serie ec
WP-EC 2014-05

Why did Spanish regions not converge before the Civil War? Agglomeration and (regional) growth revisited: Spain, 1870-1930

Alfonso Díez-Minguela, Julio Martínez-Galarraga and Daniel A. Tirado-Fabregat

Ivie

Versión: noviembre 2014 / Version: November 2014

Edita / Published by:
Instituto Valenciano de Investigaciones Económicas, S.A.
C/ Guardia Civil, 22 esc. 2 1º - 46020 Valencia (Spain)

DOI: http://dx.medra.org/10.12842/WPEC_201405

WP-EC 2014-05

Why did Spanish regions not converge before the Civil War? Agglomeration and (regional) growth revisited: Spain, 1870-1930

Alfonso Díez-Minguela, Julio Martínez-Galarraga and Daniel A. Tirado-Fabregat *

Abstract

In this paper we explore the relationship between the spatial agglomeration of economic activity 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 agglomeration of production was a key element to explain the upswing in regional income inequality in Spain during the country's early stages of development. In doing this, we present alternative indicators for agglomeration and estimate conditional growth regressions at province (NUTS3) level. The results show the existence of a positive, robust relationship between the initial levels of regional agglomeration (mainly in the industrial sector) and subsequent growth trajectories. In line with new economic geography (NEG) models, we suggest that the presence of agglomeration economies in a context of market integration favoured the emergence of a cumulative causation process that 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 classification numbers: N13, N93, O10, O40, R10.

Resumen

En este artículo se analiza la existencia de una relación entre la aglomeración espacial de la actividad y el crecimiento económico regional en España durante el periodo 1870-1930. El estudio permite revisitar la existencia de un *trade-off* entre crecimiento económico y cohesión territorial y, además, examinar si la aglomeración productiva fue 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. Los resultados muestran la existencia de una relación positiva y robusta entre el nivel inicial de aglomeración (principalmente en el sector industrial) y la posterior trayectoria de crecimiento regional. 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.
Clasificación JEL: N13, N93, O10, O40, R10.

* Universitat de València. Corresponding author: Daniel A. Tirado-Fabregat (daniel.tirado@uv.es). Authors acknowledge the financial support from the project ECO2012-39169-CO3-02.

1. Introduction

In this article we explore the existence of a relationship between the spatial agglomeration of economic activity 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 (Hirschman, 1958; Myrdal, 1957) has tended to highlight that, from a spatial perspective, economic growth has been characterized by the appearance of agglomerations in production (the formation of clusters 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 & Ottaviano, 1999; Fujita & Thisse, 2002; Baldwin & 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 Crozet & Koenig (2005), Brülhart & Sbergami (2009) and Gardiner et al. (2011).

Crozet & Koenig (2005) 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 & 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 agglomeration of population and production. They then repeat the same kind of exercise using a sample of EU countries for which they construct different indicators of agglomeration based on data for population and economic activity by branch of activity and region (NUTS2). Their results support the existence of a positive relationship between agglomeration 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) have questioned the results obtained by Brülhart & 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, as a result of their paper it is possible to identify the elements that need to be considered when it comes to suggesting how to move forward in the empirical analysis of this relationship because it has served to identify some of the problems typical of the evidence used until now.

In these circumstances, studying the case of Spain for the period 1870-1930 implies a dual contribution on 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 which, because it was still in the early stages of economic development, was characterised by the presence of high transport costs (Prados de la Escosura & Rosés, 2009; Herranz, 2007). Also, Spain has all its regions at income levels that are 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, i.e. 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)[1]. 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). On these aspects, 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. The analysis carried out by Rosés et al. (2010) shows that those Spanish regions that specialised in industrial production were the ones that achieved the highest levels of income during these years. Martínez-Galarraga et al. (2014) also 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 brief 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

[1] Specifically, the new data is for 1870, 1880 and 1890. The estimates were made following the methodology used by Rosés et al. (2010). We therefore have 10-year intervals of time for the whole period between 1860 and 1930.

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.

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 underwent what can be considered modest rates of growth. However, this trend would change after the First World War when both GDP and GDP per capita registered growth rates substantially higher than in previous periods (Table 1).

Table 1. Growth of real GDP, population and GDP per capita, 1850-1929

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

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

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 and 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 2[2].

