European
Historical
Economics
Society

EHES WORKING PAPERS IN ECONOMIC HISTORY | NO. 119

The long-term relationship between economic development and regional inequality: South-West Europe, 1860-2010

Alfonso Díez-Minguela[(a)]

Rafael González-Val[(b)]

Julio Martinez-Galarraga[(a)]

M. Teresa Sanchis[(a)]

Daniel A. Tirado[(a)]

[(a)] Universitat de València
[(b)] Universidad de Zaragoza, IEB

DECEMBER 2017

EHES Working Paper | No. 119 |December 2017

The long-term relationship between economic development and regional inequality: South-West Europe, 1860-2010

Alfonso Díez-Minguela,[(a)] Rafael González-Val,[(b)] Julio Martinez-Galarraga, [(a)] M. Teresa Sanchis[(a)] and Daniel A. Tirado [(a)]

[(a)] Universitat de València
[(b)] Universidad de Zaragoza, IEB

Abstract

This paper analyses the long-term relationship between regional inequality and economic development. Our data set includes information on national and regional per-capita GDP for four countries: France, Italy, Portugal and Spain. Data are compiled on a decadal basis for the period 1860-2010, thus enabling the evolution of regional inequalities throughout the whole process of economic development to be examined. Using parametric and semiparametric regressions, our results confirm the rise and fall of regional inequalities over time, i.e. the existence of an inverted-U curve since the early stages of modern economic growth, as the Williamson hypothesis suggests. We also find evidence that, in recent decades, regional inequalities have been on the rise again. As a result, the long-term relationship between national economic development and spatial inequalities describes an elephant-shaped curve.

JEL classification: N9, 018, R1

Keywords: Economic development, regional inequalities, Kuznets curve, Europe, economic history

Corresponding author:
Julio Martinez-Galarraga
Departament d'Anàlisi Econòmica - Universitat de València
Avda. dels tarongers s/n 46022 València (Spain)
Julio.Martinez-Galarraga@uv.es

Notice
The material presented in the EHES Working Paper Series is property of the author(s) and should be quoted as such. The views expressed in this Paper are those of the author(s) and do not necessarily represent the views of the EHES or its members

1. Introduction

Regional income inequality is a cause for concern to both academics and policymakers. In recent decades emerging economies have undergone rapid economic growth which, together with large positive effects, has generated unease regarding the spatial implications of economic integration and growth. As the World Development Report (WDR, 2009, 1) stated, "…*economic growth will be unbalanced, but development still can be inclusive. As economies grow from low to high income, production becomes more concentrated spatially. Some places —cities, coastal areas, and connected countries— are favored by producers*". In this regard, China stands as a paradigmatic case among emerging economies. Unprecedented growth there has brought with it a significant increase in regional disparities between coastal and inland regions, although these now seem to be coming to a halt (Kanbur and Zhang, 2005; Kanbur et al., 2017).

However, spatial inequalities are also a concern in more developed economies. The European Union has introduced numerous territorial cohesion programmes aimed at reducing regional disparities.[1] Nevertheless, as the Eurostat Regional Yearbooks repeatedly remind us, imbalances persist. As regards NUTS2, per-capita GDP in 2014 in the wealthiest region, Inner London (UK), was 5.39 times the EU-28 average and around 18 times that of the poorest region, Severozapaden (Bulgaria). Furthermore, 78 out of 276 European regions had income levels below 75% of the EU-28 average. In addition, the geography of inequality shows a clear core-periphery pattern. While the centre of Europe is characterized by high per-capita incomes, as one moves towards the southern and eastern peripheries of the continent, average regional incomes become gradually lower. The magnitude of these disparities and the apparent lack of policy effectiveness have prompted numerous studies on their evolution and causes.[2]

In explaining growth differentials across countries and regions, the literature points to the major role played by technological progress and knowledge spillovers. If these two engines of growth do not spread evenly across both time and space, the outcome may be divergence (Pritchett, 1997; Lucas, 2000), in which case the convergence process may display a non-linear evolution. In this respect, Williamson (1965) paved the way. Inspired by Kuznets (1955) and his seminal contribution on the dynamics of economic inequality, Williamson suggested that throughout the economic development process regional inequality exhibited an inverted U-shaped pattern.[3] He observed that in the early stages of modern economic growth industrial activity was concentrated in specific

[1] The European Regional Development Fund and the European Committee of the Regions were created in 1975 and 1994 respectively.

[2] See Magrini (2004) or Breinlich et al. (2014) for reviews of the empirical literature.

[3] Williamson (1965) used regional data for the United States and cross-country evidence for the late nineteenth and early twentieth centuries. After World War II regional studies gained ground, but data availability was still a serious limitation. Standing out among the pioneering studies suggesting a potential increase in regional income inequality in the early stages of economic growth are the classic works by Myrdal (1957) and Hirschman (1958).

locations while the rest of the regions remained largely agricultural. This in turn increased per-capita income inequality across regions. However, over the long term these disparities eventually disappeared. Regional economic convergence was thus related to the uneven spread of industrialization. Market integration, capital and, above all, labour flows reinforced and accelerated this process. Specialization and divergence in economic structures would therefore explain the rise in inequality in the early stages of modern economic growth. Further progress and national market integration would then be accompanied by a reduction in regional disparities, which could be explained by the homogenization of economic structures and labour productivity convergence.

For a long time the Williamson hypothesis had no sound theoretical backing. However, Barrios and Strobl (2009), following the Lucas (2000) growth model, recently suggested a theoretical framework for studying the evolution and determinants of regional economic inequality. In their model regional inequality emerges as a result of technological shocks (i.e. industrialization), which are concentrated in specific locations. To be more precise, they assume that, due to region-specific factor endowments and/or institutions, technological shocks only occur in some particular regions. These *lucky* regions will initially grow more rapidly, and spatial inequality will thus arise. Other regions will eventually catch up, either through the spread of technology or as a result of factor flows (capital, labour). Given these circumstances, Barrios and Strobl (2009) provide a theoretical foundation for the Williamson hypothesis in line with classic theorizing contributions on regional economic growth, such as those arising from neo-classical growth theory (Barro and Sala-i-Martin, 1991; Barro et al., 1995).[4]

A small but slowly growing body of empirical works have tested the existence of a Kuznets curve in spatial inequalities. Importantly, due to limitations in the availability of historical data, studies have focused on recent decades. Two works in particular stand out. Barrios and Strobl (2009) have tested the Williamson hypothesis econometrically using parametric and semiparametric techniques, relying on regional data for a sample of developed economies between 1975 and 2000. The authors find strong evidence in support of an inverted U-shaped curve in the relationship between regional inequality and economic development. In a similar vein with the aim of accounting for different levels of development, Lessmann (2014) uses panel data covering 56 developed and emerging economies between 1980 and 2009 and confirms the existence of an inverted U-shaped pattern. He also shows that spatial inequalities increase again at high levels of economic development.

These studies examine the evolution and causes of regional inequality with cross-sectional or panel data since 1975-80, but three decades can hardly capture deep structural changes. As Lessmann (2014, 36) himself recognizes: "*the major problem for this kind of research is that it is essential either to have historical data for single countries or to include poorer countries in a cross-country data set, since the theories of*

[4] Barrios and Strobl (2009) also provide an economic foundation for the relative growth dynamics identified by the technological gap and catching-up literature (Abramovitz, 1994).

Kuznets and Williamson point at the deep structural changes associated with the industrialization process". A more suitable way of approaching the topic would therefore be to study the evolution of regional inequality during the whole process of economic development, i.e. examining the long-term dynamics of countries since the early stages of modern economic growth. This would make it possible to account for the transition from agrarian to industrial and service-based economies. Industrialization processes in Europe began during the nineteenth century when the novel technologies of the first and second industrial revolutions spread across the continent. As economic historians have already documented, the processes of industrialization were characterized by their regional nature (Pollard, 1981). Technological shocks, industrialization and structural change did not arrive in all regions at the same time and therefore not all regions became richer at the same time, and this led to an initial upswing in regional disparities, as suggested by Williamson.

The main contribution of the present study is to conduct an analysis of the evolution and determinants of regional income inequality throughout the historical process of economic development in four south-west European countries: France, Italy, Portugal and Spain. Our data set, which borrows from works of economic history, includes information on regional per-capita GDP for these four countries on a decadal basis beginning in 1860 and ending in 2010. This is, to our knowledge, the first empirical study of the evolution of regional inequality to cover such a long time span and thus the first to capture the whole process of economic development from the early stages of modern economic growth. This empirical strategy seems more suitable than using a panel data set of countries at different stages of development. While that strategy enables an examination of the determinants of regional economic inequality based on information for recent decades, it lacks a wide time perspective and thus makes a long-term interpretation of the results difficult. All in all, our study seeks to overcome this time limitation, which is present in all the existing contributions. Additionally, following Barrios and Strobl (2009) and Lessmann (2014), it makes it possible to carry out an empirical strategy based on parametric and semiparametric techniques in historical perspective and to interpret our results in the light of a sound economic foundation.

Our study offers some appealing results. In particular, it shows that during the historical process of economic development in the four nations that make up South-West Europe, regional income inequality displays a U-shaped evolution. Interestingly, in more recent stages of development the trend is seen to be increasing again. This result is obtained from parametric regressions with polynomial functions for income, and also when we use a semiparametric approach. Moreover, it is robust to the inclusion of confounding factors that may have a (changing) effect (over time) on regional inequality, such as trade openness, the presence of agglomeration economies and/or the role of public policies. Our findings are also robust to the consideration of alternative inequality measures. They show that economic growth has been more intense in the most populated regions, which, in the context of these four countries of South-West Europe, are those in which the capital

cities are located. In particular, the dynamism of the regions containing the capital cities would explain the recent upsurge in regional inequalities. Finally, in line with Kuznets (1955) and Williamson (1965), we identify structural change as a significant transmission mechanism of the inverted-U relationship found between economic development and regional inequality in the long run.

The paper is structured as follows. Section 2 presents the theoretical foundation supporting the empirical analysis and introduces the main empirical contributions used to analyse the link between economic development and regional inequality. Section 3 details our data set on regional GDP and population for the period 1860-2010 and adds some descriptive evidence. Section 4 introduces the empirical strategy, presents the main results of our analysis and provides a series of robustness checks. Finally, Section 5 contains our conclusions.

2.- Economic foundation and empirical contributions

As noted earlier, the inverted U-shaped relationship between economic development and regional inequalities suggested by Williamson (1965) was not backed up by a theoretical framework. In order to provide a sound economic foundation to this popular hypothesis, Barrios and Strobl (2009) developed a model to give support to the so-called Williamson hypothesis. They built on the growth model based on technology diffusion developed by Lucas (2000). This setting makes it possible to describe the transition from pre-industrial stagnant economies to modern economies over the course of the long-term economic growth process. The dynamics of this transition rely on technological shocks or innovations that, beginning during the onset of industrialization, push productivity levels up and accelerate growth at a national level.

These technological shocks are, however, initially concentrated in a particular region due to region-specific factor endowments (e.g. mineral wealth, human capital) and/or institutions. Thus economic growth is also spatially localized, and consequently in the early stages of economic growth and industrialization there tends to be an increase in regional inequalities. While initially the forces of divergence predominate, at some point there occurs a process of technology diffusion from the leading region to other regions within the country. The pace and timing of this diffusion depends on the technological capabilities of other regions and country-level knowledge. Follower regions, each at a different point in time, gradually adopt the innovations of the leading region and shift from stagnation to a growth regime. Regional convergence then begins. Eventually other regions will also catch up either due to the spread of technology or as a result of factor flows (capital, labour). Under these circumstances, the relationship between national average per-capita income and regional inequalities becomes non-monotonic, depicting an inverted-U or bell-shaped

curve over time. This model thus provides a theoretical framework in line with the Williamson hypothesis.

The work by Barrios and Strobl (2009) also allows certain issues to be inferred relating to the forces that may affect regional income inequality during the economic development process. These authors suggest that spatial inequalities would be greater if national markets were globally integrated. This points to the presence of a positive relationship between trade and regional economic inequality within a given country. Several studies have examined this particular link. In most cases studies go back to 1980 and therefore analyse recent decades, but the evidence points to there being a clear effect of trade on higher regional inequalities.[5] Rodriguez-Pose and Gill (2006) and Hirte and Lessman (2014) analyse the effect of trade openness on regional inequality, including both developed and developing economies in their sample.[6] The evidence provided in these studies confirms that the effect exists. In addition, Rodriguez-Pose (2012) shows that this effect is stronger in poorer countries. Likewise Ezcurra and Rodriguez-Pose (2014), who focus on 22 emerging economies between 1990 and 2006, again find that changes in international trade have brought about a significant increase in interregional inequalities and that the impact has been bigger in poorer countries. Consequently a greater exposure to trade, while possibly benefiting these emerging economies in aggregate terms, generates winning and losing regions.

