Journal of Regional Science

WILEY

JOURNAL OF REGIONAL SCIENCE

RESEARCH ARTICLE OPEN ACCESS

Market Access, the Skill Premium and Human Capital in Spain (1860–1930)

Rafael González-Val[1,2] | Pau Insa-Sánchez[3] | Julio Martinez-Galarraga[4] | Daniel A. Tirado-Fabregat[5]

[1]Universidad de Zaragoza, Zaragoza, Spain | [2]Institut d'Economia de Barcelona (IEB), Barcelona, Spain | [3]University of Siena, Siena, Italy | [4]Universitat de Barcelona, Barcelona, Spain | [5]Universitat de València, Valencia, Spain

Correspondence: Julio Martinez-Galarraga (julio.martinez@ub.edu)

Received: 19 October 2022 | **Revised:** 11 June 2025 | **Accepted:** 12 August 2025

Keywords: history | human capital | market access | skill premium | Spain | wages

ABSTRACT

This paper explores the relationship between market access and human capital in the context of an industrializing economy, in this case Spain between the late 19th and early 20th centuries. Specifically, we examine whether differences in regional accumulations of human capital could be related to market access. To do this, we empirically test the relationship between education variables and market access for Spanish provinces between 1860 and 1930. We then focus on the mechanism that may be mediating this relationship, that is, the skill premium. The results suggest that there were sizeable provincial differences in the skill premium, the explanation for which would be that those provinces with the highest market access specialized more in skill-intensive sectors in which higher wages were paid.

JEL Classification: I25, N90, O15, R12

1 | Introduction

The theoretical and empirical literature of the new economic geography (NEG) has given us an economic interpretation of the spatial inequality that characterizes the world today. The basis for this is the belief that interaction between the economies of scale typical of certain activities and lower trade costs makes it more likely that companies and workers will become concentrated in those locations with the best access to the markets of goods and inputs, this being a cumulative process (Krugman 1991; Fujita et al. 1999; Puga 1999; Combes et al. 2008). The existence of this relationship would help us to understand not only the marked inequality that characterizes the distribution of activity across territories, but also the appearance of divergent growth paths within them (Krugman and Venables 1995; Redding and Venables 2004; Ottaviano and Pinelli 2006; Combes and Gobillon 2015).

Using this economic foundation as a starting point, the empirical literature has made it possible to establish two relevant aspects. First, there is evidence pointing to the existence of a relationship between market access or market potential (the two terms are used interchangeably in the context of this paper) and regional production specialization (Midelfart-Knarvik et al. 2002), the direction of capital flows (Crozet et al. 2004) and workers (Crozet 2004), and regional differences in wage levels (Hanson 1998, 2005; Roos 2001; Brakman et al. 2004) and income (Redding and Venables 2004; Breinlich 2006; Niebuhr 2006; Head and Mayer 2011). And second, these processes appear to be cumulative, which implies that the inequality that characterizes most developed economies today may have its roots in the early stages of the economic development processes (Crafts and Venables 2003; Wolf 2007; Liu and Meissner 2015). In this regard, for many of today's developed countries the 19th century was the first time the two fundamental elements of the NEG models coincided: technological change, typical of the first and second industrial revolutions, and market integration (reduction in trade costs) for goods and inputs.

The NEG literature has thus sufficiently proven that today's regional inequality is linked to differences in market access resulting from a process that has its roots in the past. However, a great deal less interest has been shown in identifying and studying the mechanisms that would become the basis of any such sustained relationship over time. In the search for these mechanisms, one body of work has connected two apparently disparate lines of literature: economic growth and economic geography. Firstly, endogenous growth theory has repeatedly pointed out that the availability of human capital is a determinant of growth over the long term (Romer 1986; Lucas 1988; Galor 2011; Barro 2013; Gennaioli et al. 2013). Thus the existence of marked differences in the regional availability of human capital could be one of the main reasons for the differences in regional economic trajectories and therefore a key element for understanding current regional inequality.

And secondly, the new economic geography literature. Since this first made its appearance, progress has been made in the theoretical formulation and empirical testing of new hypotheses. As pointed out by Redding and Schott (2003), given that better market access may encourage the accumulation of human capital, the incentive for individuals to invest in education should be related to their geographical location. The economic mechanism behind this would be as follows. Greater market access would favor regional specialization in those sectors characterized by the presence of increasing returns. In a two-sector theoretical framework in which manufacturing uses both skilled and unskilled labor and is the sector with increasing returns, and in which agriculture has constant returns to scale and uses unskilled labor for production purposes, it is to be expected that the higher the market potential, the higher the specialization in industrial production and the higher the wage premium for skilled workers. In these conditions, market potential would increase the incentives to invest in human capital in a particular location. As a result, the rate of human capital accumulation would be endogenous to the model and be related to each region's market potential.

Empirical analysis of this possible relationship has mainly involved recent periods and large samples of countries (Redding and Schott 2003) or European regions (Faíña and López-Rodríguez 2006; López-Rodríguez et al. 2007; Karahasan and López-Bazo 2013; Matas et al. 2015). The results generally indicate that levels of human capital (calculated using different measures of educational attainment) are highest in those regions with the greatest market access and decrease in line with lower market access. Thus it has been shown that market access is a factor that influences human capital accumulation in the present time. However, the study of whether or not this relationship existed during the early stages of economic development and what form it took is still in its infancy (Diebolt and Hippe 2018).

As its contribution to this area of study, this paper aims to analyze whether or not there was a relationship between market access and human capital accumulation in Spanish regions during the first major stage of the industrialization process, which in the case of Spain extends from the mid-19th century to the outbreak of the Civil War in 1936. Our research strategy is as follows. Firstly, we examine whether those provinces (NUTS3) with the greatest market potential also had the greatest accumulations of human capital, as measured by literacy rates and numeracy skills. To put it another way, we test whether or not one of the hypotheses deriving from the NEG literature can be confirmed, that is, that there is a link between a region's market potential and its accumulation of human capital (Redding and Schott 2003). Secondly, we focus on the economic mechanism that may explain this relationship, looking in particular at the connection between market potential and the skill premium, defined as the ratio of the wages of skilled industrial workers to those of unskilled workers employed in both agriculture and industry. The skill premium as a means of explaining the positive effect of market access on human capital was originally developed in the theoretical model presented by Redding and Schott (2003) and then expanded and empirically tested by Fallah et al. (2011). We follow this latter work to analyze the relationship between market access and the skill premium in Spain's provinces in the early decades of the 20th century. Finally, to complete our analysis, using semiparametric techniques we also examine the existence of possible nonlinearities between market potential, the skill premium and our human capital indicators.

Taking into account the theoretical modeling of the NEG and basing our study on a developing economy like Spain over the period 1860–1930, we aim to contribute to the literature in several ways. Firstly, in this particular case the period analyzed saw the beginning of the transition from an agrarian to an industrial economy that took place in parallel with an increasingly rapid integration of regional markets.[1] It has already been verified that industrialization did not occur homogeneously across Spain and that there was in fact a marked spatial concentration of industrial production. It could therefore be said that the evolution of the Spanish economy over the period reflects a reality that is in line with that described in the NEG theoretical literature (Martinez-Galarraga et al. 2021).[2]

Secondly, approaching the analysis of the relationship between market potential and human capital accumulation from a historical and regional perspective framed in a national case offers various advantages. The first is that it ensures that the institutional framework that regulates the educational supply by the State is common to all the regions analyzed, given that the presence of diverse education policies is an inherent limitation to those investigations that have addressed this type of analysis in cross-country studies. In addition, since this is a historical study, the time frame for the analysis begins before the Spanish central government started its direct funding of education. Regional variations in education would therefore be fundamentally related to the conditions that existed in each region and would not be the result of government policy aimed at achieving greater territorial balance.