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

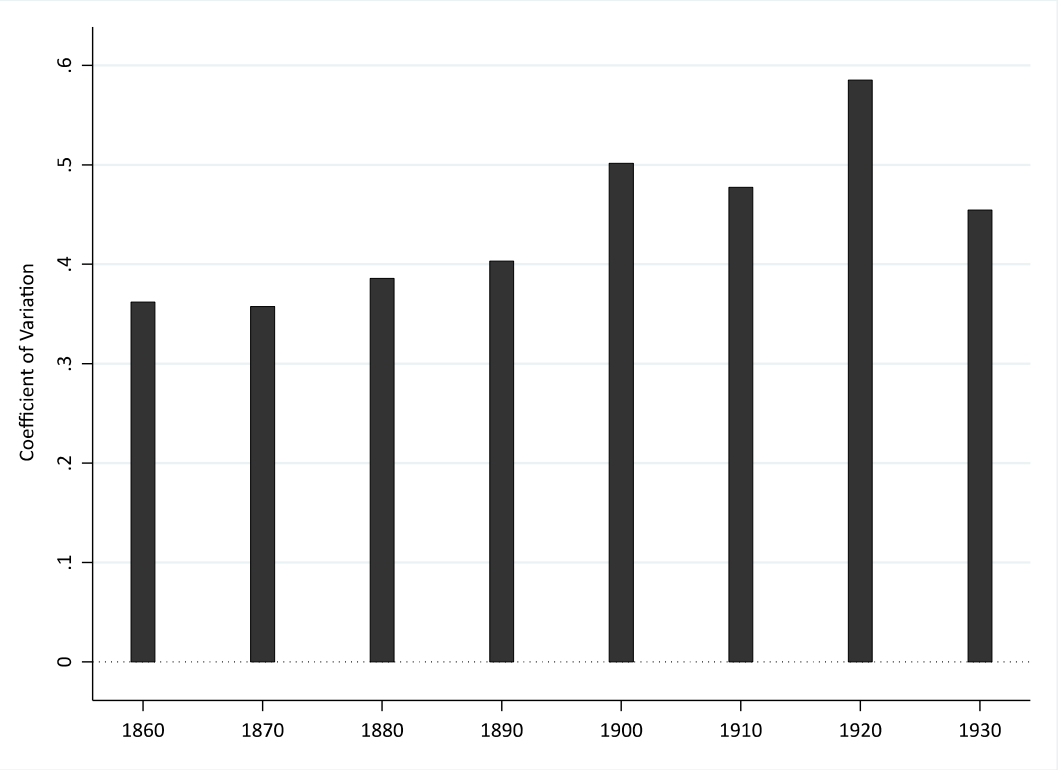
Source: Paluzie et al. (2004).

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 & 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 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), Martinez-Galarraga (2012) confirmed and extended the previous findings of Tirado et al. (2002). As the domestic market became integrated and industrialization 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.

[2] 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 been found in the spatial concentration of the industrial sector in the US (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).



Source: For 1860 and 1900-1930, Rosés et al. (2010); for 1870-1890, see text.
Note: population-weighted coefficient of variation for NUTS3 provinces.

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 new economic geography. The results suggest that structural change, i.e. 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 interwar years (Betrán, 1999; Tirado & Martínez-Galarraga, 2008). Nevertheless, the impact of the forces of new economic geography became stronger over these years since differences in productivity were more important in explaining regional inequality, and these differences could have 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 agglomeration 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 various Spanish provinces during the relevant period, we propose two different indicators. Firstly, as is common in the literature, to measure absolute agglomeration (Brülhart & Sbergami, 2009; Ottaviano & Pinelli, 2006) we use the provincial urbanisation rate, 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. Secondly, we aim to construct an indicator of economic agglomeration. In this case we provide two provincial indexes of relative agglomeration, one at sector level and one at aggregate level. Both were computed using data for gross value added (GVA) at factor cost and employment (EMP)[3].

First, taking GVA as the reference, the sectoral agglomeration index (SAGI) for each province i (with I being the total for Spain) and each economic sector r (agriculture, industry and services) can be expressed as:

$$\mathrm{SAGI}_{\mathrm{r}}^{\mathrm{i}} = \log\left(\frac{\frac{\mathrm{GVA}_{\mathrm{r}}^{\mathrm{i}}}{\mathrm{area}^{\mathrm{i}}}}{\frac{\mathrm{GVA}_{\mathrm{r}}^{\mathrm{I}}}{\mathrm{area}^{\mathrm{I}}}}\right) \qquad (1)$$

Second, the total agglomeration index (TAGI) can be defined as:

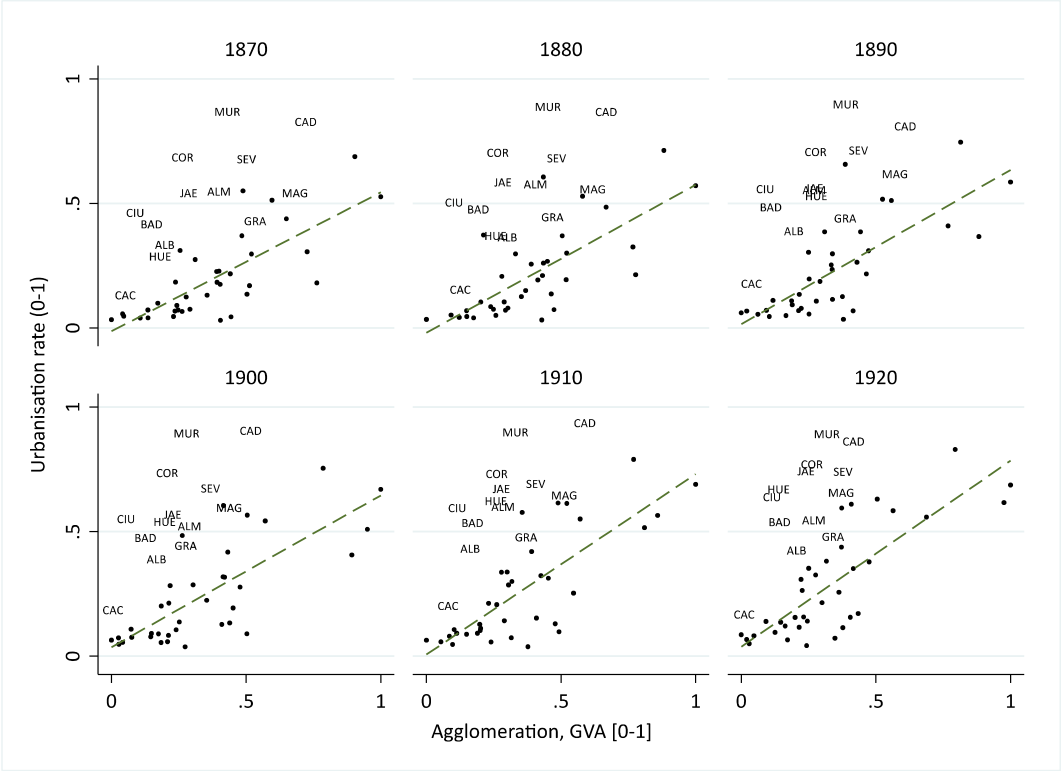
$$\mathrm{TAGI}^{\mathrm{i}} = \sum_{r=3}\left[\log\left(\frac{\frac{\mathrm{GVA}_{\mathrm{r}}^{\mathrm{i}}}{\mathrm{area}^{\mathrm{i}}}}{\frac{\mathrm{GVA}_{\mathrm{S}}^{\mathrm{I}}}{\mathrm{area}^{\mathrm{I}}}}\right)\cdot\left(\frac{\mathrm{GVA}_{\mathrm{r}}^{\mathrm{i}}}{\mathrm{GVA}_{\mathrm{S}}^{\mathrm{i}}}\right)\right] \qquad (2)$$

[3] 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.

where S represents the total GVA for each of the 49 provinces *i* considered and each province's area is expressed in square kilometres[4]. The provincial indexes of relative agglomeration have been normalised in the range [0,1].

Our two types of indicator capture different aspects linked to agglomeration. While the urbanisation rate is more associated with population, our relative indicator (for both GVA and employment) allows 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 indicates that economic agglomeration (GVA) in these provinces during the period studied is lower than that reflected in the rates of urbanisation.

Figure 2. Urbanisation rate and agglomeration (GVA) in Spain by year, 1870-1930 (NUTS3 provinces)