If, however, economic growth is considered endogenous and there are economies of scale in the production of technology, we would expect public policies to channel funds towards education and research and development (Baldwin and Martin, 2004). The design of networks and reductions in transport costs would become more relevant. In this respect, new economic geography models provide a strong foundation regarding the effect of declining transport costs and changes in market accessibility on the geographical location of economic activity (Krugman, 1991; Puga, 1999; Fujita and Thisse, 2002). Hence the presence of agglomeration economies might also hinder the process of regional income per capita convergence.[7] Ciccone and Hall (1996) and Ciccone (2002) identify an agglomeration effect linking the density of economic activity with inter-regional differences in labour productivity, thereby establishing a link between economic density and agglomeration effects. Also, Brülhart and Sbergami (2009) carry out a Barro-style empirical analysis of the determinants of economic growth for both a large sample of countries around the world and

[5] The exception would be Milanovic (2005), who analyses the five most populated countries in the world (China, India, the United States, Indonesia and Brazil) between roughly 1980 and 2000, but finds no link between trade openness and regional inequalities.

[6] Hirte and Lessman (2014) estimate a model derived from a structural economic geography approach in which interregional inequality depends on weighted trade shares and trade costs with a sample of 208 countries for 1948-2006. IV and dynamic panel regressions show that trade increases interregional inequalities but that an increase in openness is probably neutral. If openness were a proxy for global integration, they conclude that more integration would neutralize the negative interregional distribution effect.

[7] In a somewhat related matter, NEG models provide ambiguous results as regards the impact of trade openness on the internal economic geography of countries (Krugman and Livas, 1995; Paluzie, 2001; Crozet and Koenig, 2004; Hanson, 2005). A survey of this literature can be found in Brülhart (2011).

another sample restricted to EU countries over the period 1960-2000. Alongside the explanatory variables traditionally included in this type of exercise, the authors introduce different indicators for the spatial agglomeration of production and population. 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, this relationship disappears or becomes negative. In this respect they point out that their empirical analysis provides evidence in support of the Williamson hypothesis.[8]

Another relevant dimension is the pace of change along the inverted U-shaped curve. This is an aspect that has also been considered by Barrios and Strobl (2009). As modern economic growth spread across regions, region-specific factor endowments (physical and human capital) and public policies gained momentum. The latter were aimed at reducing regional disparities in an attempt to maintain territorial cohesion, for instance by funding the construction of infrastructures. Since 1986 the European Union has prioritized and allocated resources to regional policies aimed at strengthening economic and social cohesion. With the Lisbon Treaty came a new *high-level strategy* (Europe 2020) which introduced the term territorial cohesion.

Two more elements emanating from Barrios and Strobl (2009) also deserve comment. Firstly, if technological change is concentrated in specific sectors and therefore locations, structural differences could increase the extent of regional income inequality. In other words, we might witness a direct relationship between regional production specialization and spatial income inequalities. This is particularly relevant as it is one of the main mechanisms discussed in the works of Kuznets (1955) and Williamson (1965). And secondly, the model allows for technological shocks (ICT revolution) and major institutional changes (supranational market integration), and these might have a relevant impact on the long-term dynamics of regional inequality. If, for example, the ICT revolution or greater market integration benefits the richest regions most, a rise in per-capita GDP inequality across regions is to be expected. Gianneti (2002) analyses the European experience in which, since the early 1980s, convergence among countries has not been accompanied by convergence among regions. She argues that sectoral specialization may help to explain this outcome. Economic integration accelerates growth at country level but increases within-country spatial disparities given that regions that are specialized in high-tech sectors benefit more from knowledge spillovers, whereas the opposite happens in regions that are specialized in more traditional sectors.

[8] Gardiner et al. (2011) questioned the results obtained by Brülhart and Sbergami (2009). They explore the relationship between agglomeration and growth in the EU-15 (1981-2007) 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 also point out that the limited period studied (1960-2000) reduces the robustness of the results.

Within this general framework there is a growing empirical literature aimed at testing the central hypothesis deriving from Williamson's (1965) seminal work and subsequent approaches such as Barrios and Strobl (2009). Indeed Barrios and Strobl (2009) themselves empirically tested their model employing parametric and semiparametric techniques to confirm the presence of an inverted U-shaped relationship between regional inequality and development. They used a panel of European countries covering the period 1975-2000. Overall their findings provide strong evidence of a bell-shaped curve; regional inequalities initially increase and, after a certain threshold is reached, convergence begins. The results are robust to changes in regional administrative units and time periods, as well as to the inclusion of control variables and non-European countries.

Lessmann (2014) also studies the presence of an inverted U-shaped relationship between spatial inequality and economic development using a panel data set of spatial inequalities in 55 countries at different stages of economic development, covering the period 1980-2009. The author supplies strong support for an inverted U-shaped relationship but, importantly, he also indicates that regional inequality has recently been rising. This would suggest the existence of an N-shaped relationship between spatial inequalities and economic development. Lessmann and Seidel (2017) conduct a similar exercise but expand the sample of countries by using regional incomes based on night-time light satellite data.[9] This allows them to cover 180 countries for the period 1992-2012. They again find evidence in favour of the Williamson hypothesis. When they analyse the determinants of regional inequality, their results lead them to suggest that trade openness increases regional inequality but that this effect can be counterbalanced by regional transfers, infrastructure investments and more democratic institutions. Also, importantly, at high levels of income there is a high degree of dispersion in regional incomes, which, as in Lessmann (2014), points to the presence of an N-shaped evolution. These works argue that, although there are many potential explanations for the presence of an N-shaped relationship between regional inequality and development, tertiarization or the structural shift from industrial production towards a service base in the highest-developed economies appears to predominate.

The N-shaped hypothesis is in line with previous findings in Amos (1988), which focused on the United States between 1950 and 1983 and showed that the increase-decrease pattern in the evolution of regional disparities suggested by Williamson had turned into an increase-decrease-increase pattern, thereby supporting the hypothesis that regional inequality increases in later stages of development.[10] He attributed this new increase in regional inequality after completion of the Williamson curve to changes such as suburbanization and the transition to a service-based

[9] This methodology follows Gennaioli et al. (2014), who, working with 83 countries over recent decades and depending on data availability, find evidence of regional convergence at a rate close to the growth literature's canonical 2%.
[10] He finds that this was the case in 37 of the 50 US states by the early 1980s.

economy.[11] However, Ezcurra and Rapún (2006) questioned these results in a study of 14 Western European countries in 1980-2002. Using nonparametric methodology and controlling through the inclusion of additional variables in the spirit of Barrios and Strobl (2005), these authors concluded that territorial imbalances stabilize in the later stages of development.

The relationship between regional inequality and development has also attracted the interest of economic historians. Several studies in recent years have provided new regional GDP estimates at country level, enabling examination of the long-term evolution of regional inequality. Some authors have provided further evidence of the existence of an inverted-U pattern. Using new regional per-capita GDP estimates for the UK, Crafts (2005) found support for this hypothesis. Regional inequality increased after 1871, reached its highest point in the early 20th century and then declined until the 1970s.[12] Combes et al. (2011) explored the long-term evolution of spatial inequalities across French *départements* and observed an inverted U-shaped pattern in manufacturing and services between 1860, 1930 and 2000.[13] In Spain there was an upswing in regional inequality in the second half of the nineteenth century that came to an end in the early decades of the twentieth. From then on, a convergence process began that lasted until the 1980s (Martinez-Galarraga et al., 2015). For Portugal, Badia-Miró et al. (2012) provide empirical evidence in support of an inverted U-shaped pattern, noting that regional inequality reached its peak during the 1970s.

Other studies, however, do not fully support an inverted U-shaped pattern over the long term. In their pioneering work, Barro and Sala-i-Martin (1991) showed that the dispersion of income per capita across states in the US decreased steadily between 1880 and the 1970s, with the exception of the 1920s. Additionally, they pointed out that the reduction in regional inequality came to a halt in the 1980s (see also Kim, 1998; Mitchener and McLean, 1999; Caselli and Coleman, 2001; Kim and Margo, 2004). Felice (2011) found that regional disparities in Italy reached a peak in the aftermath of the Second World War.[14] Since then the dynamics of regional inequality have followed a pattern of convergence between the northern and central regions, but not between these and the *Meridionale*, thereby generating a striking north-south divide. For Sweden, Enflo et al. (2014) observe a strong regional convergence from the mid-nineteenth century up to the 1980s, and an upswing afterwards. Lastly, Buyst (2010, 2011) points out a reversal of fortunes between north and south in Belgium during the twentieth century.

[11] See also List and Gallet (1999).

[12] Geary and Stark (2015, 2016) have questioned these results and argue that disparities were diminishing before the First World War. Previous work by these authors (Geary and Stark, 2002) proposed a methodology for estimating regional GDP that has become the reference for other authors/countries.

[13] Recent studies, however, show a more complex view when per-capita GDP is used. While Caruana-Galizia (2013) finds an increase in regional income inequalities between 1840 and 1911, Bazot (2014), Díez-Minguela et al. (2016) and Rosés and Sanchis (*forthcoming*) conclude that the downward trend of regional per-capita income inequality started at the turn of the century. For a critical review of the different estimations available for France, see Díez-Minguela and Sanchis (2017).

[14] A similar long-term evolution has been found by Daniele and Malanima (2014), although these authors suggest some differences in regional GDP estimates at the time of the post-unification period.

All in all, the existing empirical literature usually follows parametric and nonparametric methodologies to examine the evolution of regional inequality over time and to study its main drivers. It is therefore possible to empirically test whether or not regional productive specialization, openness, public expenditure, urbanization rates, market potential or economic agglomeration, among other things, stimulate regional convergence/divergence. However, these empirical papers are restricted to recent decades and thus fall short of providing a long-term perspective. Also, the analysis of the relationship between economic growth and regional inequalities should be conducted considering the entire economic development process, i.e. from the early stages of modern economic growth, as in Kuznets (1955) and Williamson (1965). In addition, we have also seen that research into economic history often focuses on describing the long-term patterns (and sometimes the main drivers) of regional income inequality for individual countries, but does not empirically test for the existence of an inverted-U curve. In short, this article aims to bridge the gap between these different approaches with an analysis supported by a theoretical model, following the empirical methodology employed in the papers reviewed in this section and using data generated by economic historians. Hence we can contribute to the existing literature on regional studies, economic history and development. Our data set, which contains long-term regional per-capita GDP estimates for France, Italy, Portugal and Spain going back to the nineteenth century, is described in the next section.

3. Data and descriptive evidence

3.1. The data set

Recent developments in economic history make it possible to study the long-term evolution of regional income inequality. In this paper we have collected figures for regional population and regional gross domestic product (GDP) for France, Italy, Portugal and Spain at the NUTS2 level, on a decadal basis between 1860 and 2010.[15] There is mixed coverage among the countries in our sample. With a potential maximum of 16 benchmark years per country (16 for Spain, 15 for Italy, 13 for Portugal and 12 for France), we have compiled an unbalanced panel data set.[16] To evaluate national and regional income we use per-capita GDP in a common unit: 1990 Geary-Khamis US dollars.[17] Once we have the national average incomes (Maddison Project Database), we use historical regional GDP estimates to provide the share of national GDP for each NUTS2 region.

[15] See statistical appendix.

[16] For Spain, the sample covers all benchmarks. For the other countries, due to limitations in the estimates, the number of years available decreases. As regards Italy, there is no data for 1860 previous to unification. For Portugal estimates begin in 1890, while in the case of France the two decades between 1870 and 1890 are missing, as is the year 1940, in the middle of World War II.

[17] Per-capita GDPs for France, Italy, Portugal and Spain in 1990 Geary-Khamis US$ come from the Maddison Project Database (Bolt and Van Zanden, 2014). Population at country level for 1860-1990 comes from the original Maddison database, and for 1990-2010 from the Total Economy Database.