Finally, exploring the case of Spain over the period 1860–1930 makes it possible to avoid the effects of another element that makes empirical analysis of the model proposed in Redding and Schott (2003) difficult. As López-Rodríguez et al. (2007) point out, that model assumes an absence of mobility of qualified labor, which means that testing it involves considering evidence relating to countries or regions in which few migratory flows are recorded. There was little interprovincial migration in Spain during the second half of the nineteenth and the early twentieth centuries, so we are working in a historical context in which it is reasonable to maintain the assumption of no worker mobility.[3]

Bearing these objectives and expected contributions in mind, we have assembled a provincial (NUTS3) dataset that includes a measure of market access, a series of basic human capital indicators (literacy, numeracy) and wages for different levels of skill for five benchmark years (1860, 1900, 1910, 1920, and 1930).[4] In a first stage, the results we obtain prove the existence of a relationship between provincial market potential and our human capital indicators (Redding and Schott 2003). Then, in a second stage, a mechanism that would mediate such a relationship is also identified, this being regional differences in the wage levels of qualified industrial workers compared to those of agricultural or untrained industrial workers (Fallah et al. 2011). By showing that market access has a positive effect on the accumulation of basic human capital in the early stages of economic development, the analysis therefore provides an explanation linking market access to the unequal growth trajectory of Spanish provinces over the long term that has been demonstrated by previous work (Martinez-Galarraga et al. 2015; Díez-Minguela et al. 2018).

The work is structured as follows. Section 2 provides a brief historical background describing how the location of industry and regional growth in Spain evolved over the period analyzed. Section 3 introduces the variables used and the construction of the dataset, while Section 4 shows the characteristics of the main study variables and presents preliminary non-parametric evidence of the relationship between them. Section 5 develops the empirical test to prove the existence of a relationship between market potential and basic human capital (literacy, numeracy). Then, Section 6 studies the mechanism that underlies this relationship by examining the link between market access and the skill premium. In Section 7, a semiparametric analysis is carried out to reveal possible nonlinearities in the relationship between market potential and the different dependent variables considered. Finally, Section 8 concludes.

2 | Market Access, Industry Location and Regional Inequality in Spain

The economic integration of the Spanish market was completed during the second half of the 19th century (Rosés et al. 2010). Indeed, throughout the 19th century institutional obstacles to interregional trade were being removed through a series of reforms aimed at strengthening property rights and encouraging a reduction in the trade costs that interfered with economic relations and impeded the free movement of goods within Spain's borders. Likewise, improvements to the transport system proved to be a determining factor. The construction of the railways, in particular, gave a huge boost to the integration of the Spanish market.[5] In this context, the convergence of regional prices in the grain markets was a sign that the Spanish goods market was becoming more and more integrated over time (Peña and Sánchez-Albornoz 1983). The integration of the factors markets also advanced notably. On the capital market the interregional variation in the interest rates of short-term bills of exchange decreased (Castañeda and Tafunell 1993; Nogués-Marco et al. 2019), while the reduction in interregional differentials in real wages between 1860 and 1930 has been seen as proof of the gradual integration of the labor market over that period (Rosés and Sánchez-Alonso 2004).

In addition, it is a well-established fact in Spanish economic history that from the mid-19th century to the outbreak of the Civil War (1936–1939), industrial production gradually agglomerated in a small number of provinces (Nadal 1987; Paluzie et al. 2004). Economic historians have thoroughly investigated the roots and causes of the specific location of industrial activity and of the notable increase in the spatial concentration of manufacturing over this period (Figure 1). Rosés (2003), following Davis and Weinstein (1999, 2003), finds evidence that the home market effect was behind early Catalan industrialization (around the 1860s). Tirado et al. (2002), in line with Kim (1995), identify scale economies and market size as determining Spain's industrial geography in the mid-19th century. By the end of the century, the explanatory power of these NEG effects had increased in parallel with advances in the economic integration process.

Adopting the approach developed by Midelfart-Knarvik et al. (2002), Martinez-Galarraga (2012) confirms and extends the previous findings of Tirado et al. (2002).[6] In short, it has been concluded that, as the domestic market gradually integrated and industrialization progressed in Spain during the second half of the 19th century, NEG forces grew to be the main determinant of Spain's industrial landscape. In particular, it has been proven that the interaction between economies of scale and market potential favored the concentration of industries with increasing returns in those provinces with the best access to demand up to the 1930s. In addition, Martinez-Galarraga et al. (2008), in line with Ciccone and Hall (1996) and Ciccone (2002), suggest that agglomeration economies were present in Spain's manufacturing during the early stages of industrialization, as those regions with a higher spatial density of economic activity also enjoyed higher industrial labor productivity.

Researchers have additionally tested not only the wage equation, that is, the existence of higher wages in regions with greater market potential resulting from the agglomeration of manufacturing (backward linkages), but also the ability of these wages to generate migratory flows of workers (forward linkages). On this subject and following Hanson (1998, 2005) influential research based on the Krugman wage equation, Tirado et al. (2013) examine the existence of a spatial structure in nominal wages in industry in 1920s Spain. The results confirm that wages were higher in regions with greater market potential. In addition, extending the work done by Crozet (2004), Pons et al. (2007) establish a direct relationship between migration decisions and the market potential of the host regions during the 1920s, thus verifying the presence of forward linkages in the internal migrations between Spain's provinces in the interwar years.

Beyond the manufacturing sector, Rosés et al. (2010) focus on regional inequality in terms of GDP per capita, constructing a population-weighted coefficient of variation to measure the long-term evolution of disparities in regional income per capita at provincial level. They show that, clearly echoing the growing spatial concentration of manufacturing, the second half of the

