOMERATION AND GDP PER CAPITA GROWTH IN SPAIN, 1870-1930 (SYSTEM GENERALISED METHODS OF MOMENTS)

	(1)	(2)	(3)	(4)	(5)	(6)
	One step	Two step	One step	Two step	One step	Two step
GDP <i>per capita</i>	-0.031***	-0.029***	-0.022***	-0.023***	-0.027***	-0.027***
Agglomeration, GVA	0.027***	0.026***				
Agglomeration, EMP			0.014*	0.011		
Urbanisation rate					0.018*	0.016
Literacy rate	0.024***	0.024**	0.022***	0.025**	0.041***	0.038**
Stock of infrastructure	0.003	0.003	0.003	0.004	0.003	0.003
Share of mining	-0.027	-0.011	-0.013	0.050	-0.028	-0.004
Time dummies	YES	YES	YES	YES	YES	YES
Constant	YES	YES	YES	YES	YES	YES
Instruments	64	64	64	64	64	64
Provinces	49	49	49	49	49	49
<i>N</i>	294	294	294	294	294	294
Hansen <i>J</i> test	43.73 (0.814)	43.73 (0.814)	41.45 (0.875)	41.45 (0.875)	41.97 (0.862)	41.97 (0.862)
AR (1)	-4.63 (0.000)	-4.34 (0.000)	-4.51 (0.000)	-4.36 (0.000)	-4.68 (0.000)	-4.17 (0.000)
AR (2)	1.97 (0.049)	1.85 (0.064)	1.93 (0.053)	1.79 (0.073)	1.98 (0.048)	1.80 (0.072)

Notes: GVA: gross value added, EMP: employment.

Dependent variable: annual growth rate of GDP *per capita*; independent variables are lagged; all variables are in logarithmic scale except for rates/shares (0-1).

Statistically significant at *10%, **5%, ***1% levels.

TABLE 6
 SECTORAL AGGLOMERATION AND GDP *PER CAPITA* GROWTH IN SPAIN, 1870-1930 (SYSTEM GENERALISED METHODS OF MOMENTS)

Variables	(1)	(2)	(3)	(4)	(5)
	One step	One step	One step	One step	One step
GDP <i>per capita</i>	-0.031***	-0.015*	-0.030***	-0.027***	-0.030***
Agglomeration, GVA	0.027***				
GVA, Agriculture		-0.003			-0.013
GVA, Industry			0.020*		0.039*
GVA, Services				0.011	-0.001
Literacy rate	0.024***	0.013	0.024**	0.024**	0.012
Stock of infrastructure	0.003	0.003	0.003	0.003	0.004
Share of mining (GDP)	-0.027	-0.026	-0.037	-0.026	-0.114*
Time dummies	YES	YES	YES	YES	YES
Constant	YES	YES	YES	YES	YES
Instruments	64	64	64	64	92
Provinces	49	49	49	49	49
<i>N</i>	294	294	294	294	294
Hansen <i>J</i> test	43.92 (0.809)	42.79 (0.841)	43.38 (0.824)	41.52 (0.873)	45.80 (0.999)
AR (1)	-4.63 (0.000)	-4.48 (0.000)	-4.73 (0.000)	-4.64 (0.000)	-4.82 (0.000)
AR (2)	1.97 (0.049)	2.01 (0.044)	1.98 (0.048)	1.96 (0.050)	2.06 (0.039)

Notes: GVA: gross value added.

Dependent variable: annual growth rate of GDP *per capita*; independent variables are lagged; all variables are in logarithmic scale except for rates/shares (0-1).

Statistically significant at *10%, **5%, ***1% levels.

the one-step system GMM estimator¹⁴. Column (1) shows our baseline dynamic panel data model. Columns (2), (3) and (4) illustrate our chosen specification with a proxy for agglomeration economies (GVA) by economic activity. Column (5) presents the results of the inclusion of the three different proxies. On the whole, we found a stable empirical model. Once again conditional convergence ($\alpha < 0$) is supported, with the estimates ranging from 1.5 per cent to 3.1 per cent. This is in line with the existing literature on regional convergence. Literacy and the stock of infrastructures have a positive impact, though the latter is statistically insignificant as in Table 5. As regards our main variable of interest, only the presence of «industrial» agglomeration (GVA) is statistically significant. This supports our prior hypothesis. These findings can also be seen in column (5), where our three proxies for agglomeration are included. To conclude, the existence of a positive relationship between agglomeration economies and economic growth across Spanish provinces (NUTS3) for the period 1870-1930 emerges as our main finding. This relationship seems robust and closely related to the presence of agglomeration economies in industry.