Together with the information on regional population available from the Population Censuses, we compute regional per-capita income. Although this method allows us to make interregional comparisons across countries, it assumes no variation in regional prices. In their absence we partially tackle the worrisome matter of differences in the cost of living across regions by employing national deflators. [18]

Our data set on regional per-capita GDP borrows from previous research. In particular, for France we take data from Combes et al. (2011) Díez-Minguela and Sanchis (2017) and Rosés and Sanchis (*forthcoming*); for Italy data is obtained from Felice (2011, 2015); for Portugal Badia-Miró et al. (2012) provide regional GDP estimates, while for Spain we use the data set from Martinez-Galarraga et al. (2015) and Tirado et al. (2016).[19] Information on the current NUTS2 regions is obtained for the four countries: France (22 regions), Italy (20), Portugal (5) and Spain (17). In short, our data set includes 65 regions in four countries that cover a total area of 1,543,265 square km, i.e. more than 35% of today's EU-28. In terms of people, our sample ranges from a population of 83 million in 1860 to 180 million in 2010. That is around 35% of the total EU-28 population at both dates and over 25% of the EU-28's present-day GDP. In addition, our data set includes regions that belong to the economic core of Europe (Île-de-France, Northern Italy) and to the economic periphery of the southernmost parts of Europe.[20]

3.2. Long-term national and regional economic development

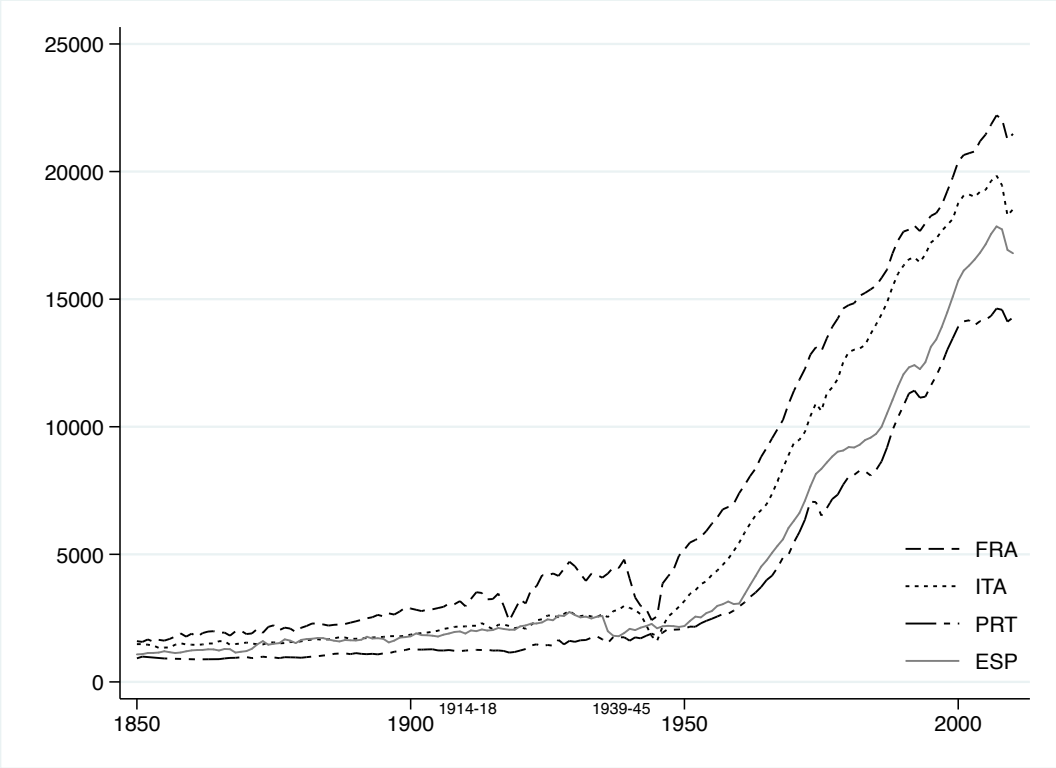
To begin with, a general view of economic evolution in our area of study can be explored by looking at the long-term dynamics of the individual countries. Figure 1 shows the evolution of per-capita GDP in the four countries that make up South-West Europe. Each country has indeed followed a different growth path over time. France stands out for its higher per-capita income since the earliest stages of development. It was an early follower of the British Industrial Revolution and its advanced position with respect to other South-West European countries intensified during the *Belle Époque* in the late nineteenth and early twentieth centuries. Despite the negative impact of the Great War, France's lead increased in the interwar years and it has remained ahead throughout the economic expansion of the *Golden Age* (1945-1973) and over recent decades.

[18] In order to account for price differences, Eurostat created the purchasing power standard (PPS). This is constructed by comparing price levels of a representative basket of goods and services across countries. However, within-country price dispersion is not accounted for. In other words, the price level for Spain is used for the capital-region of Madrid and for all the other regions.

[19] We would like to thank Marc Badia-Miró, Emanuele Felice, Jordi Guilera, Pedro Lains and Joan R. Rosés for kindly sharing unpublished data with us.

[20] South-West Europe has come to be seen as an economic area in its own right as shown by its being taken into account in current EU development policy and in historical studies such as the Cambridge Economic History of Modern Europe (Broadberry and O'Rourke, 2010).

Fig. 1 Per-capita GDP by country (1990 Geary-Khamis US$)



Source: Maddison Project Database, Bolt and Van Zanden (2014).

As regards Italy and Spain, which in the context of Western Europe were slow to industrialize, both follow a similar trajectory up to World War II. In Italy growth accelerated during the so-called '*Giolitti Age*' (1901-1913) (Felice and Vecchi, 2015), while Spanish economic growth did not reach its full potential until the twentieth century, particularly after the Great War (Prados de la Escosura, 2003, 2016). While Spain's per-capita GDP growth stagnated during the Civil War (1936-39) and the period of autarky that followed during the early decades of the Franco regime (1940s-1950s), Italy's growth was more intense in the aftermath of World War II. Likewise, in Portugal economic growth before the Great War was somewhat poor and only began to increase slightly in the interwar years (Lains, 2003). Over the second half of the twentieth century Portuguese growth resembled that of Spain, although in recent decades a new gap has opened between these two Iberian countries. The long-term evolution shows the different paths followed by each country during the process of modern economic growth and the transition from agrarian economies to modern societies. As these country-specific histories have already been extensively documented in the literature, we next concentrate on the regional dimension.

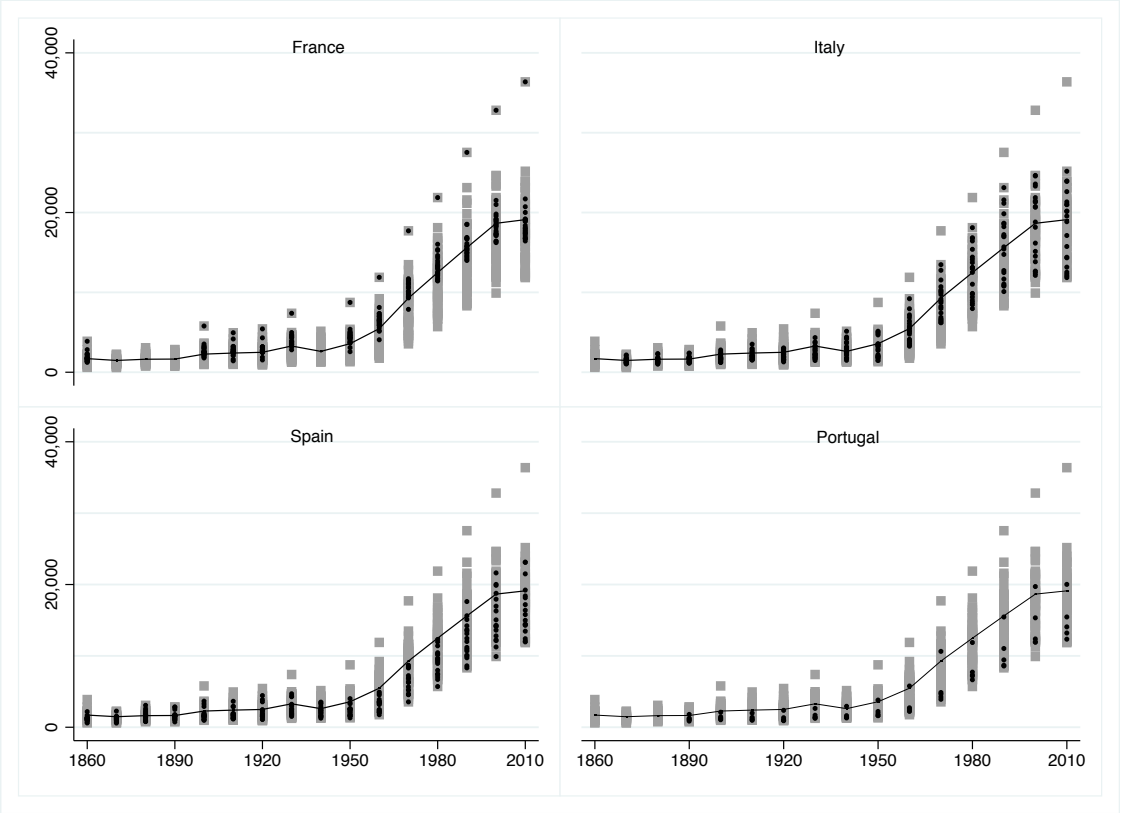
Figure 2 shows the regional per-capita GDP in each country (dark-coloured dots) compared to the average per-capita GDP for South-West Europe as a whole (solid line) for 1860-2010. The grey bars represent all the regions of the four countries taken as a whole. As can be seen, the graphs provide a more complex picture than the one described previously at the national level. Indeed

there are big differences within countries, and regions can be found in all of them with levels of income both above and below the average for South-West Europe.

In France, for example, which is the richest country, most regions had values above the South-West European average until the 1980s. Since then, however, the number of regions above that average has decreased, to the point where in 2010 only four regions exceeded it. Hence a numerous group of French regions is currently below the South-West European average, although these are still far from the bottom of the distribution of South-West European regions represented by the grey bars. In addition, the increasing lead position of the Île-de-France is noteworthy. As regards Italy's regions, throughout the period of study they are fairly equally distributed above and below the solid line that marks the South-West average, showing the traditional North-South divide. By 2010, the richest Italian regions were among the richest of South-West Europe (though far behind the Île-de-France), while the poorest were among the poorest (grey bars).

As in the case of Italy, up to the 1940s in Spain there are regions both above and below the South-West European average. Spanish regions then fall below that average and occupy the lowest positions among the regions of South-West Europe. In the twenty-first century only three regions (Madrid, the Basque Country and Navarre) enjoy a per-capita GDP above the South-West European average. Finally, Portuguese regions, with the exception of Lisbon since 1940, repeatedly fall below the average per-capita GDP for South-West Europe and, together with some Spanish regions, occupy the lowest segments of the South-West Europe distribution marked by the grey bars.

Fig. 2. Regional per-capita GDP (1990 $ Geary-Khamis) by country, 1860-2010



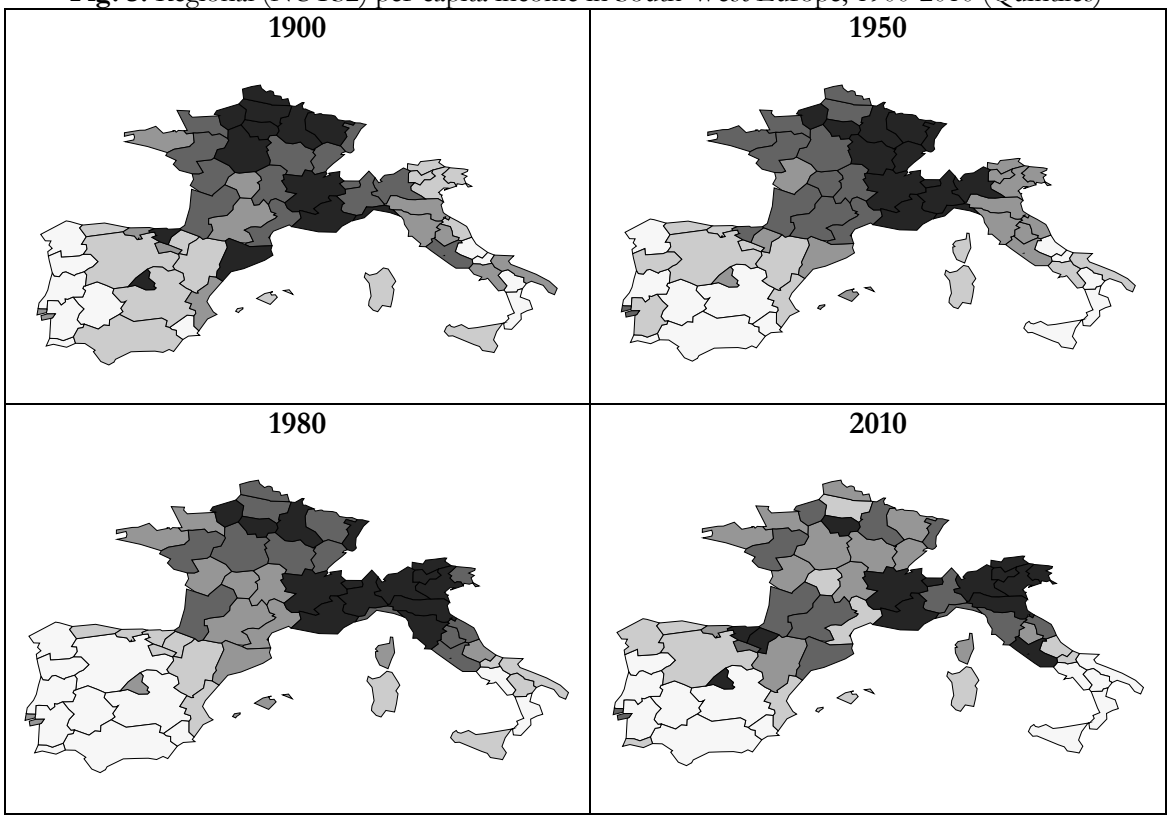
Source: main text. Note: Each black dot represents a region, while the solid line is the average per-capita GDP in South-West Europe as a whole. South-West European regions taken as a whole are in grey.