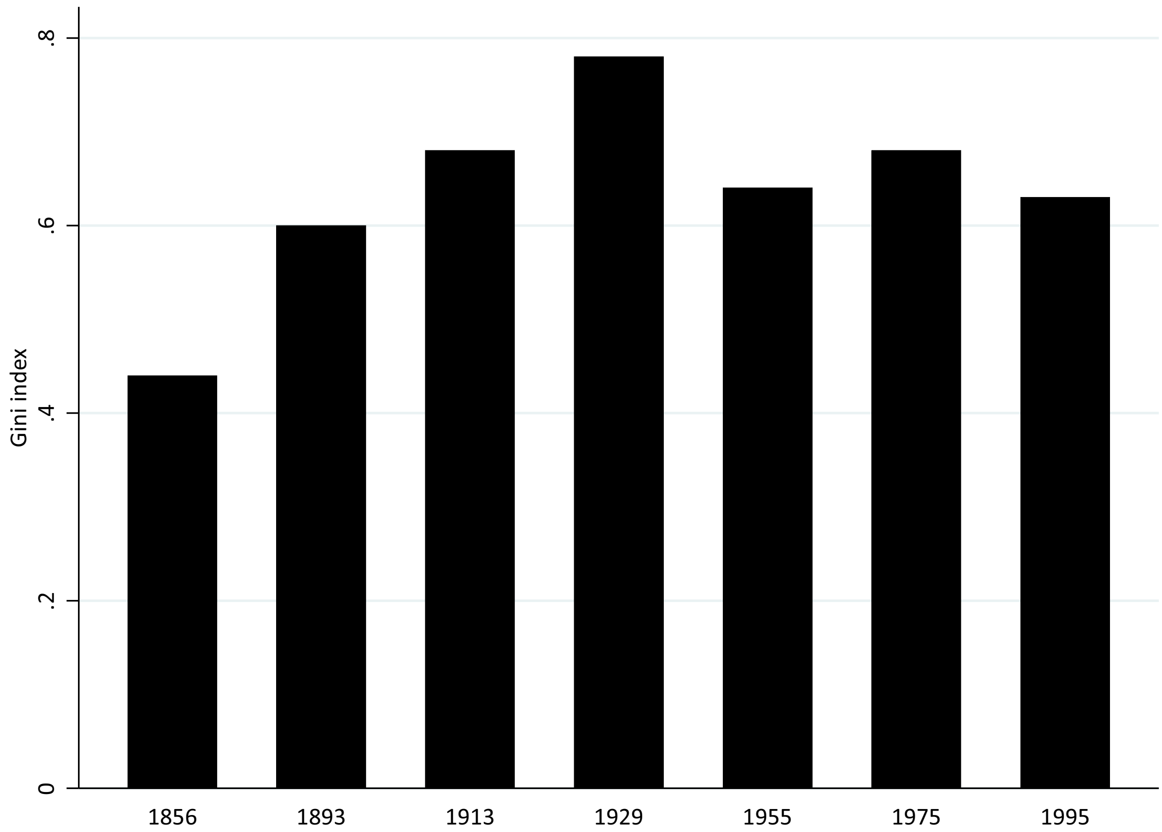


FIGURE 1 | Concentration of manufacturing in Spain, 1856–1995 (NUTS3). *Source*: Paluzie et al. (2004).

19th century and the first decades of the 20th also saw a remarkable increase in regional income inequality (Figure 2). Going one step further, Díez-Minguela et al. (2018) claim that the growth in regional inequality generated a geographical pattern that appears to be well-established in 1930 and has indeed persisted right up to the present.

Finally, Martinez-Galarraga et al. (2015), in line with Ottaviano and Pinelli (2006), show how the regional inequality that characterized the Spanish economy in 1930 was due to the existence of sizeable differences in regional growth trajectories during the period 1860–1930 and the fact that these differences were related to heterogeneous access to markets. Building on Brülhart and Sbergami (2009), Díez-Minguela et al. (2016) reinforce this result, finding that agglomeration economies are important when it comes to explaining regional inequality in Spain before the Civil War (1936–39) and that industrial agglomeration during the second half of the 19th century and the first decades of the twentieth was therefore crucial to this process.

Given these conditions, the question we aim to investigate revolves around the economic mechanism that connects the concentration of manufacturing with unequal regional economic growth. Using Redding and Schott's (2003) hypothesis as a starting point, we suggest that it was differential access to markets that determined which regions would see greater advances in industrial production. In turn, the greater or lesser industrial specialization of the regions brought about the appearance of regional differences in the wages for skilled work. Under these circumstances, the incentives to invest in education were greater in those regions with the greatest market potential, an element that ultimately gave rise to differences in regional rates of human capital accumulation. To further investigate this hypothesis, the following sections present the dataset we have constructed and some descriptive evidence regarding regional differences in market access and human capital stock.

3 | Data

With the aim of studying regional differences in human capital endowment and its determinants in greater detail, our dataset includes different human capital indicators (literacy, numeracy), a measurement for the market potential of Spanish provinces for the period 1860–1930 and both industrial and agrarian wages for 1910–1930. Other variables used as controls such as province size, measured by total population, population density and the educational infrastructure are also included. The dataset is in the form of a panel spanning five benchmark years (1860, 1900, 1910, 1920, and 1930) and covering a total of 47 Spanish provinces.[7]

Human capital is a broad, complex concept (Goldin 2016) and there are therefore different ways of approaching it (Barro and Lee 2013; Hanushek and Woessmann 2012; Abraham and Mallatt 2022). Here we focus on basic human capital and begin by calculating provincial literacy rates. Spain's population censuses provide information on self-reported literacy rates back to the mid-19th century, so we have been able to gather data for our benchmark years, broken down by sex. These data refer to

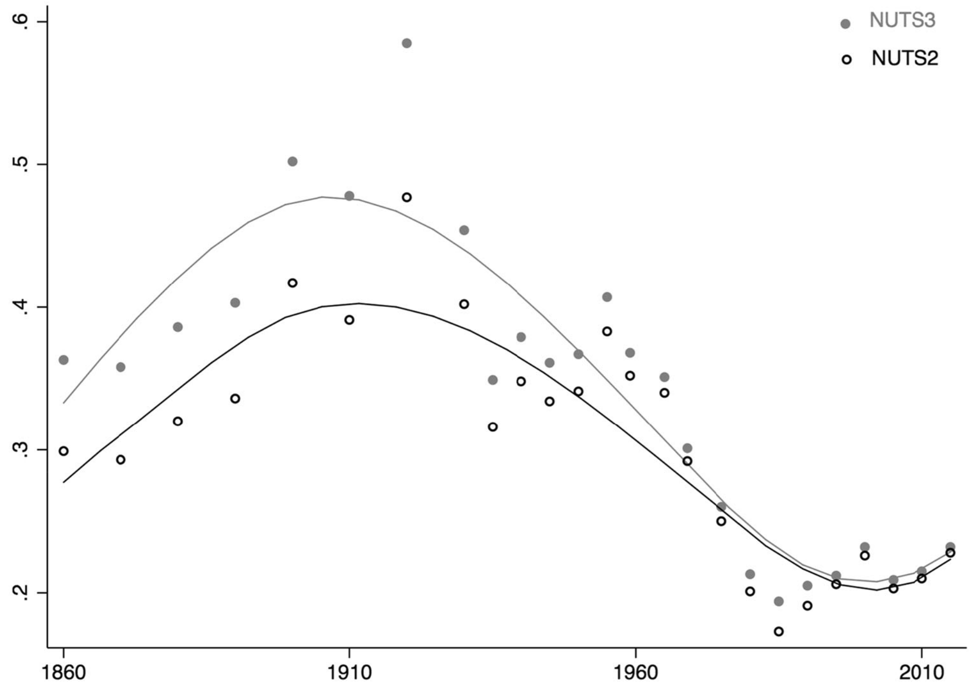


FIGURE 2 | Regional per-capita income inequality in Spain, 1860-2015. Note: Population-weighted coefficient of variation. *Source*: Díez-Minguela et al. (2018).

the percentage of the population over age 10 that could read and write. Taking the population over the age of 10 into account is common in historical studies (e.g., Núñez 2003; Diebolt and Hippe 2018), given that the ability to read and write was mainly acquired during primary education. It should be noted that these rates vary greatly according to sex and therefore the aggregate or total literacy rates actually capture two rather different stories, one for males and one for females. Thus we focus on male literacy rates and then use the female and total literacy rates for the purposes of robustness.