6. CONCLUSIONS

This article has explored the existence of a relationship between the presence of agglomeration economies and regional economic growth in Spain. Following the methodology proposed by Brülhart and Sbergami (2009), we have carried out an analysis of regional convergence in Spain between 1870 and 1930, a period in which the Spanish economy underwent the early stages of its development process. Along with control variables characteristic of this literature such as initial provisions for cumulative factors like human capital and infrastructures, the study has considered the explanatory potential of different indicators for the agglomeration of population and production. For this exercise, we have used a data set that includes estimates of regional GDP *per capita* in Spain for 1870, 1880 and 1890 for the first time.

The results indicate that regional growth in Spain during these years followed a path of conditional convergence in which initial differences in the provision of human capital played a particularly important role, enabling us to understand the different growth trajectories of the different Spanish regions. In addition, and in connection with the central aspect of this study,

¹⁴ Although the two-step system GMM estimator is more efficient, we use a one-step system GMM because the number of potentially endogenous variables increases with the disaggregation of agglomeration by economic activity, and given our small panel, the resulting large set of instruments becomes a serious concern. Nevertheless, columns (1) and (2) of Table 5 show little change in our coefficients of interest. We have also used the one-step GMM estimator for our proxy for agglomeration of employment (EMP). The results are fairly similar and are available on request.

the results support the existence of a positive, robust relationship between the presence of agglomeration economies (especially in the industrial activity) and their subsequent growth trajectories.

Therefore, we provide evidence in support of a trade-off between inequality in the spatial distribution of economic activity and economic growth. This, in turn, goes in line with Brülhart and Sbergami (2009). Our study, however, avoids some of the limitations identified in the literature. Spain, during the period 1860-1930, was in the early stages of modern economic growth. Like other European economies of the period, this economy was driven by industrialisation, which in the context of the second half of the 19th century and first third of the 20th century, was characterised by the generation of agglomeration economies. Also, the historical period explored in the case of Spain corresponds with the construction of the railway network, which brought about a rapid reduction in transport costs, an element that would favour the presence of agglomeration economies typical of industrial activities. Finally, the data set used to carry out the study was comprised of information involving relatively small territorial units in which, these being regions belonging to the same state, institutional differences appear to be less important when it comes to explaining the different growth trajectories.

From the standpoint of Spanish economic history, the exercise provides a relevant contribution as it is one of the first to discuss the reasons for regional growth in Spain during the early stages of the country's economic development using an analytical framework typical of the empirical literature on economic growth. The new evidence shows that the increase in regional economic inequality between 1870 and 1930 was due to two central elements. On the one hand, it is related to the unequal initial provisions of cumulative production factors such as human capital, while on the other the exercise shows that the original economic differences between territories could also have increased due to the direct relationship existing between the initial levels of agglomeration of production and the subsequent rates of growth. Therefore, the evidence presented indicates that, in line with the explanations supplied by NEG, the presence of agglomeration economies in some production processes, especially industrial ones, in a context of market integration brought about the start of a cumulative causation process that increased regional economic inequality in Spain during the second half of the 19th century and hindered its reduction during the first third of the 20th century.

SUPPLEMENTARY MATERIAL

To view supplementary material/s for this article, please visit <http://dx.doi.org/10.1017/S0212610915000300>