In addition, relevant changes to the economic geography of South-West Europe may have occurred during the period of study. To explore the main geographical patterns in the spatial distribution of income, Figure 3 shows maps of regional per-capita GDP. The regions are grouped in quintiles for 1900, 1950, 1980 and 2010.[21] Black indicates "very rich", while the lightest grey indicates "very poor". By 1950, most of the southern regions of Italy and the vast majority of Spanish and Portuguese regions were at the bottom of the income distribution. In both Spain and Italy there had emerged a clear north-south divide pattern, while in Portugal only Lisbon had respectable income levels. Furthermore, the rich regions were clustered in a gradient centred on the north of France around Paris (Île-de-France) and northern Italy. In short, it seems that a core-periphery pattern already existed in 1950. In recent decades, however, as the map for 2010 shows, the high-income cluster stretching from northern Italy to northern France has weakened, with only the Île-de-France remaining among the high per-capita GDP regions in the north of France.[22] A capital-city effect also appears to be clearly visible. Figures 2 and 3 show that differences between the richest and poorest regions have tended to grow over time, pointing to an increase in the dispersion

[21] The unbalanced nature of our panel makes 1900 the first year with data for all four countries in our sample.
[22] By this date, as noted in Figure 2, the Spanish regions of the Basque Country and Navarre have joined the club of wealthy regions.

of regional per-capita income. To shed further light on this issue, we next focus on the evolution of regional disparities within countries.

Fig. 3. Regional (NUTS2) per-capita income in South-West Europe, 1900-2010 (Quintiles)



Source: main text. Note: The darkest colour corresponds to the top quintile.

3.3. Long-term national per-capita income and regional inequality within countries

While the level of development in each country and region is captured by the GDP per capita expressed as purchasing power parity (PPP) in 1990 Geary-Khamis US$, we measure regional inequality using the population-weighted coefficient of variation (or the Williamson coefficient of variation - WCV), which can be computed as follows[23]:

$$WCV = \sqrt{\sum_{i=1}^{n}\left(\frac{y_i}{y_m} - 1\right)^2 \cdot \frac{p_i}{p_m}} \qquad (1)$$

where y and p stand for per-capita GDP and population, while i and m refer to regional and national values respectively. The WCV ranges from 0 to 1, from low to high inequality.

[23] As in Williamson (1965), Ezcurra and Rapún (2006), Rodriguez-Pose (2012) and Lessmann (2014). Later on we also conduct our analysis employing alternative inequality measures such as the single coefficient of variation and the Gini Index, as described in the robustness section below.

Table 1 shows the long-term evolution of spatial disparities in these four countries of South-West Europe. There seems to be a general trend towards increasing regional inequalities in the early stages of modern economic growth. While convergence does at some point begin, the timing of the reversal from divergence to convergence varies across countries. In France and Spain there are signs of regional convergence from the early twentieth century, and after that time regional disparities gradually decrease.[24] In Italy and Portugal, after the initial rise in regional income inequality, the process of convergence begins later, in the 1950s in the case of Italy and in the 1970s in Portugal. By 1980 the inverted U seems to be complete in all four countries - over the last three decades regional catching-up has come to an end or even reversed (with the exception of Portugal). Here, the case of France stands out for the sharp increase in regional inequality it experienced from 1980 to 2010.

Table 1. Regional inequality (WCV)

	1860	**1870**	**1880**	**1890**	**1900**	**1910**	**1920**	**1930**	**1940**	**1950**	**1960**	**1970**	**1980**	**1990**	**2000**	**2010**
France	0.35				0.37	0.30	0.32	0.30		0.33	0.31	0.25	0.23	0.26	0.27	0.32
Italy		0.17	0.19	0.17	0.20	0.20	0.23	0.26	0.30	0.36	0.29	0.24	0.25	0.25	0.25	0.25
Portugal				0.21	0.23	0.25	0.38	0.28	0.34	0.36	0.41	0.43	0.26	0.27	0.24	0.21
Spain	0.30	0.29	0.32	0.34	0.42	0.39	0.48	0.40	0.35	0.34	0.36	0.27	0.22	0.22	0.24	0.24

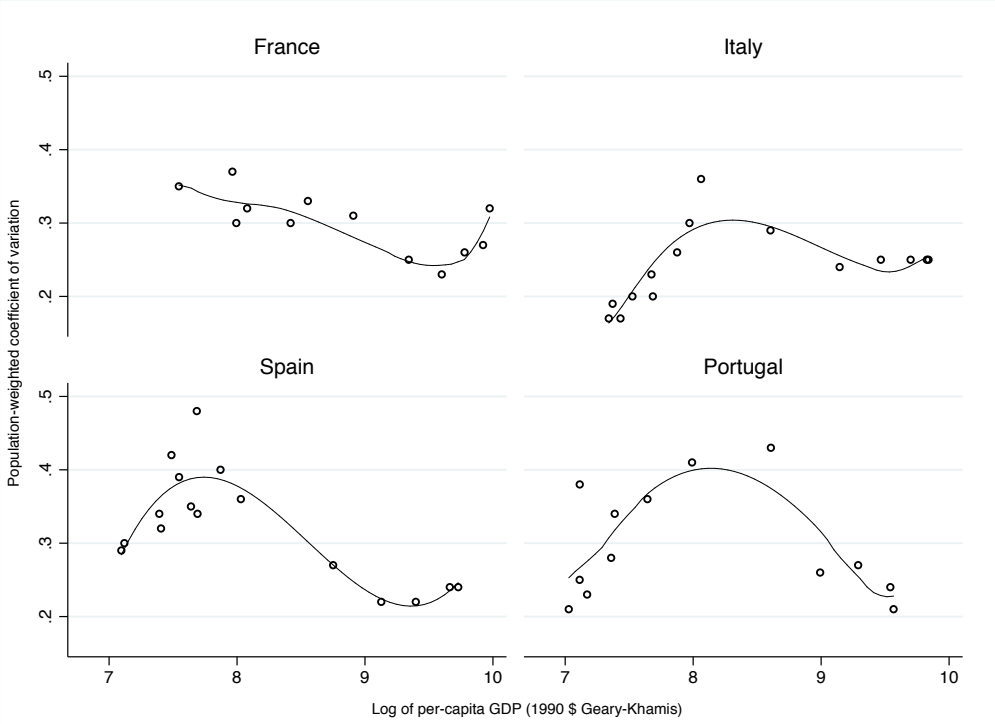
Source: main text.

All in all, the relationship between economic development and regional disparities can be summarized in the scatter plots depicted in Figure 4. The inverted-U relationship between development and regional inequality can be plainly identified, although some country-specific features merit comment. First, in the case of France the upward trend of the inverted U-curve does not show up. It could be argued that this might be due to the fact that modern economic growth in this early-industrialized country started before 1860, the first year with available estimates of per-capita GDP (and an absence of data for 1870-1890). Second, at high levels of per-capita income, regional disparities show different patterns. While in the case of France there is a notable increase in spatial inequality, in Italy and Spain the convergence process came to a halt in the later stages of development, and this led to a stabilization of regional disparities (or a slight increase). Finally in Portugal, while the presence of an inverted U-shape is clearly shown, there is (still?) no evidence of an increase in regional inequalities in recent decades. This initial exploration seems to point to the existence of a non-monotonic bell-shaped curve throughout the development process in the countries of South-West Europe, with mixed evidence as to the evolution of regional inequality in the later stages of economic growth. In order to further investigate these issues, in the next section

[24] Regional GDP estimates (at current prices) for 1920 may suffer an upward bias due to the impact of the Great War on prices.

we conduct a sound econometric analysis to empirically test the relationship between economic development and regional inequality.

Fig. 4. Regional income dispersion (WCV) and per-capita GDP (All sample, 1860-2010)



Source: main text.

4. Econometric analysis

4.1. Methodology

A first way to test the relationship between regional inequality and economic development (measured by per-capita GDP) is to run a simple parametric OLS regression for our pool of four countries in South-West Europe (France, Italy, Portugal and Spain). We follow a parametric specification similar to that proposed by Barrios and Strobl (2009) and Lessmann (2014), which takes the following form:

$$WCV_{it} = \alpha + \sum_{j=1}^{k} \beta_j Y_{it}^{j} + \sum_{m=1}^{q} \gamma_m X_{mit} + \varepsilon_{it} \qquad i = 1,2,...,N,\ t = 1,2,...,T \qquad (2)$$

where WCV_{it} is the Williamson coefficient of variation (population-weighted coefficient of variation) of the national per-capita GDP for country i at time t; Y_{it} is the logarithm of per-capita GDP at country level, which enters the regression in a polynomial form of degree j; and X_{mit}

represents q different control variables at country level.[25] ε_{it} is the error term. As in Barrios and Strobl (2009) and Lessmann (2014), we run pooled regressions for the period 1860–2010 including all the observations.[26]

The control variables are similar to those selected by Lessmann (2014), based on previous empirical literature on spatial inequalities[27]. These controls include, in log scale, the number of regions by country (NUTS2 regions), the country size in square kilometers, and the ratio of both variables, measuring average sub-national unit size. As well as these geographical variables, we also add a set of economic variables related to the economic size of the country that may have an impact on regional inequalities. First, we include a measure of national trade openness. A fair amount of research has empirically established a relationship between trade and regional inequality (Rodríguez-Pose, 2012; Ezcurra and Rodríguez-Pose, 2014; Hirte and Lessmann, 2014). We therefore include the national trade openness ratio, defined as the sum of imports and exports as a share of GDP, to measure the evolution of trade, in our case since the wave of globalization of the late nineteenth century (O'Rourke and Williamson, 1999).[28]

Second, we take into account the influence of agglomeration economies on spatial disparities (Puga, 2002; Baldwin and Martin, 2004; Brülhart and Sbergami, 2009; Desmet and Henderson, 2015). We use the Krugman Specialization Index (KSI), an indicator that shows how similar or dissimilar the employment shares are across industries. Indeed, as defined by Kuznets (1955), structural change is a key characteristic in the long process of modern economic growth. In such a context, regions that have over time specialized in the manufacturing and service sectors (where agglomeration tends to occur) would be in a better position to take advantage of the positive effect that the adoption of technologies may have on economic growth. By contrast, an agricultural specialization may have a negative effect on the regional growth trajectory (Barrios and Strobl, 2009).[29]

---

[25] As in Lessmann's (2014) cross-sectional model, our pooled regression has no time or country dummies. In some preliminary estimations we introduced country and time fixed effects to control for country-specific time-invariant unobservables and time-specific factors common to all the countries in the sample. Nevertheless, as in Barrios and Strobl (2009), these time and country dummies were not significant in most of the parametric regressions and in all the semiparametric models. The results of these tests are available from the authors on request. Moreover, we cannot include some of our time-invariant controls at country level with country fixed effects.

[26] Lessmann (2014) also estimates a panel data version of Eq. (1) to take advantage of the panel dimension of his data set. Nevertheless, our data show a small cross-section dimension (four countries) and low frequency in the time dimension (decade-by-decade observations, with some missing data), while the desirable analysis is with the time dimension fixed and the cross-section dimension tending to infinity (Wooldridge, 2002). Therefore we discard a panel data analysis.

[27] The statistical appendix at the end of the paper details how the variables have been computed.

[28] This is a standard measure in the empirical literature for capturing trade openness (Frankel and Rose, 2002), although some authors have suggested alternative indicators (Alcalá and Ciccone, 2004; Hirte and Lessmann, 2014). Also, Rodriguez-Pose and Gill (2006) take into account trade composition using a ratio of primary sector to secondary.

[29] In addition, we also consider urban population so as to capture the potential presence of agglomeration economies. In particular, we use the fraction of population living in large urban areas, i.e. in cities with over 100,000 inhabitants. The results, available on request, are not significantly altered.

Finally, public policies and redistribution have gained momentum in more recent decades, especially in the post-WWII era (Lindert, 2004). The magnitude and evolution of the redistributive capacity of the states included in our sample may have had an impact on territorial disparities (Rodríguez-Pose and Ezcurra, 2010; Lessmann, 2012). We control for the size of the public sector, measured by the share of total governmental expenditure over GDP, as a proxy for the redistributive capacity of the countries in the sample. Table A.1 in the Appendix shows the summary statistics of all variables employed.