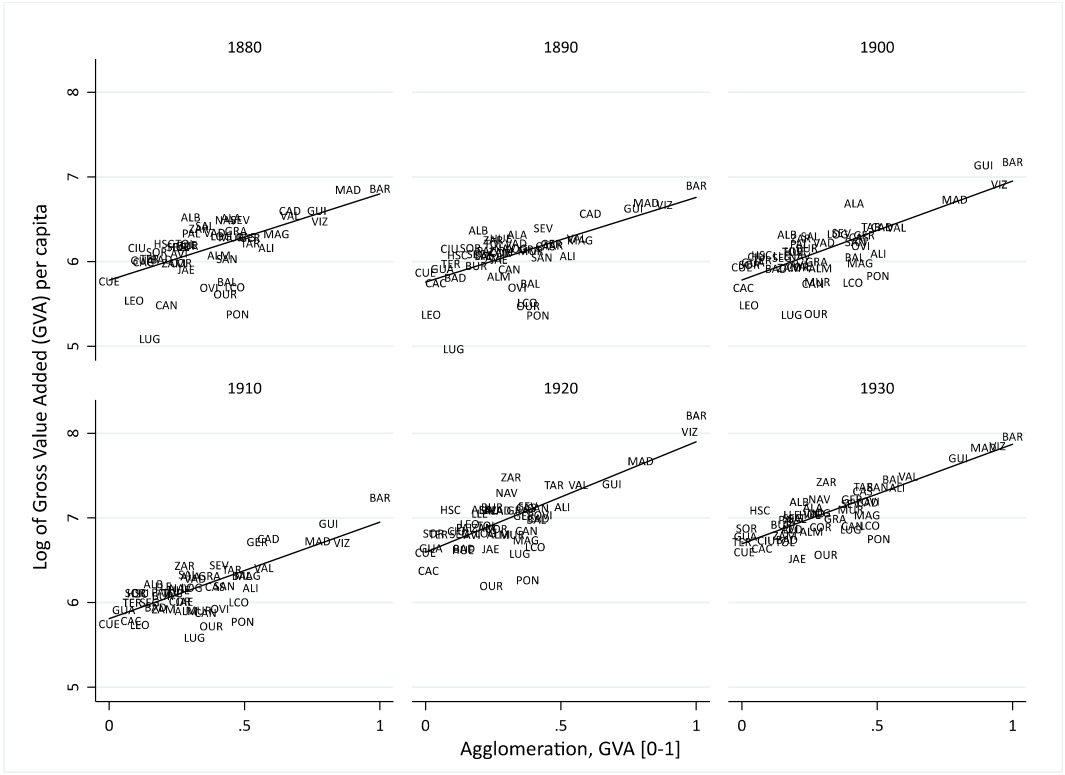


Source: see text. Dots illustrate the Canary Islands, the Balearic Islands and the northern provinces. The southern provinces (Almería, Cádiz, Córdoba, Granada, Huelva, Jaén, Málaga, Sevilla, Badajoz, Cáceres, Murcia, Albacete and Ciudad Real) are shown using a 3-digit code.

Now that our indicators have been presented, an initial approximation can be made of the relationship between agglomeration 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 was greater also registered higher levels of GDP per capita throughout the period.

[4] The two Canary Island provinces are counted as one.

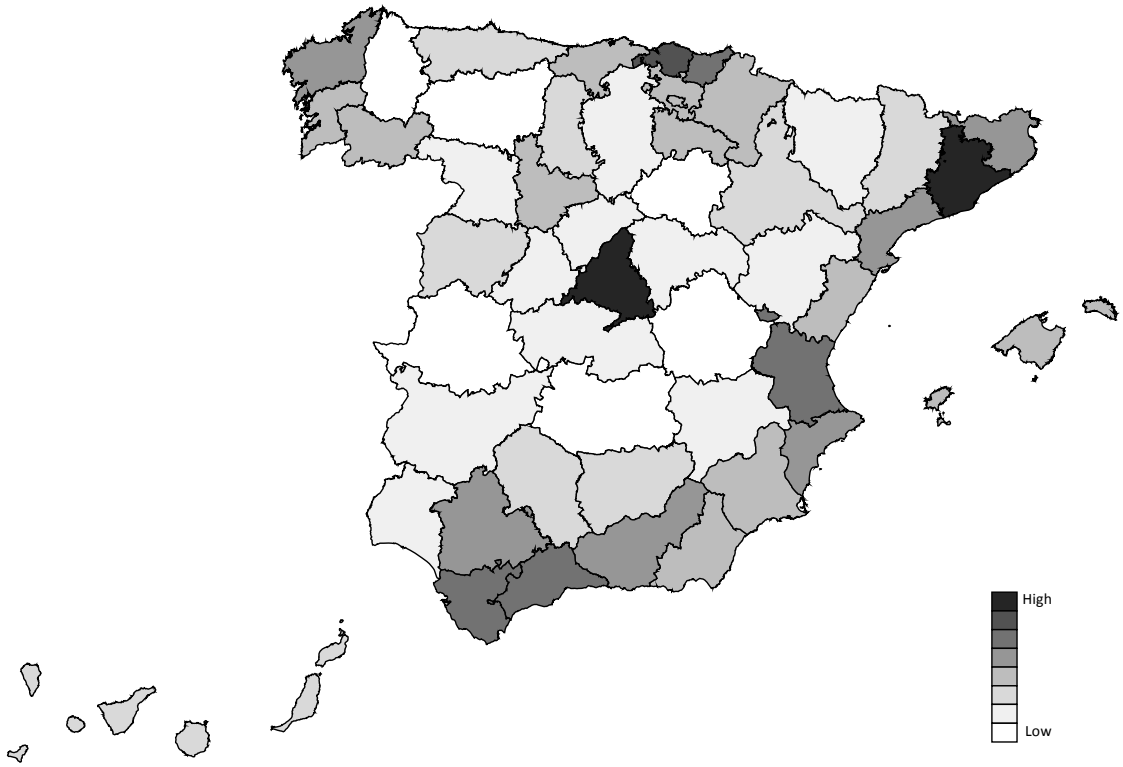
Figure 3. GDP per capita and agglomeration (GVA) in Spain by year, 1870-1930 (NUTS3 provinces)