REFERENCES

- AGUILAR-RETURETA, J. (2015): «The GDP Per Capita of the Mexican Regions (1895-1930): New Estimates». *Revista de Historia Económica/Journal of Iberian and Latin American Economic History* 33 (3), pp. 387-423.
- ARELLANO, M., and BOND, S. (1991): «Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations». *The Review of Economic Studies* 58 (2), pp. 277-297.
- ARELLANO, M., and BOVER, O. (1995): «Another Look at the Instrumental Variable Estimation of Error-Components Models». *Journal of Econometrics* 68 (1), pp. 29-51.
- BADIA-MIRÓ, M.; GUILERA, J., and LAINS, P. (2011): «Regional Incomes in Portugal: Industrialisation, Integration and Inequality, 1890-1980». *Revista de Historia Económica/Journal of Iberian and Latin American Economic History* 30 (2), pp. 225-244.
- BALDWIN, R., and MARTIN, P. (2004): «Agglomeration and Regional Growth», in J. V. Henderson, and J. F. Thisse (eds), *Handbook of Regional and Urban Economics*, vol. 4, Ch. 60. Amsterdam: Elsevier, pp. 2671-2711.
- BARRO, R., and SALA-I-MARTIN, X. (1991): «Convergence Across States and Regions». *Brookings Papers on Economic Activity* 1, pp. 107-182.
- BETRÁN, C. (1999): «Difusión y Localización Industrial en España Durante el Primer Tercio del Siglo XX». *Revista de Historia Económica* 17 (3), pp. 663-696.
- BLUNDELL, R., and BOND, S. (1998): «Initial Conditions and Moment Restrictions in Dynamic Panel Data Models». *Journal of Econometrics* 87 (1), pp. 115-143.
- BLUNDELL, R., and BOND, S. (2000): «GMM Estimation with Persistent Panel Data: An Application to Production Functions». *Econometrics Reviews* 19 (3), pp. 321-340.
- BRINGAS, M. A. (2000): *La Productividad de los Factores en la Agricultura Española (1752-1935)*. Madrid: Banco de España.
- BRÛLHART, M., and SBERGAMI, F. (2009): «Agglomeration and Growth: Cross-Country Evidence». *Journal of Urban Economics* 65 (1), pp. 48-63.
- BUYST, E. (2012): «Continuity and Change in Regional Disparities in Belgium During the Twentieth Century». *Journal of Historical Geography* 37 (3), pp. 329-337.
- CICCONI, A. (2002): «Agglomeration Effects in Europe». *European Economic Review* 46 (2), pp. 213-227.
- CICCONI, A., and HALL, R. E. (1996): «Productivity and the Density of Economic Activity». *American Economic Review* 86 (1), pp. 54-70.
- COMBES, P. P.; LAFOURCADE, M., THISSE, J. F., and TOUTAIN, J. C. (2011): «The Rise and Fall of Spatial Inequalities in France: A Long-Run Perspective». *Explorations in Economic History* 48 (2), pp. 243-271.
- CORDERO, R., and MENÉNDEZ, F. (1978): «El Sistema Ferroviario Español», in M. Artola (ed.), *Los Ferrocarriles en España: 1844-1943*, vol. 1. Madrid: Banco de España, pp. 161-338.
- CRAFTS, N. F. R. (2005): «Regional GDP in Britain, 1871-1911: Some Estimates». *Scottish Journal of Political Economy* 52 (1), pp. 54-64.
- CROZET, M., and KOENIG, P. (2005): «The Cohesion Vs. Growth Tradeoff. Evidence from EU Regions (1980-2000)». *ERSA Conference Papers*, European Regional Science Association, ersa05p716.
- DAVIS, D. R., and WEINSTEIN, D. E. (1999): «Economic Geography and Regional Production Structure: An Empirical Investigation». *European Economic Review* 43 (2), pp. 379-407.