The main coefficients of interest are the β_j. We consider different polynomial functions with values of k from 1 to 3 in order to capture any non-linearity in the relationship between WCV_{it} and Y_{it}. In the case of $k = 2$ (a quadratic function), we expect $\beta_1 > 0$ and $\beta_2 < 0$, which implies an inverted U-shaped relationship between spatial inequality and development, in line with Williamson (1965). Furthermore, a cubic term is included in some specifications to control for a possible increasing spatial inequality at high levels of development after the inverted-U pattern has been completed (Amos, 1988; Lessmann, 2014).

The model in Eq. (2) is a simple way to directly test for the existence of any non-linear behaviour by using different polynomial degrees. Nevertheless, a parametric specification implies strong assumptions; it imposes a particular structure on the underlying relationship between the variables (that may not reflect the true underlying relationship), and the coefficients are not allowed to change over time and country (the relationship is restricted to being stationary over the entire structure of the GDP distribution).

To overcome these limitations, Durlauf (2001) suggests the use of semiparametric methods. This approach allows us to tackle the possible non-linear effect of economic development on regional inequality in a flexible way. For instance, the standard correlation index and the coefficients from parametric regressions give us only an aggregate average relationship between inequality and GDP, and this relationship is restricted by the fact that it must remain unchanged through the entire distribution of GDPs. In contrast, the semiparametric estimate allows inequality to vary with GDP over the entire distribution, allowing for the linear effects of other conditioning variables. In the related literature, Barrios and Strobl (2009) and Lessmann (2014) have applied this methodology to the study of regional inequalities. We perform a semiparametric analysis using the kernel regression estimator (Robinson, 1988). This consists of taking the following specification:

$$WCV = \alpha + f(Y) + \gamma X + \varepsilon, \qquad (3)$$

in which, for the sake of clarity, we drop the subscripts it. X is a set of explanatory variables that are assumed to have a linear effect on WCV, $f(.)$ is a smooth and continuous, possibly non-

linear, unknown function of Y, and ε is a random error term. Thus the model has a parametric (γX) and a nonparametric ($f(Y)$) part. Robinson's approach is a two-step methodology. First, $\hat{\gamma}$ is estimated applying a procedure similar to that whereby variables can be partialed out of an OLS regression (but using nonparametric regressions). Second, a kernel regression of $WCV - \hat{\gamma}X$ on Y is carried out. In both stages, a Gaussian kernel-weighted local polynomial fit is used for kernel regressions.[30]

4.2. Results

Table 2 shows the results of the OLS estimation of Eq. (2). The first column corresponds to a simple bivariate regression, finding a negative but not significant impact of per-capita GDP on regional inequality, measured by the Williamson coefficient of variation. In column 2 we include both the level of per-capita GDP and its square term to capture any non-linearity in its relationship with inequality. The estimated coefficients are significant at 10%, with the GDP coefficient being positive while the square value of GDP shows a negative effect on the WCV. This result is consistent with an inverted U-shaped relationship between spatial inequality and development, although both coefficients become not significant when we include our set of controls (columns 3 and 4). Nevertheless, when we consider a third-degree polynomial function (column 5) we obtain robust evidence of the inverted U-shaped relationship between per-capita GDP and regional inequality. Moreover, the estimated coefficient of the cubic term is positive, which, as in Lessmann (2014), implies that spatial inequality increases after the inverted-U pattern has been completed.

Regarding the control variables, the only significant coefficient in all the regressions (columns 3 to 6) corresponds to country size, indicating that the bigger the country the higher the regional inequality. The other geographical controls are in no case significant, and therefore there is no significant heterogeneity in the territorial classification between countries. The set of economic variables are also not significant. First, we do not find a statistically significant relationship between globalization, measured through the rate of openness, and the level of regional inequality. An explanation for this could be that, given that the markets for goods, capital and labour evolved together in terms of their degree of integration, at least during the First Globalization (1870-1913), our indicator might be capturing the joint effect of these three elements. In this case, as pointed out by Hirte and Lessmann (2014), the effects of an increasing international integration in the capital and labour markets could be compensating for any cost in terms of regional inequality that trade growth may bring.

As regards the variable for general government expenditure as a share of GDP, which aims to capture the impact of redistributive policies on spatial inequalities, we recognize that the lack of

[30] The semiparametric models are estimated using the "semipar" Stata package developed by Verardi and Debarsy (2012). See Barrios and Strobl (2009) and Lessmann (2014) for more details.

significance may be related to the distance that exists between the theoretical effect that we aim to capture and our variable. General government expenditure might not directly capture the amount of resources devoted to territorial cohesion policies, for instance, or, more importantly, the potential changes in the relevance of these types of policy on total public expenditure throughout history. In particular, in the four economies included in our study the amount of public resources devoted to redistributive policies was rather limited before the 1950s.

Finally, the variable that aims to capture the potential effect on regional inequality of agglomeration economies in a number of specific economic sectors, the KSI, is not significant. This result holds when we use an alternative measure of agglomeration, such as urbanization rates. However, as highlighted by Brülhart and Sbergami (2009), the link between agglomeration and regional inequality may not be linear throughout the long process of economic development, and this would explain the non-significant relationship between these two variables.

Table 2. Parametric estimates (pooled regressions, 1860–2010)

	(1)	(2)	(3)	(4)	(5)
ln(GDPpc)	-0.020	0.573*	0.602	-0.067	13.994***
	(0.016)	(0.241)	(0.279)	(0.323)	(2.197)
ln(GDPpc) ^{2}		-0.035*	-0.036	0.004	-1.642***
		(0.014)	(0.016)	(0.021)	(0.239)
ln(GDPpc) ^{3}					0.064***
					(0.008)
Number of regions (ln scale)			0.067	-0.100	-0.136
			(0.164)	(0.262)	(0.245)
Country size (km^2, ln scale)			0.096**	0.102***	0.111***
			(0.017)	(0.013)	(0.012)
ln(area)/ln(units)			0.094	0.005	0.015
			(0.070)	(0.121)	(0.118)
Openness ratio				-0.000	-0.001
				(0.002)	(0.002)
General Gov't Expenditure /GDP				-0.001	0.001
				(0.002)	(0.002)
Krugman Specialization Index				1.010	0.477
				(0.776)	(0.872)
Observations	56	56	56	56	56
R^2	0.075	0.177	0.379	0.408	0.565

Note: Coefficient (robust standard errors clustered by country). Dependent variable: Williamson coefficient of variation (population-weighted CV). Significant at the *10%, **5%, ***1% level. All the models include a constant.

The results make it possible to trace the relationship between economic development and regional inequality according to the South-West European experience. Based on the estimations reported in column (5) of Table 2, regional inequalities increase up to an income of ln(GDPpc)=8.03, which

corresponds to approximately 3,000 US$1990. That level of income was reached in France by 1910, in Italy in 1950 and in Portugal and Spain in 1960. From then on spatial disparities fall, reaching their minimum at an income of ln(GDPpc)=9.5, i.e. around 13,000 US$1990. While France reached that threshold of income around 1975 and Italy in 1980, the economies of Portugal and Spain had to wait until the late 1990s to reach that level of income per capita. Beyond that level, a new upsurge in regional inequalities is observed. Hence these results provide evidence in support of the idea that, in the course of the long-term process of economic development, spatial disparities evolve in the shape of an inverted-U curve. In the early stages of modern economic growth during the start of industrialization, inequality increases. Convergence then begins as industrialization spreads to a larger number of regions; the Kuznets curve was completed in the final decades of the twentieth century. Interestingly, since then a new upsurge in spatial disparities has appeared, now in the context of service-based economies.

Next we conduct a semiparametric analysis. As mentioned above, we obtain two outputs from Eq. (3): a set of regression coefficients for the linear part of the model and a graph which shows the nonlinear relationship between spatial inequality and development. Table 3 reports the results of the linear part of the model. As in the parametric regressions (Table 2), the only significant coefficient corresponds to country size. Figure 5 shows the nonparametric part of the estimation, including the 95% confidence bands.

Table 3. Semiparametric estimates, linear part of the model (pooled regressions, 1860–2010)

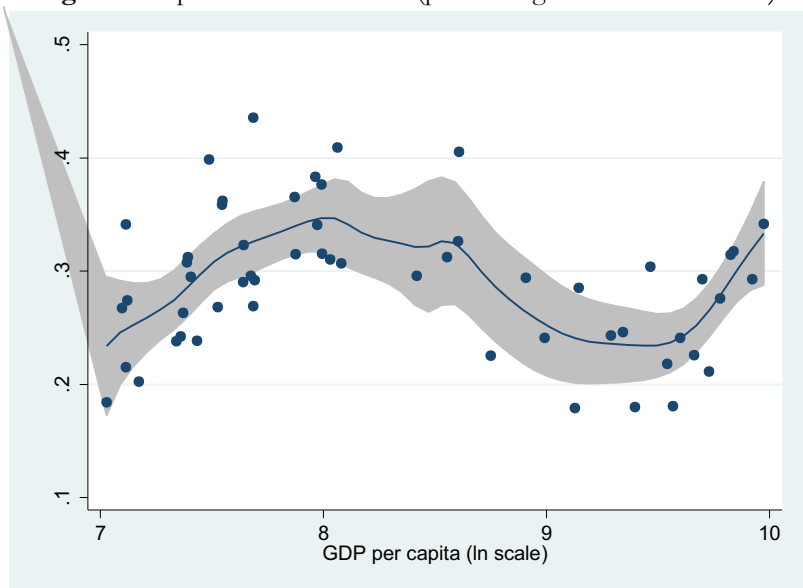
	(1)
Number of regions (ln scale)	-0.261
	(0.269)
Country size (km², ln scale)	0.125***
	(0.016)
ln(area)/ln(units)	-0.041
	(0.128)
Openness ratio	-0.001
	(0.002)
General Gov't Expenditure /GDP	0.001
	(0.002)
Krugman Specialization Index	0.354
	(0.743)
Observations	56
R^2	0.410

Note: Coefficient (robust standard errors clustered by country). Dependent variable: Williamson coefficient of variation (population-weighted CV). Significant at the *10%, **5%, ***1% level. All the models include a constant.

The graph (Fig. 5) shows a clear bell-shaped curve, supporting the idea of an inverted U-shaped relationship between spatial inequality and development over time. It also shows increasing

inequalities at high levels of economic development at the upper tail of the GDP distribution. The thresholds of per-capita GDP at which the relationship between spatial inequality and development reverses are similar to those calculated for the parametric regression.

Fig. 5. Semiparametric estimates (pooled regressions 1860–2010)



Source: main text.

Our results can be interpreted as follows. The initial upsurge in regional income inequality is linked to the different timing and intensity of the take-off of industrialization across the regions of South-West Europe. Technological shocks came about with the adoption of the new technologies that characterized the First and Second Technological Revolutions. This favoured an increase in regional inequality up to the 1950s, from which time a process of convergence began. The diffusion of the new technologies and production sectors to a greater number of regions would be responsible for the reduction in regional inequality. However, convergence came to a halt in the 1980s and a new upsurge in regional income inequality followed. The last three decades correspond to a period of high levels of income and a deeper process of tertiarization in the most advanced regions in our sample. The new technological shock arising from the Third Revolution, associated with information and communication technologies (ICTs), would be the reason behind the uneven regional growth.

Indeed, throughout the whole process of economic development, the structural transformation of the regional economies has proceeded with different timings and in an economic context characterized by changing conditions affecting, among other things, international trade and public policies. In order to further explore this issue, we next examine the potential mechanism behind the inverted-U curve suggested by Kuznets (1955) and Williamson (1965). These authors established that structural change is a major driver of regional inequalities, with the transition from agrarian to industrial economies being responsible for the evolution of these inequalities. To investigate this

connection, we modify our baseline specification and consider the percentage of non-agricultural gross value added (GVA) over total GDP as the main explicative variable instead of per-capita GDP. In our case, non-agricultural GVA includes industry and service sectors.

Table 4 shows the results of the parametric regressions using the non-agricultural GVA. It can be seen that spatial inequalities increase as the relative size of the non-agricultural sectors increases, but the negative coefficient of the square non-agricultural GVA/GDP points to a nonlinear relationship, indicating the existence of a reverse point after which spatial inequalities decrease as the relative size of the modern sectors increases. To confirm this result, we again estimate the semiparametric model. Table 5 reports the semiparametric estimates for the linear part of the model using sectoral data, while Figure 6 shows the nonparametric part of the estimation. A bell-shaped curve is again obtained, and the results are consistent with those obtained using per-capita GDP. However, in this case the increase in more developed stages of development does not appear in the curve, although signs of an increase are visible. Thus, as in Lessmann (2014), this result suggests that structural change (the shift from agriculture to industry) has caused the inverted-U pattern observed in the data.