In addition, we also consider provincial-level numeracy indexes as a complementary measure of human capital. Numeracy skills can be defined as the ability to understand and work with numbers, so ultimately numerical knowledge and number discipline may also be crucial for industrial activities. To obtain a better picture of these basic numerical skills, a growing number of works have focused on age heaping by making use of historical data on ages (Tollnek and Baten 2016). It has been suggested that an irregular pattern of age distribution or a preference for certain digits may indicate the presence of misreported age, which can become an indicator of a society's aggregate numeracy skills. Our data come from Beltrán Tapia et al. (2022), who calculate age heaping on the basis of the year-by-year age of the population included in the censuses published since the late 19th century for the Spanish provinces, broken down by sex. It should be noted that as regards Spain the first population census that collected year-by-year age information by province was that of 1877.

The Whipple index is possibly the most popular indicator for measuring digit preference and age heaping. It can be computed as the ratio of the number of reported ages ending in 0 or 5 to a fifth of the given population, excluding the oldest and youngest individuals. It is therefore calculated for specific age cohorts (23–32, 33–42...) or for a whole range (23–62; 23–72). So, if P_{25} denotes the total number of respondents who reported an age of 25, the Whipple index for multiples of 5 for the range 23–62 years would be computed as follows:

$$W = \left(\frac{(P_{25} + P_{30} + P_{35} + P_{40} + P_{45} + P_{50} + P_{55} + P_{60})}{1/5(P_{23} + P_{24}... + P_{61} + P_{62})} \right) \cdot 100 \quad (1)$$

By definition, the index varies between 100—which indicates no preference for ages ending in 0 or 5—and 500, which implies perfect heaping. For the sake of clarity, following A'Hearn et al. (2009) the original Whipple index is converted into a new measure, the ABCC index, that ranges from 0 to 100:

$$\begin{aligned} &\text{ABCC} = \left\{1 - \frac{(\text{W} - 100)}{400}\right\} \cdot 100 \text{ for } \text{W} \geq 100; \\ &\text{ABCC} = 100 \text{ elsewhere} \end{aligned} \quad (2)$$

We use male ABCC indexes, given that the lack of gender gaps in age-heaping in Spain's provinces throws doubt on the accuracy of female age reporting (Beltrán Tapia et al. 2022).[8]

Along with measures of human capital, our analysis also relies on an indicator of provincial market access or market potential (Martinez-Galarraga 2014). The role of market accessibility has regained significance as a factor explaining the spatial distribution of economic activity thanks to the crucial role it plays within economic geography models (Fujita et al. 1999; Combes et al. 2008). Here we measure market access using the traditional Harris market potential equation (Harris 1954), also known as the nominal market potential. This captures the idea that the accessibility of a location depends on the size of other markets, usually measured in terms of GDP, once distances (or bilateral transport/trade costs) have been deducted.

This methodology has traditionally been used by geographers and economists (Clark et al. 1969; Keeble et al. 1982), although recently it has also grown more popular among economic historians, following in the steps of Crafts (2005) and his analysis of Victorian Britain. Harris's market potential function has also appeared in empirical works within the field of economic geography. The validity of this indicator in such a context is supported by two facts: first, the results obtained from empirical exercises in economic geography do not provide conclusive evidence that alternative measures might be more effective in terms of robustness (Head and Mayer 2004),[9] and second, a direct relationship can be established between Harris's equation and the market access measures deriving directly from NEG models, thus providing a theoretical foundation for Harris's market potential equation (Combes et al. 2008).

What is known as the nominal market potential or Harris's market potential function can be defined as:

$$MP_r = \sum \frac{M_s}{d_{rs}}, \tag{3}$$

where M_s is a measure of the economic size of province s (i.e., GDP) and d_{rs} is distance. In this case, d is equal to the bilateral transport costs between provinces r and s. Drawing on this expression, we computed a measure of Spanish NUTS3 market potential for the years 1860, 1900, 1910, 1920, and 1930, following Crafts (2005).[10] The figures were obtained as follows. First, the market potential of a Spanish province r was disaggregated into two components: the domestic market potential (DMP_r), which incorporated each province's self-potential (SP_r), and the foreign market potential (FMP_r) between the provincial and the international nodes. Thus the market potential of a province r (MP_r) was calculated as the sum of the domestic and foreign market potential:

$$MP_r = DMP_r + FMP_r \tag{4}$$

According to this expression, the domestic market potential for each of the 47 provinces r is calculated as follows:

$$DMP_r = \sum_{1}^{s=46} \frac{M_s}{d_{rs}} + SP_r \tag{5}$$

with

$$SP_r = \frac{M_r}{d_{rr}} \tag{6}$$

being the measure of the self-potential of each province r, where d_{rr} is calculated by taking a distance θ_{rr} equivalent to a third of the radius of a circle with an area equal to that of the province:

$$\theta_{rr} = \frac{1}{3}\sqrt{\frac{(\text{area of the province}_r)}{\pi}}. \tag{7}$$

The next expression yields the foreign market potential of province r (FMP_r), where d_{rp} captures the distance from the inland provincial node to the nearest Spanish sea port:

$$FMP_r = \sum_{1}^{f=4} \frac{M_f}{d_{rp}} \cdot Distance_{pf}^{\delta} \cdot Tariff_f^{\gamma} \tag{8}$$

with $d_{rp} = 1$ if r is a coastal province, and $d_{rp} = d_{rs}$ if r is an inland province. In this case, M_f is the economic size of the foreign market; d_{rp} captures the transport costs from the inland provincial node to the nearest Spanish sea port p; $Distance_{pf}$ is the distance between the Spanish sea port and the international node f; $Tariff_f$ is the mean tariffs applied in the foreign country f; and δ and γ are the elasticities – estimated by international trade gravity equations – associated with the coefficients for distance and tariffs respectively.

Hence the total market potential of province r (MP_r) is the sum of the following terms, the first two of which correspond to domestic market potential (including the self-potential of province r) and the third to foreign market potential:

$$MP_r = \sum_{1}^{s=46} \frac{M_r}{d_{rs}} + SP_r + \left[\sum_{1}^{f=4} \frac{M_f}{d_{rp}} \cdot Distance_{pf}^{\delta} \cdot Tariff_f^{\gamma}\right]. \tag{9}$$

The economic size of the provincial markets (M_r) is measured by the aggregate income. Data on nominal GDP at a NUTS3 level are obtained from Rosés et al. (2010). To measure d_{rs} we considered transport costs. In such a case we need data on distances and average transport rates for commodities. Internal transport was assumed to be by railway (using a network that was constantly expanding over the relevant period) and coastal shipping.[11] For railway distances, our sources were Ministerio de Obras Públicas (1902) and Wais (1987). For distances between ports, electronic atlases provided information on the length of sea journeys. As regards transport costs, data on railway freight rates came from Herranz (2005b) and coastal shipping rates in 1865 from Nadal (1975). To take into account the reduction in sea transport costs, the data were corrected using the freight rate indexes calculated by Mohammed and Williamson (2004). In the first benchmark year in the 1860s, however, only 32 out of the 47 provinces considered were connected to the railway network. Road transport was therefore also included in the internal market potential estimates for that time. Distances by road were taken from Dirección General de Obras Públicas (1861). For road transport prices we used the information in Barquín (1999). Finally, the relative weight of each transport mode in the coastal provinces came from Frax (1981).