- DAVIS, D. R., and WEINSTEIN, D. E. (2003): «Market Access, Economic Geography and Comparative Advantage: An Empirical Test». *Journal of International Economics* 59 (1), pp. 1-23.
- ENFLO, K., and ROSÉS, J. R. (2015): «Coping with Regional Inequality in Sweden: Structural Change, Migrations, and Policy, 1860-2000». *Economic History Review* 68 (1), pp. 191-217.
- FELICE, E. (2011): «Regional Value Added in Italy, 1891-2001, and the Foundation of a Long-Term Picture». *Economic History Review* 64 (3), pp. 929-950.
- FUJITA, M., and THISSE, J. F. (2002): «Agglomeration and Market Interaction». *CEPR Discussion Papers*, no. 3362.
- GARDINER, B.; MARTIN, R., and TYLER, P. (2011): «Does Spatial Agglomeration Increase National Growth? Some Evidence from Europe». *Journal of Economic Geography* 11 (6), pp. 979-1006.
- GEARY, F., and STARK, T. (2002): «Examining Ireland's Post-Famine Economic Growth Performance». *Economic Journal* 112 (482), pp. 919-935.
- GEARY, F., and STARK, T. (2015): «Regional GDP in the UK, 1861-1911: New Estimates». *Economic History Review* 68 (1), pp. 123-144.
- GOLLIN, D. (2002): «Getting Income Shares Right». *Journal of Political Economy* 110 (21), pp. 458-474.
- GRUPO DE ESTUDIOS DE HISTORIA RURAL. (1991): *Estadísticas Históricas de la Producción Agraria Española, 1859-1935*. Madrid: Ministerio de Agricultura, Pesca y Alimentación.
- HANSEN, L. P. (1982): «Large Sample Properties of Generalized Method of Moments Estimators». *Econometrica* 50 (4), pp. 1029-1054.
- HENDERSON, V. (2003): «The Urbanization Process and Economic Growth: The So-What Question». *Journal of Economic Growth* 8 (1), pp. 47-71.
- HENNING, M.; ENFLO, K., and ANDERSSON, F. N. G. (2011): «Trends and Cycles in Regional Economic Growth: How Spatial Differences Shaped the Swedish Growth Experience from 1860-2009». *Explorations in Economic History* 48 (4), pp. 538-555.
- HERRANZ, A. (2005): «La Reducción de los Costes de Transporte en España (1800-1936)». *Cuadernos Económicos del ICE* 70, pp. 183-203.
- HERRANZ, A. (2007): «Infrastructure Investment and Spanish Economic Growth, 1850-1935». *Explorations in Economic History* 44 (3), pp. 452-468.
- HERRANZ, A. (2008): *Infraestructuras y Crecimiento Económico en España (1850-1935)*. Madrid: Fundación de los Ferrocarriles españoles.
- HIRSCHMAN, A. O. (1958): *The Strategy of Economic Development*. New Haven, CT: Yale University Press.
- KIM, S. (1995): «Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987». *Quarterly Journal of Economics* 110 (4), pp. 881-908.
- KRUGMAN, P. (1991): «Increasing Returns and Economic Geography». *Journal of Political Economy* 99 (3), pp. 483-499.
- KUZNETS, S. (1955): «Economic Growth and Income Inequality». *American Economic Review* 45 (1), pp. 1-28.
- MADRAZO, S. (1984): *El Sistema de Transportes en España, 1750-1850*. Madrid: Turner.
- MARTIN, P., and OTTAVIANO, G. I. P. (1999): «Growing Locations: Industry Location in a Model of Endogenous Growth». *European Economic Review* 43 (2), pp. 281-302.
- MARTINEZ-GALARRAGA, J. (2012): «The Determinants of Industrial Location in Spain, 1856-1929». *Explorations in Economic History* 49 (2), pp. 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), pp. 195-212.
- MARTINEZ-GALARRAGA, J.; TIRADO, D. A., and GONZÁLEZ-VAL, R. (2015): «Market Potential and Regional Economic Growth in Spain, 1860-1930». *European Review of Economic History*, doi:10.1093/ereh/hev014.
- MIDELFART-KNARVIK, K. H.; OVERMAN, H., REDDING, S., and VENABLES, A. J. (2002): «The Location of European Industry», in European Commission (ed.), *European Economy, European Integration and the Functioning of Product Markets*, 2/2002. Brussels: Directorate-General for Economic and Financial Affairs, pp. 213-269.
- MYRDAL, G. (1957): *Economic Theory and Under-Developed Regions*. London: Methuen & Co.
- 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, pp. 23-61.
- NÚÑEZ, C. E. (1992): *La Fuente de la Riqueza: Educación y Desarrollo Económico en la España Contemporánea*. Madrid: Alianza.
- PALUZIE, E.; PONS, J., and TIRADO, D. A. (2004): «The Geographical Concentration of Industry Across Spanish Regions, 1856-1995». *Jahrbuch für Regionalwissenschaft* 24 (2), pp. 143-160.
- PARERO, A. (2001): «Industrialización, Desindustrialización y Nueva Industrialización de las Regiones Españolas (1950-2000). Un Enfoque Desde la Historia Económica». *Revista de Historia Industrial* 19-20, pp. 15-75.
- POLLARD, S. (1981): *Peaceful Conquest. The Industrialization of Europe, 1760-1970*. London: Oxford University Press.
- PRADOS DE LA ESCOSURA, L. (2003): *El Progreso Económico de España, 1850-2000*. Bilbao: Fundación BBVA.
- PRADOS DE LA ESCOSURA, L. (2008): «Inequality, Poverty and the Kuznets Curve in Spain, 1850-2000». *European Review of Economic History* 12 (3), pp. 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), pp. 1063-1091.
- REHER, D. S. (1994): «Ciudades, Procesos de Urbanización y Sistemas Urbanos en la Península Ibérica, 1550-1991», in M. Guàrdia; F. J. Monclús, and J. L. Oyón (eds), *Atlas Histórico de las Ciudades Europeas: Península Ibérica*. Barcelona: Centre de Cultura Contemporània de Barcelona – Salvat, pp. 1-30.
- ROODMAN, D. (2009a): «A Note on the Theme of Too Many Instruments». *Oxford Bulletin of Economics and Statistics* 71 (1), pp. 135-158.
- ROODMAN, D. (2009b): «How to do xtabond2: An Introduction to Difference and System GMM in Stata». *The Stata Journal* 9 (1), pp. 86-136.
- ROSÉS, J. R. (2003): «Why Isn't the Whole of Spain Industrialized? New Economic Geography and Early Industrialization, 1797-1910». *The Journal of Economic History* 63 (4), pp. 995-1022.
- 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 (2), pp. 244-257.
- ROSÉS, J. R., and SÁNCHEZ-ALONSO, B. (2004): «Regional Wage Convergence in Spain, 1850-1930». *Explorations in Economic History* 41 (4), pp. 404-425.
- SALA-I-MARTIN, X.; DOPPELHOFER, G., and MILLER, R. I. (2004): «Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach». *American Economic Review* 94 (4), pp. 813-835.
- SÁNCHEZ-ALBORNOZ, N. (1987): *The Economic Modernization of Spain, 1830-1930*. New York, NY: New York University Press.