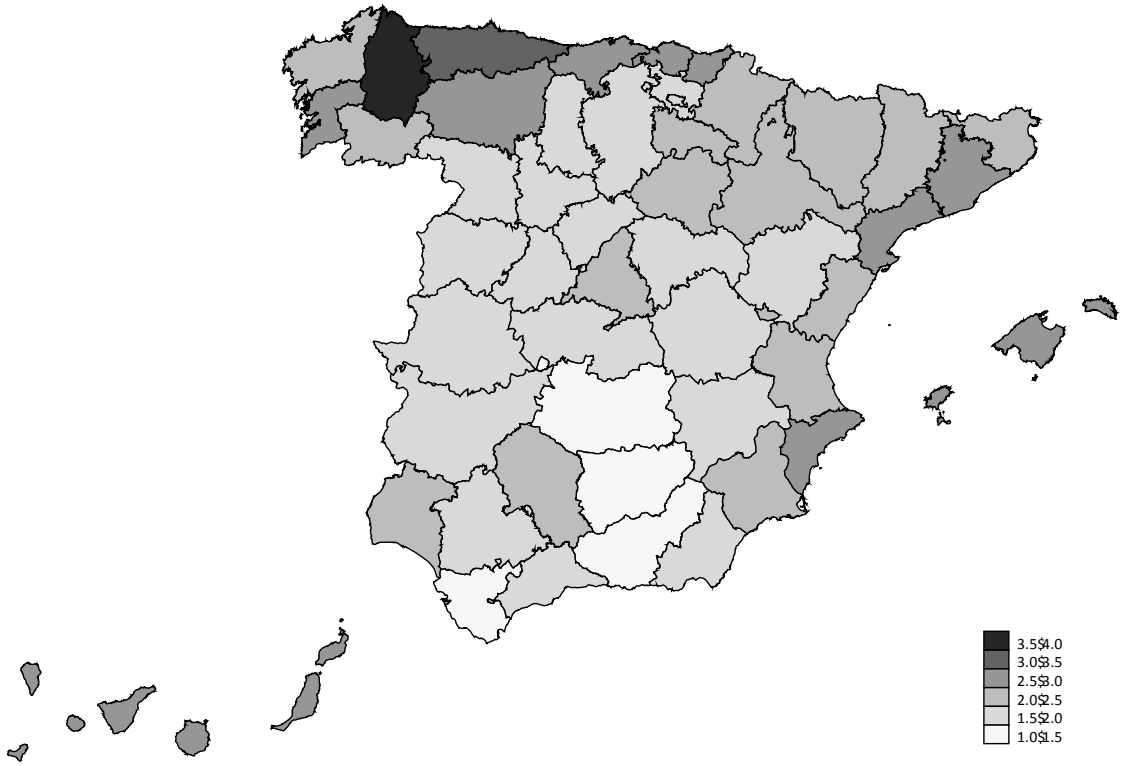
Source: see text.

As for the analysis of the relationship between agglomeration 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 Maps 1 to 3. To begin with, our indicator showing the agglomeration of total GVA in 1870, the first year of the study, can be seen in Map 1. The group of provinces with the greatest agglomeration, headed by Madrid and Barcelona, are mainly to be found along the eastern Mediterranean coast, in the Ebro valley, the north of the peninsula and in eastern Andalusia. With the exception of eastern Andalusia, Map 2 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 obtained when analysing agglomeration in industry (Map 3). 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.