Table 4. Parametric estimates using sectoral data

	(1)	(2)	(3)	(4)	(5)
Non-agricultural GVA/GDP	-0.0004	0.048**	0.043**	0.063**	0.111
	(0.001)	(0.010)	(0.011)	(0.012)	(0.082)
(Non-agricultural GVA/GDP) ^{2}		-0.0003**	-0.0002**	-0.0004**	-0.001
		(0.0001)	(0.0001)	(0.0001)	(0.001)
(Non-agricultural GVA/GDP) ^{3}					0.000
					(0.000)
Number of regions (ln scale)			-0.032	-0.020	-0.003
			(0.072)	(0.088)	(0.083)
Country size (km^2, ln scale)			0.060**	0.061*	0.055
			(0.012)	(0.023)	(0.025)
ln(area)/ln(units)			0.019	0.023	0.027
			(0.035)	(0.030)	(0.022)
Openness ratio				0.001	0.001
				(0.001)	(0.001)
General Gov't Expenditure /GDP				0.0003	-0.0001
				(0.002)	(0.002)
Krugman Specialization Index				-0.609	-0.447
				(0.299)	(0.265)
Observations	56	56	56	56	56
R^2	0.009	0.551	0.602	0.639	0.643

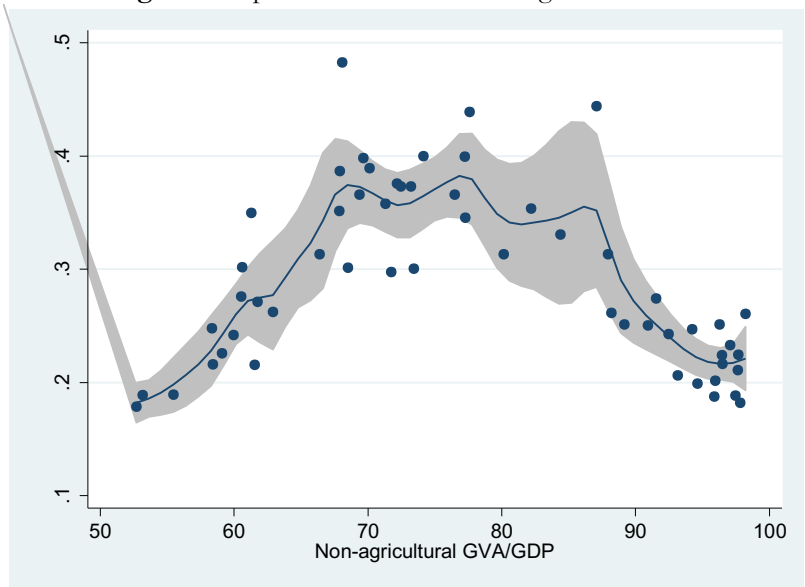
Note: Coefficient (robust standard errors clustered by country). Dependent variable: Williamson coefficient of variation (population-weighted CV). Significant at the *10%, **5%, ***1% level. All the models include a constant.

Table 5. Semiparametric estimates using sectoral data, linear part of the model

	(1)
Number of regions (ln scale)	0.056
	(0.051)
Country size (km^2, ln scale)	0.046**
	(0.013)
ln(area)/ln(units)	0.051**
	(0.015)
Openness ratio	0.001
	(0.001)
General Gov't Expenditure /GDP	-0.0005
	(0.002)
Krugman Specialization Index	-0.558**
	(0.136)
Observations	56
R-squared	0.173

Note: Coefficient (robust standard errors clustered by country). Dependent variable: Williamson coefficient of variation (population-weighted CV). Significant at the *10%, **5%, ***1% level. All the models include a constant.

Fig. 6. Semiparametric estimates using sectoral data



Source: main text.

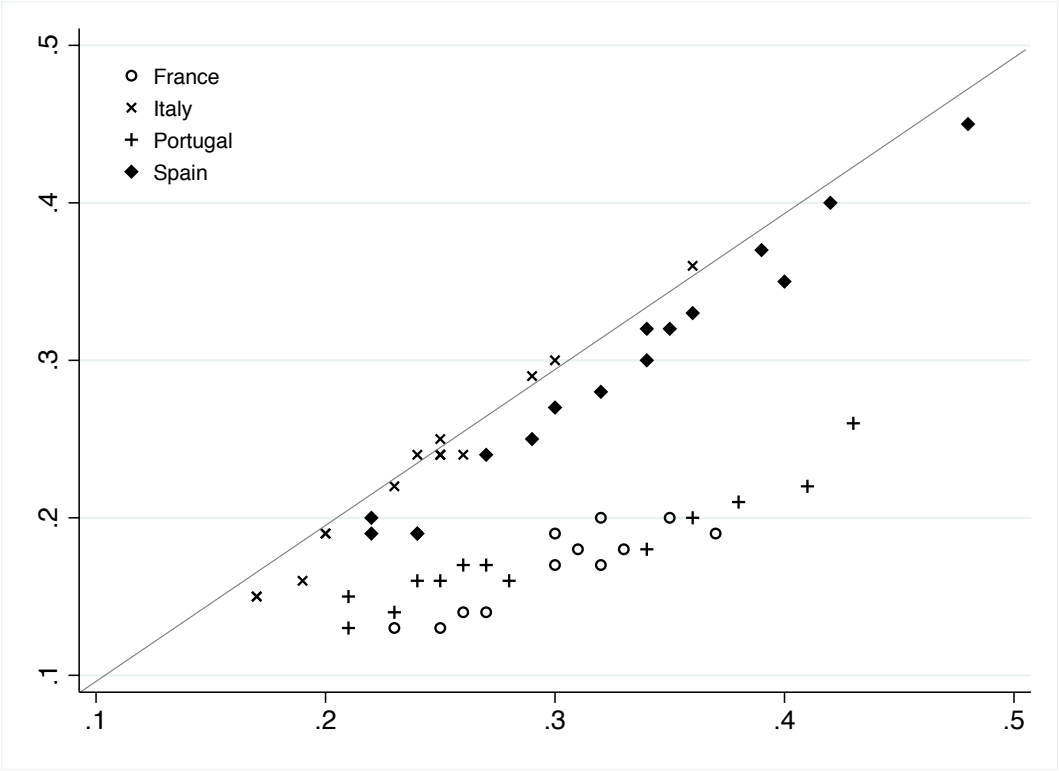
4.3. Robustness

We carry out several robustness checks. First we consider alternative measures of inequality instead of the WCV: the unweighted or single coefficient of variation (CV) and the Gini coefficient. Table A.2 in the Appendix shows the results of the parametric regressions for each endogenous variable. The results using the CV and Gini (columns 1-4) are consistent with those in Table 2. Table A.3 reports the semiparametric estimates for the linear part of the model using the two alternative

measures of inequality (columns 1-2). Again, these results are similar to those in Table 3. Figures A.1. and A.2 show the nonparametric part of the estimation. The bell-shaped curve is clear in both cases, although the pattern is more pronounced when the Gini coefficient is used.

To avoid any possible bias due to the predominance of the capital cities (Ades and Glaeser, 1995), we calculate the WCV excluding the region in which the country's capital is located. Fig. 7 shows that there is indeed a strong capital effect impacting regional inequality in the countries in our sample, especially in the case of France and Portugal. The (downward) deviation from the diagonal shows that when the region in which the capital city is located is excluded, regional inequality decreases substantially. A smaller effect is found for both Spain and Italy. As regards the results of the estimations (Table A.2, columns 5 and 6), the only difference is that, when we use the WCV excluding the region with the capital city, the GDP coefficients are not significant. Table A.3 (column 3) reports the semiparametric estimates for the linear part of the model with the measures of inequality excluding capital cities. Again, these results are similar to those in Table 3. Figure A.3 shows the nonparametric part of the estimation and the bell-shaped curve is clear.

Fig. 7. Regional inequality with capital cities included (x-axis) and excluded (y-axis).



Note: Williamson (population-weighted) coefficient of variation.

5. Conclusions

The present study has analysed the evolution and determinants of regional income inequality throughout the historical process of economic development in four countries of South-West Europe: France, Italy, Portugal and Spain. In order to do this, a new data set on regional per-capita GDP for these four countries has been assembled on a decadal basis beginning in 1860 and ending in 2010. In particular, we have analysed the existence of an inverted-U curve since the early stages of modern economic growth, as the Williamson hypothesis suggests.

Three main results are obtained. First, the study confirms that, over the course of the historical process of economic development, regional income inequality has followed a U-shaped evolution. Interestingly, in the more recent stages of development the trend is on the rise again. This result, obtained using both parametric regressions with polynomial functions for income and a semiparametric approach, is robust not only to the inclusion of confounding factors that may have an effect on regional inequality, but also to the use of alternative measures of inequality. Second, our results also show that economic growth has been more intense in the most populated regions, which, in the context of the four South-West European countries studied, correspond to those in which the capital cities are located. The dynamism of these 'capital' regions could therefore explain the recent upsurge in regional inequalities. Third, in line with Kuznets (1955) and Williamson (1965), we identify structural change as being a significant transmission mechanism of the inverted-U relationship that we have found between economic development and regional inequality in the long term.

Additionally, the results enable the historical relationship between economic development and regional inequality to be traced back in accordance with the experiences of South-West Europe. Based on the estimations, regional inequalities increased until an income of approximately 3,000 US$1990 was reached. This threshold was achieved by France around 1910, by Italy around 1950 and by Portugal and Spain around 1960. From then on spatial disparities fell, reaching a minimum at an income of around 13,000 US$1990. While this happened in France and Italy in the late 1970s, the economies of Portugal and Spain had to wait until the late 1990s before such an income per capita was attained. Beyond that level a new upsurge in regional inequalities is observed. From a historical perspective, the inverted U-shaped relationship between national economic development and spatial inequality has in the end become an elephant.

Acknowledgements

Financial support from the projects ECO2013-45969-P, ECO2015-65049-C12-1-P, ECO2015-71534-REDT and ECO2016-75941-R (*Ministerio de Economía y Competitividad*), the DGA (ADETRE research group) and FEDER is gratefully acknowledged.

References

Ades, A.F., Glaeser, E.L., 1995. Trade and circuses: explaining urban giants. *The Quarterly Journal of Economics* 110 (1), 195-227.

Abramovitz, M., 1994. Catch-up and convergence in the postwar growth boom and after. In W. Baumol, R. Nelson and E. Wolff (eds), *The convergence of productivity: cross-national studies and historical evidence* (Oxford: Oxford University Press), 86-125.

Alcalá, F., Ciccone, A., 2004. Trade and productivity. *The Quarterly Journal of Economics* 129 (2), 613-646.

Amos, O.J., 1988. Unbalanced regional growth and regional income inequality in the latter stages of development. *Regional Science and Urban Economics* 18 (4), 549-566.

Badia-Miró, M., Guilera, J., Lains, P., 2012. Regional incomes in Portugal: industrialisation, integration and inequality, 1890-1980. *Revista de Historia Económica* 30 (2), 225-244.

Baldwin, R., Martin, P., 2004. Agglomeration and regional growth. In J.V. Henderson and J.F. Thisse (eds.), *Handbook of regional and urban economics*, vol.4 (Amsterdam: Elsevier), 2671-2711.

Barrios, S., Strobl, E., 2009. The dynamics of regional inequalities. *Regional Science and Urban Economics* 39 (5), 575-591.

Barro, R., Sala-i-Martin, X., 1991. Convergence across states and regions. *Brooking Papers on Economic Activity* 1, 107-82.

Barro, R., Mankiw, G., Sala-i-Martin, X., 1995. Capital mobility in neoclassical models of growth. *American Economic Review* 85 (1), 103-115.

Bazot, G., 2014. Interregional inequalities, convergence, and growth in France from 1840 to 1911. *Annals of Economics and Statistics* 113/114, 309-345.

Bolt, J., Van Zanden, J.L., 2014. The Maddison Project: collaborative research on historical accounts. *Economic History Review* 67 (3), 627-651.

Breinlich, H., Ottaviano, G.I.P., Temple, J., 2014. Regional growth and regional decline. In P. Aghion and S.N. Durlauf (eds.), *Handbook of economic growth*, vol.2 (Amsterdam: Elsevier), 683-779.

Broadberry, S., O'Rourke, K.H., 2010. *The Cambridge economic history of modern Europe* (Cambridge: Cambridge University Press).

Brülhart, M., 2011. The spatial effects of trade openness: a survey. *Review of World Economics* 147 (1), 59-83.

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

Buyst, E., 2010. Reversal of fortune in a small, open economy: regional GDP in Belgium, 1896-2000. *Rivista di Storia Economica* 26, 75-92.

Buyst, E., 2011. Continuity and change in regional disparities in Belgium during the twentieth century. *Journal of Historical Geography* 37 (3), 329-337.

Caruana-Galizia, P., 2013. Estimating French regional income: departmental per capita gross value added, 1872-1911. *Research in Economic History* 29, 71-95.

Caselli, F., Coleman II, W.J., 2001. The U.S. structural transformation and regional convergence: reinterpretation. *Journal of Political Economy* 109 (3), 584-616.