- SÁNCHEZ-ALONSO, B. (1995): *Las Causas de la Emigración Española, 1880-1913*. Madrid: Alianza.
- SIMPSON, J. (1994): «La Producción y la Productividad Agraria Española, 1890-1936». *Revista de Historia Económica* 12 (1), pp. 43-84.
- TIRADO, D. A., and MARTÍNEZ-GALARRAGA, J. (2008): «Una Nova Estimació Retrospectiva del VAB Regional Industrial. Espanya (1860-1930)». *Documents de Treball de la Facultat de Ciències Econòmiques i Empresariales*, Universitat de Barcelona, Col·lecció d'Economia, E08/192.
- 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 (3), pp. 343-363.
- WILLIAMSON, J. G. (1965): «Regional Inequality and the Process of National Development: A Description of the Patterns». *Economic Development and Cultural Change* 13 (4, Part II), pp. 3-84.
- WINDMEIJER, F. (2005): «A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators». *Journal of Econometrics* 126 (1), pp. 25-51.

APPENDIX A

FIGURE A1
MAP OF SPANISH PROVINCES (NUTS3).



APPENDIX B: SPAIN'S PROVINCIAL GDP 1870-1890

The methodology proposed by Geary and Stark (2002) for the estimation of regional output is based on two variables: employment and productivity. To begin with, the total GDP of the Spanish economy would be the sum of its parts, or provincial GDPs:

$$Y_{\text{ESP}} = \sum_i Y_i \quad [\text{A1}]$$

where Y_i represents GDP for province i , and is defined as follows:

$$Y_i = \sum_j y_{ij} L_{ij} \quad [\text{A2}]$$

where y_{ij} stands for the average added value per worker in each province i , in sector j , whereas L_{ij} represents the number of workers in each province and sector. As, there are no direct measures of y_{ij} , it is proxied by taking the Spanish sectoral output per worker y_j , assuming that provincial labour productivity in each sector is reflected by its wage relative to the Spanish average (w_{ij}/w_j). Under these assumptions, the provincial GDP will be given by the following expression:

$$Y_i = \sum_j \left[y_j \beta_j \left(\frac{w_{ij}}{w_j} \right) \right] L_{ij} \quad [\text{A3}]$$

where w_{ij} is the wage paid in province i in sector j , w_j the Spanish wage in each sector j and β_j a scalar that preserves the relative province differences but scales the absolute values so that the provincial total for each sector adds up to the Spanish total. This model of indirect estimation, based on wage incomes, allows an estimation of GDP by province at factor cost, in current pesetas. The estimation has been undertaken for five economic sectors (agriculture, mining, manufacturing, construction and services)¹⁵. Yet, depending on the available statistical information, in some sectors we did not have to resort to the Geary-Stark methodology.