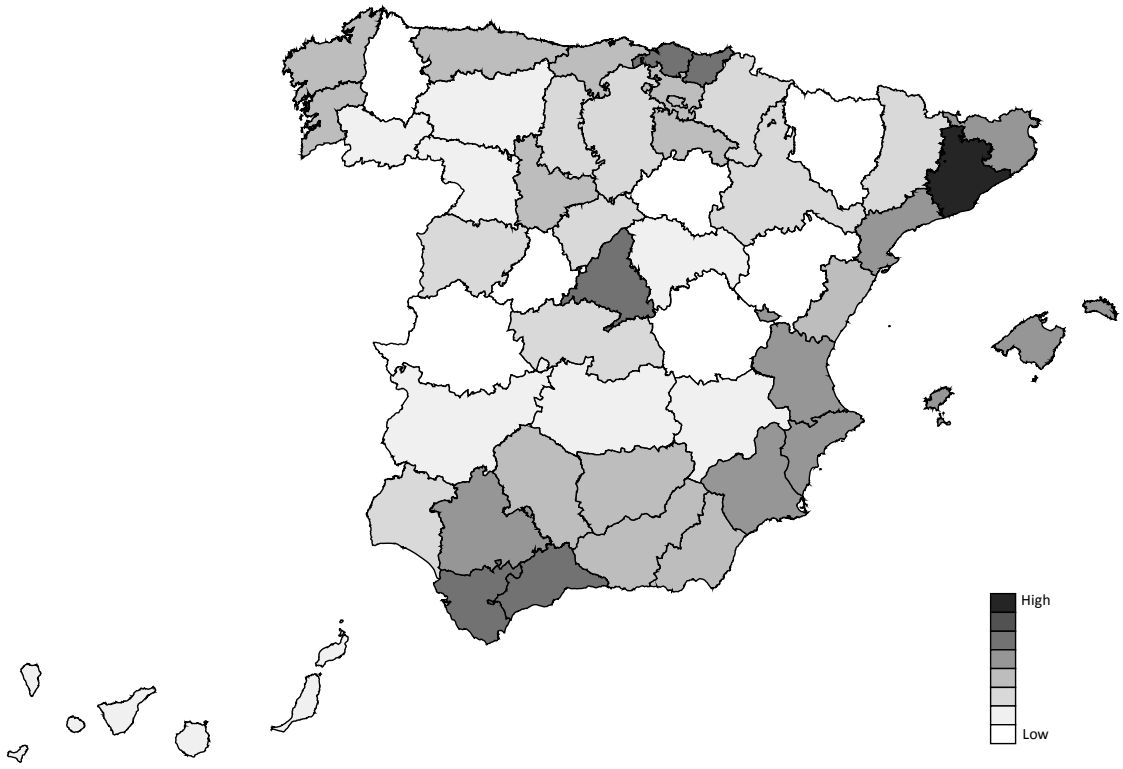
Map 1. Relative agglomeration [0-1] of gross value added (GVA) in Spain, 1870



Map 2. Annual growth rate (%) of nominal GDP per capita in Spain, 1870-1930



Map 3. Relative agglomeration [0-1] of industry gross value added (IGVA) in Spain, 1870



Source: Rosés et al. (2010) and text.

4. Empirical model

Following Brülhart & Sbergami (2009), we set out to empirically evaluate the relationship between agglomeration and economic growth across Spanish provinces (NUTS3) for the period 1870-1930. As Brülhart & 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 & Sala-i-Martin, 1991; Sala-i-Martin et al., 2004):

$$g_{i,p} = \alpha y_{i,t-T} + \beta A_{i,t-T} + \gamma Z_{i,t-T} + \mu_i + v_t + \varepsilon_{i,p} \qquad (3)$$

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

$$g_{i,p} = \left(y_{i,t} - y_{i,t-T}\right)/T, \qquad (4)$$

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, i.e. the 'catch-up' term. $A_{i,t-T}$ captures agglomeration, our variable of interest, within 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 measures of agglomeration for gross value added (GVA) and employment (EMP). Urbanisation rates have also been used as a proxy to test for robustness (Henderson, 2003; Brülhart & Sbergami, 2009). In addition we present measures of agglomeration for GVA by economic activity (agriculture, industry and services) to extend our empirical analysis.

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 capital. These variables are included so as to control for other relevant factors that positively affect the process of regional economic growth. The share of mining (as a percentage of GDP) aims to control for regional differences in natural resources. 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 GMM estimator, as we explain below. Table 3 shows the descriptive statistics of our main variables. Finally μ_i, v_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.

Table 3. Descriptive statistics.

Variables	Mean	Std. Dev.	Max.	Min.	Obs.
GDP per capita					
Annual growth rate (%)	0.021	0.034	0.131	-0.038	294
Level (pesetas)	688	474	3763	95	343
Agglomeration, normalised [0-1]					
Total, 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	1.030	0.160	343
Stock of infrastructure (millions pesetas)	61.8	42.8	384.4	2.2	343
Share of mining (as a % of GDP)	0.013	0.026	0.154	0.000	343

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 instruments for our panel regression recommends the use of the system GMM estimator, proposed by Arellano & Bover (1995) and further developed by Blundell & Bond (1998, 2000)[5].

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. The system GMM estimator provides us with the possibility of estimating a small dynamic panel without the relevant external instruments. 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 have been imposed to lessen the potential problem of instrument proliferation. We report the Hansen J tests for the joint validity of the instruments[6] and have estimated the system GMM estimator with the xtabond2 package for Stata 13 developed by David Roodman (2009b).