The construction of the measure for foreign market potential includes the main trading partners of Spain: France, the United Kingdom, Germany and the United States. These four countries accounted around 60% of the total exports in the period under study (Prados de la Escosura 1982). We take their main ports as the exterior nodes (London, Hamburg and New York). However, as a consequence of its geographical location in relation to that of the Iberian Peninsula, the French market can be accessed both via the Atlantic and the Mediterranean seaboards. For this reason, the French market is divided and three nodes are considered: Le Havre and Nantes on the Atlantic seaboard and Marseille on the Mediterranean. Then, the calculation is based on the gravity equation for international trade estimated by Estevadeordal et al. (2003). The elasticities obtained for distance and tariffs (-0.8 and -1.0 respectively) were used to reduce the size of the foreign markets. The nominal GDP of the main trading partners for Spain came from Crafts (2005), based on estimates by Prados de la Escosura (2000). Prevailing exchange rates were applied to convert GDP figures from pounds to pesetas. Maritime distances were again obtained from an electronic atlas and, finally, tariffs came from O'Rourke (2000) and Mitchell (1998a, 1998b).

Another important piece of our analysis is the skill premium in wages. To calculate the skill premium between 1910 and 1930, we use wages of skilled industrial workers, unskilled industrial workers and agricultural workers.[12] With this information we obtain two alternative skill premia, one comparing the wages of industrial and agricultural workers and the other comparing the wages of skilled and unskilled workers within the industrial sector. For industrial workers data came from Ministerio de Trabajo (1927, 1931), which shows the wage per hour (in *pesetas*) for workers with different skills employed in industry in Spain at the provincial level.[13] The source consistently reports the following job categories: skilled workers (male and female), unskilled workers (without gender differentiation) and apprentices (male and female). Each of these categories covers a number of occupations that vary depending on the province and the industry.[14] Fortunately, and although for simplicity we consider average wages, the source reports wage information for different occupational categories within each task and industry.[15] Therefore, it is likely that the skill differentiation was given by the presence of specific skills and not by the execution of different tasks, which would reduce the informative potential of the skill premium calculated using wages from different occupations.

Thus, this characteristic of the source allows us to overcome two important limitations in the calculation of the skill premium (Kunst et al. 2022): the possibility of capturing changes in returns to tasks possibly non-monotonic in levels of education; and the possibility of capturing returns to skills not related to educational performance, which is a factor to be taken into account in contexts such as the one at hand, with a formal educational system that was still in a very primitive phase. Thus, the category of skilled workers included supervisory jobs or those that required specific skills, such as masters (*maestros*), supervisors (*supervisores*) or foremen (*capataces*). On the other hand, the category of unskilled workers included positions such as assistants (*ayudantes*), guards (*guardas*) or clerks (*mozos*) who, presumably, did not provide any particular qualification or specialization in the execution of the tasks. On the other hand, agricultural wages came from Bringas (2000), who collects the average male daily wages (in *pesetas*) paid in the different Spanish provinces from official sources.[16]

Being the variables described so far the main variables of interest in our work, the analysis we carry out also requires a series of variables that are used as controls. On the one hand, we take the total population of each province, which allows us to approximate labor abundance, and, on the other, we also consider population density. In both cases the demographic information was obtained from the respective Population Censuses (1860, 1900, 1910, 1920, and 1930), while the provincial area to compute population density came from the *Instituto Nacional de Estadística* (INE). Finally, we construct a variable to capture possible differences in educational infrastructure, which allows us to take into account the existing diversity in terms of primary education supply between Spanish provinces. For this purpose, the indicator chosen is the ratio of the total number of schools, both public and private, over the school-age population, for which we consider children aged between 6 and 14 years. While the population data came, once again, from the respective Population Censuses, the data referring to schools was obtained from different official education statistics.[17]

Table A1 in the Appendix provides information on these variables as well as the sources from which they are drawn.

4 | Human Capital and Market Access: Background and Preliminary Evidence

To understand the differences in education between regions we first need to examine the regulatory framework for primary education in Spain in 1860, which was the result of a long gestation process that culminated in 1857 with the passing of the Public Instruction Act (PIA), also known as the Moyano Law. This regulated the Spanish education system from 1857 to 1970, when it was replaced by the General Education Act (GEA).[18] This framework organized the education system, setting out its stages and establishing the content of each subject. It also listed the obligations that fell to families (the schooling of their children) and municipalities (the opening of schools and their conservation).

Primary education was split into elementary (*elemental*) and higher (*superior*) education (Art. 1), with a curriculum being established for both. At the same time, primary education would become "*compulsory for all Spaniards*" (Art. 7) between the ages of 6 and 9, and free for those whose parents or guardians could not afford it (Art. 9).[19] Thus the Moyano Law aimed to give structure and organization to every level of education. Nevertheless, public schools, that is, those that were "*fully or partly maintained by public funds or by religious or other similar foundations*" (Art. 97), continued to be the responsibility of local councils and religious organizations.

Certainly the provision of education infrastructures in 1860 was the responsibility of the municipalities and its funding was expected to come from these entities and families and, to a lesser extent, from secular or religious foundations. It has been argued that in this period the financial difficulties experienced at the local level would have made it difficult to produce substantial improvements in school provision (Núñez 1991). Hence the Moyano Law of 1857 could not break the secular trend, and as a result literacy levels tended to improve more in those places where they were already high in the mid-19th century (Beltrán Tapia et al. 2021).

Then, from the mid-19th century and during the early twentieth, the Spanish economy and society underwent a number of

TABLE 1 | Literacy rates (over age 10), Spain.

	Male	Female	Total
1860	41.7	11.9	26.5
1900	57.0	33.3	44.7
1910	62.7	42.2	52.0
1920	69.6	52.6	60.7
1930	81.9	65.2	73.3

Source: Population censuses.

important changes, including rapid population growth and increasing structural transformation (Pérez Moreda et al. 2015; Prados de la Escosura 2017). It was in this context of far-reaching socioeconomic change that the Royal Decree of 18 April 1900 created the Ministry of Public Instruction and Fine Arts, and this was followed 2 years later by the central government taking over the funding of primary education. Overall, during this period, primary education ended up reaching, to a greater or lesser extent, most of the territories, so that the extension of the network and of the school supply would have reduced the differences in the cost to acquire education. In relation with Redding and Schott's (2003) model, this supply-side argument would complement the main idea of their theoretical model, that is, that the demand for education at the intrahousehold level will be determined by the evolution of the skill premium and by the fact that an increase in wages in areas with greater access to markets will offer relatively higher wages for skilled workers which wll be an incentive to invest in education.