Agriculture

As in Rosés *et al.* (2010) for 1860, we have adopted a modified version of the Geary-Stark methodology. First, we take provincial agrarian wages and the male agrarian active population. Some concerns nonetheless arise. From the statistical information available it is difficult to disentangle the size of the female workforce in agriculture and the yearly number of working days, as well as its variation over provinces. Hence, we rely on the direct estimation of agrarian production by province in 1910 (Rosés *et al.* 2010 based on Simpson 1994 and Grupo de Estudios de Historia Rural 1991) and divide it by the agrarian GVA obtained using the Geary-Stark indirect estimation.

¹⁵ We then obtain the industrial GVA aggregating mining, manufacturing and construction.

The resulting scalar is used to correct our initial figures, thus assuming that the relative female workforce and the yearly days worked in each province remained constant for 1870-1890. Then, we scaled the values obtained so that the provincial total agrarian GVA adds up to the Spanish totals offered by Prados de la Escosura (2003). Wages come from Bringas (2000)¹⁶ and the male agrarian population is compiled from the Population Censuses of 1877 and 1887, respectively. Given the absence of information for 1870, the provincial agrarian GVA for that year has been computed interpolating the results for 1860 and 1880.

Mining

For the mining sector, a direct estimation is undertaken. The provincial mining production comes from the Spanish Mining Statistics (*Estadística Minera de España*) for the years 1870, 1880 and 1890. Then, we use the provincial values obtained to distribute Spain's mining GVA at factor cost (Prados de la Escosura 2003).

Manufacturing and Public Utilities

For this sector, we estimate the provincial value added taking a production function with constant returns to scale where the output is obtained from two production factors: labour and capital. Hence, the estimation is based both on wage income (Geary and Stark 2002) and capital income (following the refinement suggested by Crafts 2005). The manufacturing GVA (GVA_{it}^{MFG}) can be thus expressed as

$$GVA_{it}^{MFG} = \alpha_{it}(\omega_{it} \times L_{it}) + (1 - \alpha_{it})(r_{it} \times K_{it}) \quad [A4]$$

with α_{it} being the share of the wage income in manufacturing GVA in region i at time t , ω_{it} the wage, L_{it} the active population, r_{it} captures the returns to capital in manufacturing, whereas K_{it} stands for the capital stock. For the Spanish case, there is information available for each of the components of equation [A4], except for r_{it} . For this reason, we had to assume perfect capital mobility.

For the wage income we take the manufacturing employment in each province from the Population Censuses of 1877 and 1887. Given the absence of wage data for these years we make an interpolation using the wages available for 1860 and 1900 (Rosés *et al.* 2010 based on Madrazo 1984 and Sánchez-Alonso 1995). The provincial capital income is obtained from a fiscal source, the *Estadística Administrativa de la Contribución Industrial y de Comercio* (EACI) for the years 1878-1879 and 1890. This manufacturing tax, established in 1845, consisted of a fixed rate over the main means of production in use, and varied for each type of machinery and branch, and

¹⁶ For the estimation of 1880, we use the closest available wage data, which corresponds to 1887.

therefore it is particularly appropriate to capture the provincial capital stock. The addition of the Basque Country and Navarre absent from this fiscal source is based on the data provided in Parejo (2001). This author used the available historical indices of production to estimate the contribution of these regions to the Spanish industry in 1850 and 1900. Given that industry in the Basque Country took off in the last decades of the 19th century, we make a geometric interpolation using Parejo (2001) data to capture this trend. Then, the regional information for the Basque Country is divided using the active population in manufacturing by province in the Population Censuses of 1877 and 1887.

In order to calculate the factor shares in equation [A4], we combine the data from two sources. First, the proportions in labour and capital used in nine manufacturing sectors at the Spanish level are obtained from the first available Input-Output Table in 1958¹⁷. Second, the provincial structure for these comes from the EACI¹⁸. With this information, specific factor shares for each province and for each benchmark were constructed¹⁹. The provincial share of manufacturing in 1870 is obtained as an average of that of 1860 and 1880 and then re-scaled to the total GVA in manufacturing and public utilities in Spain according to the figures provided by Prados de la Escosura (2003).