Ciccone, A., 2002. Agglomeration effects in Europe. *European Economic Review* 46 (2), 213-227.

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

Combes, P.P., Lafourcade, M., Thisse, J.F., Toutain, J.C., 2011. The rise and fall of spatial inequalities in France: a long-run perspective. *Explorations in Economic History* 48 (2), 243-271.

Comín, F., Díaz, D., 2005. Sector público administrativo y estado del bienestar. In A. Carreras and X. Tafunell (eds.), *Estadísticas históricas de España, siglos XIX-XX* (Bilbao: Fundación BBVA), 873-964.

Crafts, N., 2005. Regional GDP in Britain, 1871–1911: some estimates. *Scottish Journal of Political Economy* 52 (1), 54-64.

Crozet, M., Koenig, P. 2004. Trade liberalization and the internal geography of countries. In T. Mayer, M. Mucchielli (eds.), *Multinational firms' location and economic geography* (Cheltenham: Edward Elgar), 91-109.

Daniele, V., Malanima, P., 2014. Falling disparities and persisting dualism: regional development and industrialization in Italy, 1891-2001. *Investigaciones de Historia Económica* 10 (3), 165-176.

Desmet, K., Henderson, J.V., 2015. The geography of development within countries. In G. Duranton, J.V. Henderson and W.C. Strange (eds), *Handbook of regional and urban economics*, vol. 5 (Amsterdam: Elsevier), 1457-1517.

Díez-Minguela, A., Rosés, J.R., Sanchis, M.T., 2016. French and the Paris desert revisited: regional income polarization in France, 1860-2010. *Paper presented at 56th European Regional Association Congress (ERSA), Vienna 23th to 26th August.*

Díez-Minguela, A., Sanchis, M.T., 2017. Regional Income inequality in France: what does history teach us? *Paper presented at XLIII Reunión de Estudios Regionales, Pablo de Olavide Unversity, Seville 16th-17th November.*

Durlauf, S.N., 2001. Manifesto for a growth econometrics. *Journal of Econometrics* 100 (1), 65-69.

Enflo, K., Henning, M., Schön, L., 2014. Swedish regional GDP 1855-2000: estimations and general trends in the Swedish regional system, *Research in Economic History* 30, 47-89.

Ezcurra, R., Rapún, M., 2006. Regional disparities and national development revisited: the case of Western Europe. *European Urban and Regional Studies* 13 (4), 355-369.

Ezcurra, R., Rodríguez-Pose, A., 2014. Trade openness and spatial inequality in emerging countries. *Spatial Economic Analysis* 9 (2), 162-182.

Felice, E., 2011. Regional value added in Italy, 1891-2001, and the foundation of a long-term picture. *Economic History Review* 64 (3), 929-950.

Felice, E., 2015. *Ascesa e declino. Storia economica d'Italia* (Bologna: il Mulino).

Felice, E., Vecchi, G., 2015. Italy's growth and decline, 1891-2011. *Journal of Interdisciplinary History* 45 (4), 507-548.

Flora, P., 1983. *State, economy and society in Western Europe, 1815-1975* (London: Macmillan).

Frankel, J.A., Rose, A., 2002. An estimate of the effect of common currencies on trade and income. *The Quarterly Journal of Economics* 117 (2), 437-466.

Fujita, M., Thisse, J.F., 2002. *Economics of agglomeration* (Cambridge: Cambridge University Press).

Gardiner, B., Martin, R., Tyler, P., 2011. Does spatial agglomeration increase national growth? Some evidence from Europe. *Journal of Economic Geography* 11 (6), 979-1006.

Geary, F., Stark, T., 2002. Examining Ireland's post-famine economic growth performance. *Economic Journal* 112, 919-35.

Geary, F., Stark, T., 2015. Regional GDP in the UK, 1861–1911: new estimates. *Economic History Review* 68 (1), 123-144.

Geary, F., Stark, T., 2016. What happened to regional inequality in Britain in the twentieth century? *Economic History Review* 69 (1), 215-228.

Gennaioli, N., La Porta, R., Lopez de Silanes, F., Schleifer, A., 2014. Growth in regions. *Journal of Economic Growth* 19, 259-309.

Giannetti, M., 2002. The effects of integration on regional disparities: convergence, divergence or both? *European Economic Review* 46, 539-567.

Hanson, G.H., 2005. Market potential, increasing returns, and geographic concentration. *Journal of International Economics* 67, 1-24.

Hirschman, A.O., 1958. *The strategy of economic development* (New Haven: Yale University Press).

Hirte, G., Lessmann, C., 2014. Trade, integration, and interregional inequality, *CESifo Working Papers Series* 4799.

INSEE. National accounts – 2010 base (Table: t-1105). https://www.insee.fr/.

Kanbur, R., Venables, A.J., (eds.) 2005. *Spatial inequality and development* (Oxford: Oxford University Press).

Kanbur, R., Zhang, X., 2005. Fifty years of regional inequality in China: a journey through central planning, reform, and openness. *Review of Development Economics* 9 (1), 87-106.

Kanbur, R., Wang, L., Zhang, X., 2017. The great Chinese inequality turnaround. *CEPR Discussion Paper* 11892.

Kim, S., 1998. Economic integration and convergence: US regions, 1840–1987. *Journal of Economic History* 58 (3), 659-683.

Kim, S., Margo, R., 2004. Historical perspectives on U.S. economic geography. In J.V. Henderson and J.F. Thisse (eds.), *Handbook of regional and urban economics*, vol.4 (Amsterdam: Elsevier), 2981-3019.

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

Krugman, P., Livas Elizondo, R., 1996. Trade policy and the Third World metropolis. *Journal of Development Economics* 49 (1), 137-150.

Kuznets, S., 1955. Economic growth and income inequality. *American Economic Review* 45 (1), 1-28.

Lains, P., 2003. Catching-up the European core: Portuguese economic growth, 1910-1990. *Explorations in Economic History* 40, 269-386.

Lains, P., 2007. Growth in a protected environment: Portugal, 1850-1950. *Research in Economic History* 24, 119-160.

Lessmann, C., 2012. Fiscal decentralization and regional disparity: evidence from cross-section and panel data, *Environment and Planning A* 44 (6), 2455-2473.

Lessmann, C., 2014. Spatial inequality and development. Is there an inverted-U relationship. *Journal of Development Economics* 106, 35-51.

Lessmann, C., Seidel, A. 2017. Regional inequality, convergence, and its determinants. A view from outer space. *European Economic Review* 92, 110-132.

Lindert, P.H., 2004. *Growing public. Social spending and economic growth since the eighteenth century* (Cambridge: Cambridge University Press).

List, J.A., Gallet, C.A., 1999. The Kuznets curve: what happens after the inverted-U?, *Review of Development Economics* 3 (2), 200-206.

Lucas, R.E., 2000. Some macroeconomics for the 21st Century. *Journal of Economic Perspectives* 14 (1), 159-168.

Maddison Project Database: http://www.ggdc.net/maddison/maddison-project/data.htm

Magrini, S., 2004. Regional (di)convergence. In J.V. Henderson and J.F. Thisse (eds.), *Handbook of regional and urban Economics*, vol. 4 (Amsterdam: Elsevier), 2741-2796.

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

Milanovic, B., 2005. Half a world: regional inequality in five great federations. *Journal of the Asia Pacific Economy* 10 (4), 408-445.

Mitchell, B.R., 2007. *International historical statistics: Europe, 1750-2005* (Basingstoke: Palgrave Macmilllan).

Mitchener, K.J., McLean, I.W., 1999. U.S. regional growth and convergence, 1880–1980. *The Journal of Economic History* 59 (4), 1016-1042.

Myrdal, G., 1957. *Economic theory and under-developed regions* (London: Methuen & Co).

OCDE. National accounts at a glance. Government. https://stats.oecd.org.

O'Rourke, K.H., Williamson, J.G., 1999. *Globalization and history. The evolution of a nineteenth-century Atlantic economy* (Cambridge: The MIT Press).

Paluzie, E., 2001. Trade policy and regional inequalities. *Papers in Regional Science* 80, 67-85.

Pollard, S., 1981. *Peaceful conquest: the industrialization of Europe, 1760-1970* (Oxford: Oxford University Press).

Prados de la Escosura, L., 2003. *El progreso económico de España, 1850-2000* (Bilbao: Fundación BBVA).

Prados de la Escosura, L., 2010. Spain's international position, 1850-1913. *Revista de Historia Económica / Journal of Iberian and Latin American Economic History* 28 (1), 173-215.

Prados de la Escosura, L., 2016. Spain's historical national accounts: expenditure and output, 1850-2015, *EHES working papers*, No. 103.

Pritchett, L., 1997. Divergence, big time. *Journal of Economic Perspectives* 11 (3), 3-17.

Puga, D., 1999. The rise and fall of regional inequalities. *European Economic Review* 43 (2), 303-334.

Puga, D., 2002. European regional policy in light of recent location theories. *Journal of Economic Geography* 2 (4), 373-406.

Robinson, P.M., 1988. Root-N consistent semiparametric regression. *Econometrica* 56 (4), 931-954.

Rodríguez-Pose, A., 2012. Trade and regional inequality. *Economic Geography* 88 (2), 109-136.

Rodríguez-Pose, A., Gill, N., 2006. How does trade affect regional disparities? *World Development* 34 (7), 1201-1222.

Rodríguez-Pose, A., Ezcurra, R., 2010. Does decentralization matter for regional disparities? A cross-country analysis. *Journal of Economic Geography* 10 (5), 619-644.

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

Rosés, J.R., Sanchis, M.T., (*forthcoming*). A long run perspective on French regional income inequalities, 1860-2010. In N. Wolf and J.R. Rosés (eds). *The economic development of Europe's regions. A quantitative history since 1900* (London: Routledge).

Singer, J.D., Bremer, S., Stuckey, J., 1972. Capability distribution, uncertainty, and major power war, 1820-1965. In B. Russett (ed.), *Peace, war, and numbers* (Beverly Hills: Sage), 19-48.

Tena, A., 2005. Sector exterior. In A. Carreras and X. Tafunell (eds.), *Estadísticas históricas de España, siglos XIX-XX* (Bilbao: Fundación BBVA), 573-644.

Tirado, D.A., Díez-Minguela, A., Martinez-Galarraga, J., 2016. Regional inequality and economic development in Spain, 1860-2010. *Journal of Historical Geography* 54, 87-98.

Total Economy Database: https://www.conference-board.org/data/economydatabase/

Toutain, J.C., 1987. *Le produit interieur brut de la France de 1789 á 1982. Économies et Sociétés* (Grenoble: Presses Universitaires de Grenoble), Cahiers de l'ISMEA, série AF, 15.

Valério, N. (ed.), 2001. *Estatísticas Históricas Portuguesas* (Lisboa: Instituto Nacional de Estatística).

Verardi, V., Debassy, N., 2012. Robinson's square root of N consistent semiparametric regression estimator in Stata. *The Stata Journal* 12 (4), 726-735.

Williamson, J.G., 1965. Regional inequality and the process of national development: a description of the patterns. *Economic Development and Cultural Change* 13 (4), 1-84.

Wooldridge, J.M., 2002. *Econometric analysis of cross section and panel data* (Cambridge: The MIT Press).

World Bank. https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS

World Bank, 2009. *World Development Report: Reshaping economic geography* (Washington: World Bank).

Appendix

Table A.1. Summary descriptive statistics

Variable	Mean	Standard deviation	Minimum	Maximum
WCV	0.290	0.070	0.168	0.477
WCV (excluding capital region)	0.222	0.074	0.126	0.455
CV	0.291	0.081	0.166	0.494
Gini	0.143	0.039	0.070	0.224
ln(GDPpc)	8.322	0.958	7.028	9.975
Number of regions (ln scale)	2.648	0.584	1.609	3.091
Country size (km², ln scale)	12.601	0.706	11.396	13.177
ln(area)/ln(units)	5.010	1.162	4.211	7.081
Openness ratio	30.195	15.247	4.067	67.084
General Gov't Expenditure /GDP	24.000	15.614	5.795	56.373
Krugman Specialization Index	0.113	0.029	0.057	0.187

Source: see statistical appendix.