Given the virtual absence of public funding for education and the limited development of the Spanish economy, it is understandable that education levels were very low at the beginning of the period under study and that, in addition, there was much regional variation in the provision and demand for schooling (Núñez 1992). On the one hand, Table 1 shows that the male literacy rate in Spain was just 41.7% in 1860.[20] On the other, as Figure 3 illustrates, regions in the center-north of the country, such as Castile-Leon, Cantabria, the Basque Country, La Rioja and Navarre, had the highest levels of male literacy in the mid-19th century. At the other extreme were the southern regions, that is, Andalusia, Extremadura, Castile-la Mancha, Murcia and Valencia, along with those in the north-west of the peninsula, in Galicia, for example.[21]

Male literacy rates improved over the period studied. By 1900 they had increased to 57.0% for the country as a whole and would continue growing to reach 81.9% in 1930.[22] However, this improvement was not equally distributed across the country. To begin with, the spatial patterns identified for male literacy rates in 1860 simply became more pronounced during the final decades of the century. It appears that the Moyano Law of 1857 was unable to iron out the pre-existing imbalances in education infrastructure endowments. Indeed, it could be argued that the opposite was true. Regional differences were stronger or more obvious at the turn of the century and regional disparities were still sizeable in 1930. The image of literacy rates in 1900 had evolved towards a clear north-south divide, with the northern provinces moving towards universal literacy and the southern provinces continuing at levels which were sometimes below 60%.

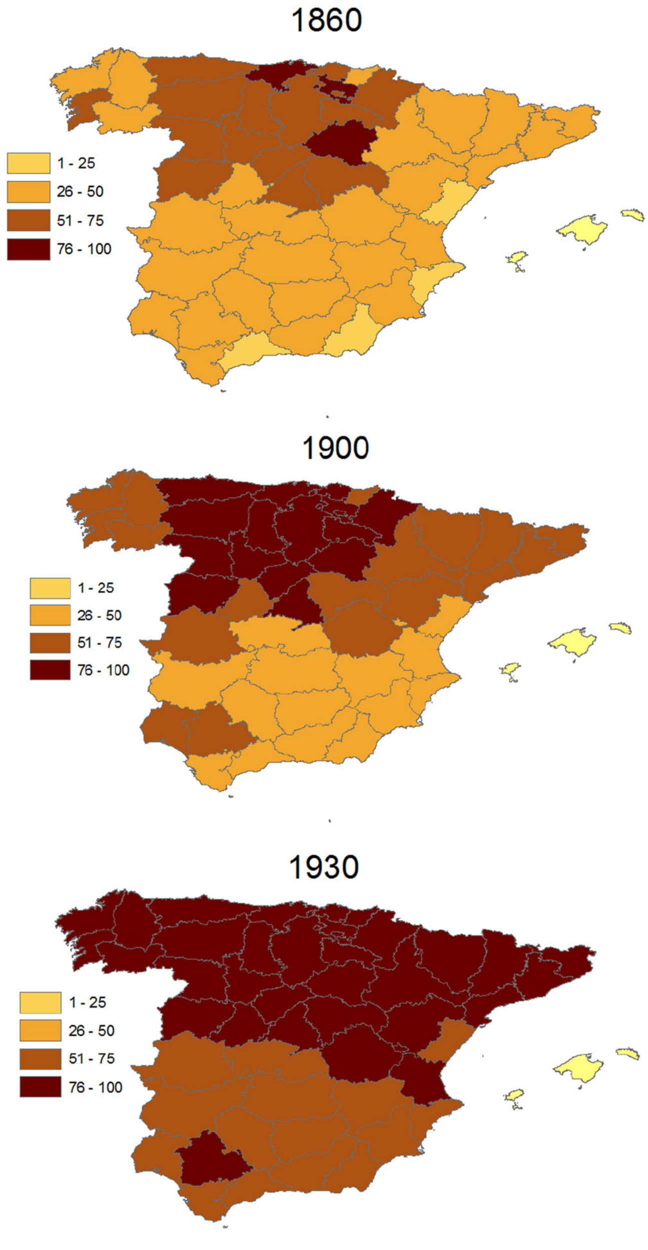


FIGURE 3 | Male literacy rates, 1860–1930. *Source*: Population censuses. [Color figure can be viewed at wileyonlinelibrary.com]

These results of the evolution of literacy show that the effectiveness of the school was different between the north and the south of Spain, which could be explained by different idiosyncratic characteristics, be they cultural, social or territorial (Beltrán Tapia and Martinez-Galarraga 2018; Reher 2023). The social attitude towards education and culture, especially in the rural contexts that were the majority in Spain at the time, has been recurrently pointed out as one of these factors. In these circumstances there could be a lack of incentives or even a rejection of school or culture in general, whether due to mere ignorance, lack of opportunities for social mobility or the opportunity cost generated by the possibility of resorting to child labor, a situation which could have generated a phenomenon of spatial inequality of opportunity in access to the educational system (Insa-Sánchez 2024). We intend to deepen into this line of argument by incorporating an economic explanation: the importance of access to markets and the evolution of wages.

In addition, the spatial patterns and evolution of the male ABCC indexes are shown in Figure 4. As explained, the first available year in this case is 1877. Generally speaking there tends to be a geographic core, namely the center and center-north, where literacy was relatively high and age-heaping practically non-existent. By 1930, the lower male ABCC indexes were still mainly concentrated in south-eastern (and north-western) Spain.

In turn, the maps in Figure 5 show the geographical pattern of regional accessibility and how it evolved from 1860 to 1930.[23] Throughout this period Barcelona had the greatest market potential, and therefore the maps are expressed in relation to this province. The evidence shows that the most significant changes in the relative accessibility of the provinces occurred in the second half of the 19th century in parallel with domestic market integration. A centrifugal tendency can be seen as the geographical structure evolved towards a clear division between inland and coastal provinces, with the latter showing a higher market potential than their inland counterparts, with the sole exception of Madrid. It could be hypothesized that the expansion of the railway network—all provincial capitals were connected to the network by 1901—was responsible for much of the change in the pattern of market potential. Once this dual structure had taken shape at the end of the 19th century, the division between inland and coastal provinces persisted over the first few decades of the 20th.

Now that all the main variables have been presented, the next step is to perform a preliminary exploratory approximation to