Construction

In Spain's GVA series offered by Prados de la Escosura (2003), the construction sector includes both residential construction and public works. Spain's GVA for residential construction is distributed using the provincial urbanisation rates, calculated as the share of the population living in cities over 5,000 inhabitants (Reher 1994). The urbanisation rates for 1870 are obtained interpolating the information contained in the Population Censuses of 1860 and 1877. Public works' GVA is distributed on the basis of the provincial stock of infrastructures provided by Herranz (2008) for 1870, 1880 and 1890.

Services

For Spain, Prados de la Escosura (2003) offers information of the GVA in services disaggregated in eleven categories: transport, communications, trade, banking and insurance, housing, public administration, education, health services, hotels and restaurants, domestic services and professions. On the basis of this disaggregation, we compiled the data on the active population from the Population Censuses and then assigned to each category, according to the skills and productivity levels of the workforce, different wages.

¹⁷ In this respect, substantial international evidence shows that the output proportions in labour and capital remain relatively stable for long periods (Gollin 2002).

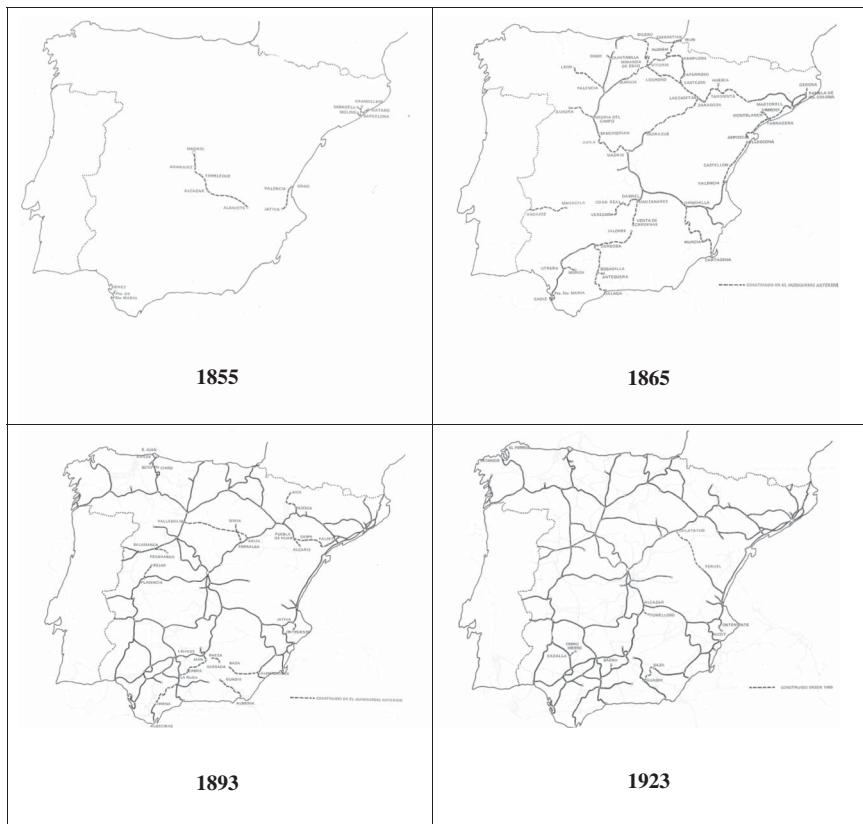
¹⁸ The industrial branches considered are food, textiles, metal, chemicals, paper, wood, ceramic, leather and miscellaneous industries.

¹⁹ For the Basque Country provinces and Navarre, given that the fiscal information is not available, we assume a labour share similar to the Spanish total.

This methodology is therefore slightly different to that applied by Geary and Stark (2002). In particular, agrarian wages were used for domestic service; an unweighted average of agrarian and industry urban wages (unskilled and skilled) for transport and communications; an unweighted average of industry urban unskilled and skilled wages for commerce, hotels and restaurants; and, finally, urban skilled wages for the remaining branches. Wages were drawn from the sources previously mentioned in other economic sectors and from Rosés and Sánchez-Alonso (2004). For 1870, the provincial values are obtained as an average of 1860 and 1880.

APPENDIX C

FIGURE A2
THE EXPANSION OF THE RAILWAY NETWORK IN SPAIN, 1855-1923.



Source: Cordero and Menéndez (1978).