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 49 Spanish provinces (NUTS3) over six decades. Columns (2) and (4) show our chosen specification with two proxies for agglomeration: gross value

[5] Arellano & Bond (1991) introduced the generalised methods of moments (GMM) estimator for dynamic panel data models and proved that it offered significant gains in efficiency when compared with conventional instrumental variable (IV) approaches. With the GMM estimator, the endogenous variables in first differences were instrumented with lagged levels. However, Blundell & 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.

[6] 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.

added (GVA) and employment (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 are 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.

Table 4. Agglomeration and GDP per capita growth in Spain, 1870-1930

Variables	**(1)**	**(2)**	**(3)**	**(4)**
GDP per capita	-0.017***	-0.029***	-0.015***	-0.020***
Agglomeration				
GVA	0.017***	0.028***		
Employment			0.019***	0.017***
Literacy rate		0.016***		0.014***
Infrastructure		0.005***		0.005***
Share of mining (GDP)		-0.022		-0.022
Time dummies	NO	YES	NO	YES
Constant	YES	YES	YES	YES
R^2	0.053	0.784	0.060	0.779
Observations	294	294	294	294

Note: 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.

Our two proxies for agglomeration do not fully capture the absolute concentration of economic activity in province i and year t. We therefore need to test our preliminary results. To this end we estimate our preferred specification with urbanisation rates as a proxy for absolute agglomeration. Table 5 shows the pooled OLS regression with our proxy for agglomeration of GVA and levels of urbanisation. Our agglomeration measure seems to perform better. This strengthens our approach because the proxies reflecting economic activity appear to be a more appropriate measure of agglomeration than population-related ones. Moreover, our proxy for agglomeration of GVA allows us to empirically examine the relative concentration of economic activity by main sector (agriculture, industry and services).

Table 5. Agglomeration/Urbanisation and GDP per capita growth in Spain, 1870-1930

Variables	**(1)**	**(2)**
GDP per capita	-0.029***	-0.022***
Agglomeration, GVA	0.028***	
Urbanisation rate		0.012***
Literacy rate	0.016***	0.023***
Stock of infrastructure	0.005***	0.003**
Share of mining (GDP)	-0.022	-0.037
Time dummies	YES	YES
Constant	YES	YES
R^2	0.784	0.770
Observations	294	294

Note: 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.

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 49 Spanish provinces (NUTS3) over six decades, 1870-1930. As stated earlier, all time-dependent variables will be treated as potentially endogenous. However, to mitigate the potential problem of instrument proliferation, given the size of our panel we have instrumented GDP per capita and agglomeration with a maximum of two lags and the rest with just one. Table 6 reports our main results. Columns (1) and (3) show the one-step estimation, while columns (2) and (4) illustrate the two-step estimation. We report both estimation procedures to shed further light on our results[7].

The results are in line with those reported above. First, the relationship between agglomeration and economic growth is empirically supported. The β coefficients are positive and statistically significant when our preferred 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% and 3%, which is in line with the existing literature on regional convergence (Barro & Sala-i-Martin, 1991). Third, the control variables perform well. Literacy is highly significant and has a positive effect on economic growth, while the stock of infrastructures also has a positive

[7] 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.

impact, although it loses its statistical significance. Lastly, the share of mining has a negative but statistically insignificant effect.

Table 6. Agglomeration and GDP per capita growth in Spain, 1870-1930 (System GMM)

Variables	**(1) (One-step)**	**(2) (Two-step)**	**(3) (One-step)**	**(4) (Two-step)**
GDP per capita	-0.031***	-0.029***	-0.022***	-0.022***
Agglomeration				
GVA	0.027***	0.026***		
Employment			0.014*	0.014
Literacy rate	0.024***	0.024**	0.022***	0.022*
Stock of infrastructure	0.003	0.003	0.003	0.004
Share of mining (GDP)	-0.027	-0.015	-0.013	0.019
Time dummies	YES	YES	YES	YES
Constant	YES	YES	YES	YES
Provinces	49	49	49	49
Observations	294	294	294	294
Hansen J test	43.92	43.92	41.49	41.49
	(0.809)	(0.809)	(0.874)	(0.874)
AR (1)	-4.63	-4.32	-4.51	-4.33
	(0.000)	(0.000)	(0.000)	(0.000)
AR (2)	1.97	1.83	1.93	1.82
	(0.049)	(0.067)	(0.053)	(0.069)

Note: 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.

As Table 6 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)[8]. 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 64 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

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

found a strong relationship between agglomeration and economic growth across Spanish provinces during the early stages of industrialisation.