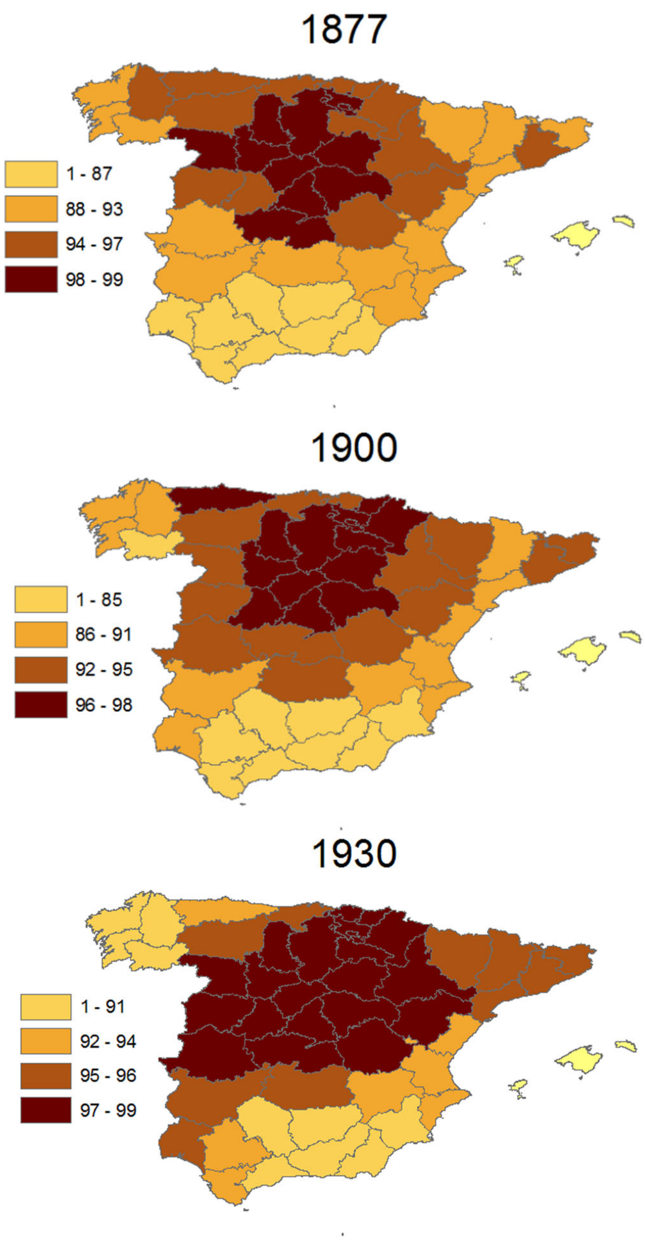


FIGURE 4 | Male ABCC index, 1877–1930. *Source*: Beltrán Tapia et al. (2022). [Color figure can be viewed at wileyonlinelibrary.com]

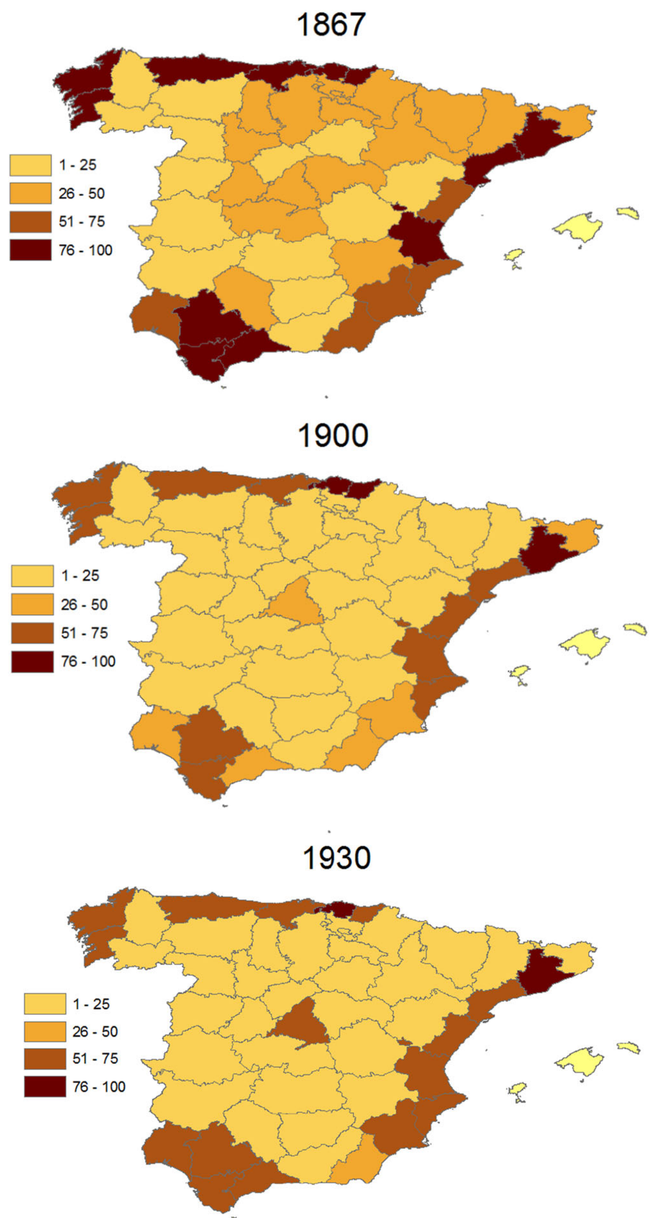


FIGURE 5 | Market potential, 1867–1930 (Barcelona = 100). *Source*: Martinez-Galarraga (2014). [Color figure can be viewed at wileyonlinelibrary.com]

discover whether there is a relationship between provincial differences in market access and the pronounced regional variability in literacy rates. First of all we present non-parametric evidence as to the existence of a relationship between the relative market potential and provincial male literacy rates (Figure 6). These graphs show how the male literacy rate varies across the market potential distribution for five different definitions of market access: total market potential, total market potential excluding self-potential, domestic market potential, domestic market potential excluding self-potential and foreign market potential.

In Figure 6, all the male literacy/market potential pairs from 1860 to 1930 are considered in a pool (taking into account the lagged market potentials in $t-1$). The graphs show a similar picture: the higher the market potential, the higher the male literacy rate. This