Table A.2. Parametric estimates (pooled regressions, 1860-2010), alternative measures of inequality

Dependent variable	Coeff. of Variation		Gini Coefficient		WCV exclud. Capital	
	(1)	(2)	(3)	(4)	(5)	(6)
ln(GDPpc)	-0.049	11.819	-0.024*	4.472**	-0.020	10.599
	(0.016)	(1.033)	(0.010)	(0.789)	(0.025)	(4.688)
$ln(GDPpc)^2$		-1.390***		-0.529***		-1.240
		(0.116)		(0.088)		(0.544)
$ln(GDPpc)^3$		0.054***		0.021***		0.048
		(0.005)		(0.003)		(0.021)
Number of regions (ln scale)		-0.668*		-0.657**		-1.541**
		(0.268)		(0.140)		(0.267)
Country size (km^2, ln scale)		0.079**		0.061***		0.193***
		(0.018)		(0.008)		(0.010)
ln(area)/ln(units)		-0.255		-0.289**		-0.671**
		(0.126)		(0.065)		(0.128)
Openness ratio		-0.001		-0.001		-0.001
		(0.002)		(0.001)		(0.002)
General Gov't Expenditure / GDP		0.000		0.000		0.001
		(0.003)		(0.001)		(0.001)
Krugman Specialization Index		0.361		0.331		0.468
		(0.987)		(0.484)		(0.842)
Observations	56	56	56	56	56	56
R^2	0.064	0.653	0.341	0.623	0.369	0.691

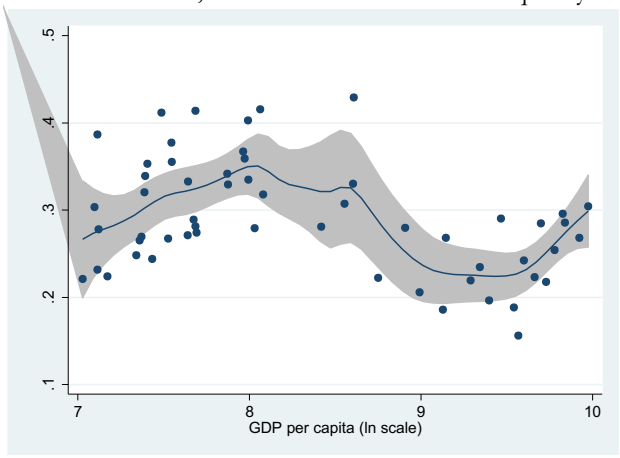
Note: Coefficient (robust standard errors clustered by country). Significant at the *10%, **5%, ***1% level. All the models include a constant. Dependent variables: WCV but excluding the region where the capital city of the country is located, Unweighted Coefficient of Variation (CV) and Gini coefficient.

Table A.3. Semiparametric estimates, linear part of the model (pooled regressions, 1860–2010), alternative measures of inequality

Dependent variable:	Coeff. of Variation	Gini Coefficient	WCV excluding capital
	(1)	(2)	(3)
Number of regions (ln scale)	-0.807*	-0.699**	-1.607**
	(0.282)	(0.152)	(0.302)
Country size (km², ln scale)	0.096***	0.066***	0.200***
	(0.016)	(0.010)	(0.019)
ln(area)/ln(units)	-0.315*	-0.307**	-0.698**
	(0.132)	(0.070)	(0.140)
Openness ratio	-0.001	-0.000	-0.001
	(0.002)	(0.001)	(0.001)
General Gov't Expenditure /GDP	0.000	0.000	0.002
	(0.002)	(0.001)	(0.001)
Krugman Specialization Index	0.227	0.198	0.252
	(0.806)	(0.352)	(0.615)
Observations	56	56	56
R^2	0.414	0.476	0.553

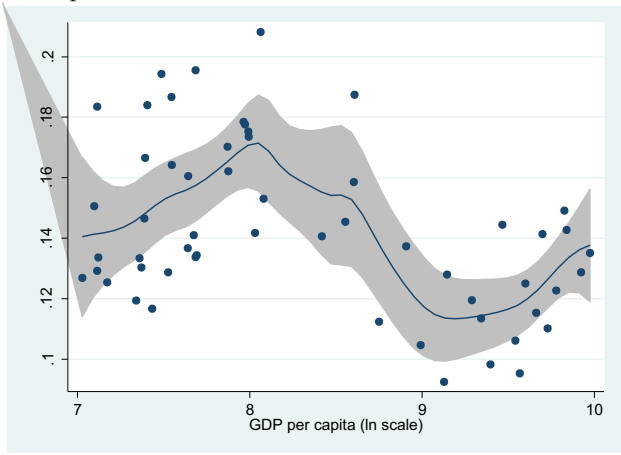
Note: Coefficient (robust standard errors clustered by country). Significant at the *10%, **5%, ***1% level. All the models include a constant. Dependent variables: WCV but excluding the region where the capital city of the country is located, Unweighted Coefficient of Variation (CV) and Gini coefficient.

Figure A.1. Semiparametric estimates, alternative measures of inequality: coefficient of variation



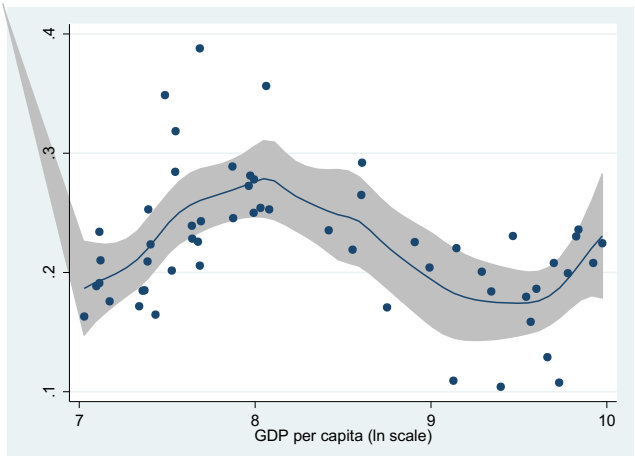
Source: main text.

Figure A.2. Semiparametric estimates, alternative measures of inequality: Gini



Source: main text.

Figure A.3. Semiparametric estimates, alternative measures of inequality: WCV excluding capital cities



Source: main text.

Statistical Appendix

Regional data

GDP and population:

France:

Data for 22 NUTS2 regions come from several sources. The *départements d'outre-mer* are excluded and also Corsica in 1860, 1900 and 1930. No data available for 1870-1890 and 1940. The regions of Alsace and Lorraine are included despite having passed to Germany in the period 1870-1919.

1860, 1896, 1930: Combes et al. (2011).

1911, 1921: Díez-Minguela and Sanchis (2017).

1954, 1962, 1975: Rosés and Sanchis (*forthcoming*).

1982: Eurostat.

1990, 2000, 2007: INSEE, Comptes régionaux. Base 2010**.**

Italy:

Data for 20 NUTS2 regions with current borders come from Felice (2015), table A.2.3. Although today Italy contains 21 NUTS2 regions, in the data set Bolzano is merged with the region of Trentino Alto Adige. GDP data are provided on a decadal basis on years ending in 1, from 1871 to 2011 (with the exception of 1938 for 1940/41). We would like to thank Emanuele Felice for kindly sharing the data with us.

Portugal:

Data for 5 NUTS2 regions between 1890-1980, on a decadal basis (excluding Madeira-Açores). These data, aggregated from the original data set for historical districts, come from Badia-Miró et al. (2012). We would like to thank Marc Badia-Miró, Jordi Guilera and Pedro Lains for kindly sharing the data with us as well as for providing us with the estimates for 1990-2010.

Spain:

Data for 17 NUTS2 regions or *Comunidades Autónomas* between 1860-2010, on a decadal basis (excluding Ceuta and Melilla). Data come from different sources: see Rosés et al. (2010), Martinez-Galarraga et al. (2015), Tirado et al. (2016).

National data

GDP per capita:

Maddison Project Database, Bolt and Van Zanden (2014).

Population:

1850-1950: Maddison: http://www.ggdc.net/maddison/oriindex.htm.

1950-2010: Total Economy Database: https://www.conference-board.org/data/economydatabase/

Openness rate:

Computed as [(X+M)/GDP]. To avoid potential fluctuations we calculate a 3-year average. For recent decades (1960-2010), data come from the World Bank: (https://data.worldbank.org/indicator/NE.TRD.GNFS.ZS). For previous periods, alternative sources by country are used.

France:

1860-1938: exports and imports come from Mitchell (2007, p. 621); GDP from Toutain (1987).[31]

1950: exports and imports, *idem*; GDP from the National Accounts (INSEE).

1960-2010: World Bank.

[31] For the years 1920 and 1938, given that data are absent due to the wars, a 3-year average cannot be computed.

Italy:
1861-1950: exports, imports and GDP come from Felice (2015), table A.3.1.
1960-2010: World Bank.

Portugal:
1861-1950: exports, imports and GDP come from Lains (2007, p. 156), table A.3.
1960-2010: World Bank.

Spain:
1860-1910: exports, imports come from Prados de la Escosura (2010, p. 210), Appendix 1; GDP from Prados de la Escosura (2003)
1920-1950: exports and imports come from Tena (2005, p. 601 and 628-631), tables 8.4 and 8.8; GDP from Prados de la Escosura (2003)
1960-2010: World Bank

(General) Government Expenditure as a % of GDP:
Computed as the share of general government expenditure in GDP, we calculate it as a 3-year average. Data for recent decades come from the OECD and can be found at https://stats.oecd.org.[32]

F*rance:*
1980-2010: OECD.
1950-1970: General Government Expenditure as a % of GDP, Flora (1983).
1870-1940: Expenditures come from Flora (1983); GDP comes from Toutain (1987).
1860: Backward projection of 1872 (Flora, 1983), based on central gov't data in Mitchell (2007).

Italy:
1980-2010: OECD.
1890-1970: Felice (2015), Table A.3.3.
1860-1880: Backward projection of 1885 (Felice, 2015), based on central gov't data in Flora (1983).

Portugal:
2000-2010: OECD.
1860-1990: Expenditure in Valério (2001, p. 702), table 9.5; GDP in Valério (2001, p. 536) table 6.6.

Spain:
2000-2010: OECD.
1960-1990: Government expenditure/GDP from Comín and Díaz (2005, p. 877), table 12.2.
1860-1950: Expenditure in Comín and Díaz (2005, p. 944), table 12.18; GDP in Prados de la Escosura (2003).

Krugman Specialization Index
For every pair of regions j and k, the KSI can be computed as:

$$KSI_{jkt} = 0.5 \sum_{s} |x_{sjt} - x_{skt}|$$

where x_{sjt} is the share of sector s in total employment of region j at time t. The indicator ranges from 0 to 1. If the two regions have a similar industrial structure the value would be close to zero, while an indicator close to one would indicate big differences in the industrial structure of the two regions. The national KSI is computed as an average value of the regional KSI. In particular, we use regional employment figures for 3 economic sectors: agriculture, manufacturing and services. These data on employment by sector are obtained from different sources:

[32] Except for France 1860-1970 and Spain 1960-1990, due to data availability.

France: Díez-Minguela and Sanchis (2017), Rosés and Sanchis (*forthcoming*).
Italy: Felice (2015), table A.4.4, data kindly supplied by the author.
Portugal: Badia-Miró et al. (2012), data kindly supplied by the authors. Data converted from historical districts to NUTS2.
Spain: Martinez-Galarraga et al. (2015), Tirado et al. (2016).

Urbanization rate
Computed as the population living in cities >100,000 inhabitants. Data come from Singer et al. (1972) and the online data set at: http://www.correlatesofwar.org/data-sets/national-material-capabilities.

Non-agrarian Gross Value Added (GVA):
France: Toutain (1987), Insee.
Italy: Felice and Vecchi (2015, 544), Table A.1.
Portugal: 1870-1950, Lains (2007, p. 149), table A.1; 1960-1990, Valério (2001, p. 520), table 6.1.b; 2000-2010: INE (https://www.ine.pt)
Spain: Prados (2016). Data set available at http://espacioinvestiga.org/bbdd-chne/?lang=en

European Historical Economics Society

EHES Working Paper Series

Recent EHES Working Papers

2017

EHES 118 Openness and growth in a historical perspective: a VECM approach
Giovanni Federico, Paul Sharp, Antonio Tena-Junguito

EHES 117 The Bank of England as Lender of Last Resort: New historical evidence from daily transactional data
Mike Anson, David Bholat, Miao Kang, Ryland Thomas

EHES 116 Harbingers of Modernity: Monetary Injections and European Economic Growth, 1492-1790
Nuno Palma

EHES 115 'Rational' Farmers and the Emergence of Modern Accounting in Danish Dairying
Markus Lampe, Paul Sharp

EHES 114 Missed opportunities? The development of human welfare in Western Europe, 1913-1950
Daniel Gallardo Albarrán

EHES.113 Portugal's wine globalization waves, 1750-2015
Pedro Lains

EHES.112 Effects of reforms and supervisory organizations: Evidence from the Ottoman Empire and the İstanbul bourse
Elmas Yaldız Hanedar, Avni Önder Hanedar and Ferdi Çelikay

EHES.111 Resource endowments and agricultural commercialization in colonial Africa: Did labour seasonality and food security drive Uganda's cotton revolution?
Michiel de Haas and Kostadis J. Papaioannou

All papers may be downloaded free of charge from: www.ehes.org
The European Historical Economics Society is concerned with advancing education in European economic history through study of European economies and economic history. The society is registered with the Charity Commissioners of England and Wales number: 1052680