Table 7. Agglomeration and GDP per capita growth in Spain, 1870-1930 (System GMM)

Variables	**(1)** **(One-step)**	**(2)** **(One-step)**	**(3)** **(One-step)**	**(4)** **(One-step)**	**(5)** **(One-step)**
GDP per capita	-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
Provinces	49	49	49	49	49
N	294	294	294	294	294
Hansen J test	43.92	42.79	43.38	41.52	45.80
	(0.809)	(0.841)	(0.824)	(0.873)	(0.999)
AR (1)	-4.63	-4.48	-4.73	-4.64	-4.82
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
AR (2)	1.97	2.01	1.98	1.96	2.06
	(0.049)	(0.044)	(0.048)	(0.050)	(0.039)

Note: 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.

Finally we disaggregate our preferred measure of agglomeration by economic activity (agriculture, industry and services). Table 7 reports the main results using the one-step system GMM estimator[9]. Column (1) shows our preferred dynamic panel data model. Columns (2), (3) and (4) illustrate our chosen specification with a proxy for agglomeration of GVA by economic activity. Column (5) presents the results of the inclusion of the three different proxies for agglomeration by sector. On the whole we found a stable empirical model. Once again conditional convergence ($\alpha < 0$) is supported, with the estimates ranging from 1.5% to 3.1%. This is in line with the existing literature on regional convergence. Literacy and the stock of infrastructures

[9] 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 6 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.

have a positive impact, though the latter losses statistical significance; whereas the share of mining maintains its negative effect on economic growth. As regards our main variable of interest, only 'industrial' agglomeration (GVA) is statistically significant. This supports our prior hypothesis and strengthens the study. Indeed 'agricultural' agglomeration appears to have a negative impact on economic growth. 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 and economic growth across Spanish provinces (NUTS3) for the period 1870-1930 emerges as our main finding. This relationship seems robust and directly related to the presence of agglomeration economies in industry.

6. Conclusions

This article has explored the existence of a relationship between the spatial agglomeration of economic activity and regional economic growth in Spain. Following the methodology proposed by Brülhart & 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 these types of exercise, 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. The exercise was carried out using a data set that for the first time included estimates of regional GDP per capita in Spain for 1870, 1880 and 1890, which complete the data set provided by Rosés et al. (2010).

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 and infrastructures 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, the results support the existence of a positive, robust relationship between the initial levels of regional agglomeration of economic activity (especially industrial activity) and their subsequent growth trajectories.

Therefore the case of Spain, in line with the results obtained by Brülhart & Sbergami (2009), provides evidence to support the existence of a trade-off between inequality in the spatial distribution of economic activity and economic growth in a case study that allows us to avoid some of the limitations identified in the evidence shown in previous papers. The present paper analyses an economy which, during the period studied, was in the early stages of its growth process. Like other European economies of the period, this economy was driven by growth in the industrial sector 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 in its production processes. 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 use 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. One the one hand it is related to the unequal initial provisions of cumulative production factors such as human capital and infrastructures, 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.

7. References

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.

BALDWIN, R and MARTIN, P. (2004): "Agglomeration and regional growth", in J.V. Henderson and J.F. Thisse (ed.), *Handbook of regional and urban economics*, vol.4, chapter 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.

BRÜLHART, M. and SBERGAMI, F. (2009): "Agglomeration and growth: cross-country evidence". *Journal of Urban Economics* 65(1), pp. 48-63.

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.

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.

FUJITA, M. and THISSE, J.F. (2002): "Agglomeration and market interaction". *CEPR Discussion Papers*, 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.

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.

HERRANZ, A. (2007): "Infrastructure investment and Spanish economic growth, 1850-1935". *Explorations in Economic History* 44(3), pp. 452-468.

HIRSCHMAN, A.O. (1958): *The strategy of economic development*. New Haven: 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.

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. (2014): "Market potential and regional economic growth in Spain, 1860-1930". *DT-AEHE*, Nº 1409.

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 XIX. Una perspectiva histórica*. Barcelona: Ariel, pp. 23-61.

OTTAVIANO, G.I.P. and PINELLI, D. (2006): "Market potential and productivity: evidence from Finnish regions". *Regional Science and Urban Economics* 36(5), pp. 636-657.

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.

POLLARD, S. (1981): *Peaceful Conquest. The industrialization of Europe, 1760-1970*. London: Oxford University Press.

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 (dirs.), *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.

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: New York University Press.

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, Universitat de Barcelona, Collecció 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, July, 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.

Ivie

Guardia Civil, 22 - Esc. 2, 1º
46020 Valencia - Spain
Phone: +34 963 190 050
Fax: +34 963 190 055

Website: http://www.ivie.es
E-mail: publicaciones@ivie.es