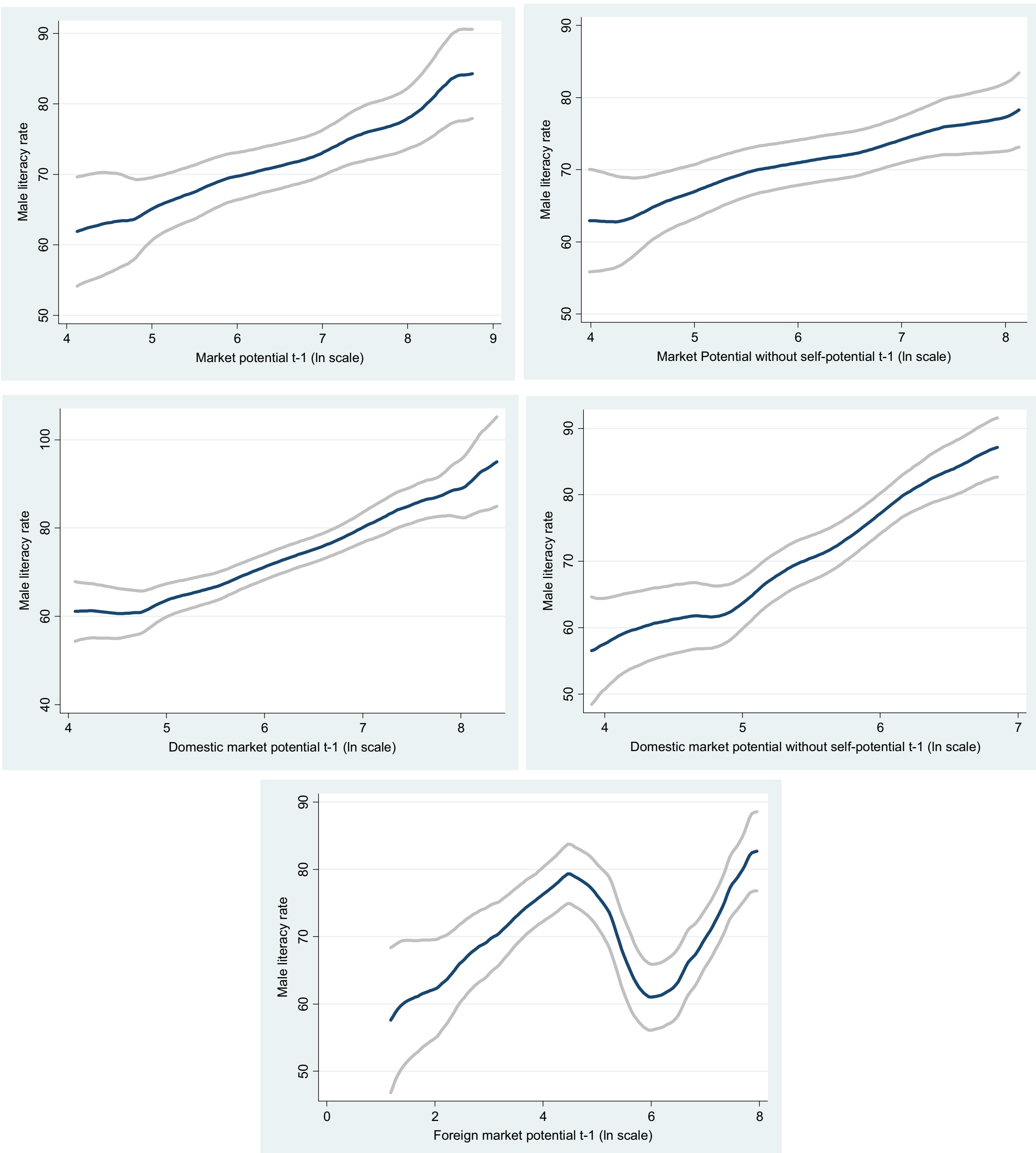


FIGURE 6 | Non-parametric relationship between male literacy rates and different measures of lagged market potential (1860–1930). *Source*: see text. [Color figure can be viewed at wileyonlinelibrary.com]

increasing relationship holds for all the market potential measures, although in the case of foreign market potential the relationship shows an inverted U-shaped pattern for some values. Overall, we can conclude that there was a positive relationship between market potential and male literacy rates in the period considered.[24] Nevertheless, this is only a preliminary descriptive analysis for illustrative purposes, as these graphs represent the unconditional relationship between literacy rates and market potential. The rest of the paper presents the results of an analysis that explores this relationship following a sounder theoretical and empirical method.

5 | Market Access and Educational Attainment

In the parametric analysis we make use of the panel structure of our data to, first, test for the existence of a relationship between market potential and male literacy rates at provincial level. Following Redding and Schott (2003) and Fallah et al. (2011), given that the proportion of the literate population is bounded between 0 and 1, we use a logistic transformation of the literacy rates. We then estimate a model in which the male literacy rate is regressed on the market potential for the previous period:[25]

$$\ln\left(\frac{literacyrate_{it}}{1 - literacyrate_{it}}\right) = \alpha + \beta \ \ln(MP_{it-1}) + \Gamma' X_{it-1} + \Pi' T_{it} + \varphi\eta_i + \varepsilon_{it}, \quad (10)$$

where our main explanatory variable is MP_{it-1}, that is, the lagged market potential. While t represents our benchmark years (1860, 1900, 1910, 1920, and 1930), i denotes the provinces ($I = 47$ provinces). As stated earlier, we consider five alternative measures of market potential corresponding to its different components: total market potential, domestic market potential, foreign market potential and a version of total market potential and domestic market potential that exclude each province's self-potential. By excluding the self-potential from the market potential variables we seek to reduce some endogeneity concerns.

Then, X_{it-1} is a vector of additional (lagged) control variables at provincial level that represent factors that could attract skilled individuals or impose restrictions on the provision of primary education (Fallah et al. 2011). As explained in the data section, we consider both the province's total population (as a measure of labor abundance) and its population density. Both variables are obtained from the respective population censuses. On this subject, the historical literature on primary education in Spain has shown that there is a positive link between population size and education levels, pointing to a relationship between the legal framework of school supply, which favored its spread across more populated areas, and the presence of economies of scale in education provision (Frago 1990; Pérez Moreda 1997; Reher 1997; Núñez 2003). It has also been suggested that the costs associated with primary education are higher in rural contexts with sparse populations, where literacy rates are consistently lower, than in urban contexts, where there would be more demand for skilled work and the provision costs would be lower. To capture differences in education infrastructure between provinces, we also include as a control variable the ratio of the total number of schools, both public and private, over the school-age population (6–14 years) (Díez-Minguela et al. 2023).[26] Finally, η_i denotes provincial fixed effects (FE) to control for other provincial characteristics not accounted for in the specification (e.g., first-nature causes and geography) and T_{it} is a set of province-specific time trends ($\sum_{i=1}^{n-1} \text{Province}_i \cdot Time_t$) that capture the particular behavior of the provinces in our panel over time[27]. ε_{it} is the error term. Robust standard errors are clustered by province. Table 2 shows the descriptive statistics for the variables used in this study.

Some endogeneity concerns may arise because an important component of Harris' market potential function is the contribution of its own GDP to the potential of region i (self-potential) and, as mentioned above, the provision of education infrastructures in 1860 was the responsibility of the local authorities. It could therefore be argued that those regions with higher GDP (and higher market potential) would spend more on education and thus have higher literacy rates. To address this, first, all explanatory variables are lagged one period to mitigate any direct endogeneity between the dependent and independent variables. In addition, as explained, we consider our market potential measures excluding self-potential, which should avoid any possible simultaneity problems. Finally, to tackle any potential endogeneity concern we estimate Equation (10) using both panel FE and instrumental variables (IV).

Columns 1–5 in Table 3 report the results of the panel FE estimates. The coefficient of market potential is positive and significant at a 1% level for all five different measures of market potential, indicating that greater market potential is associated with higher literacy rates. Moreover, the estimated coefficients across the alternative definitions of market potential are quite similar, which suggests a consistent relationship between market potential and education. Columns 6–10 of Table 3 present the IV (2SLS) results.

To deal with any potential endogeneity issues, we need to find variables that were not influenced by the basic literacy skills of the population but which did have an impact on market potential. The instruments considered for the market potential variable are the infrastructure stock (Herranz 2005b) and knowledge capital, measured as the number of patents per capita (Sáiz 2005), both in logs. The infrastructure stock is measured in 1890 constant pesetas and includes different kinds of infrastructure: railways, tramways, roads, ports, telegraph and telephone, energy distribution networks, urban infrastructure and hydraulic works. Several of these, particularly those related to transport, have a clear impact on market potential through transport costs. None, however, are directly linked to the education system. Notably, the infrastructure stock considered here focuses exclusively on economic infrastructure, deliberately excluding social infrastructure—that is, assets that enhance social welfare, such as educational and health facilities (Herranz 2005b).

One might argue that the expansion of infrastructure could also influence human capital accumulation through mechanisms beyond market access, such as capital flows and internal migration between regions. However, historical evidence indicates that infrastructure growth was not directly associated with increased capital flows or regional migration. Prados de la Escosura and Rosés (2010) identified distinct phases of capital accumulation over the long term: an intense expansion from the 1850s to the early 1880s, followed by a deceleration until World War I. Capital accumulation then resumed briskly in the 1920s, only to be disrupted by the early 1930s. Rather than infrastructure

TABLE 2 | Descriptive statistics.

Variable	Mean	Standard deviation	Minimum	Maximum
Male literacy rate	65.90	21.91	21.67	99.75
Total literacy rate	54.19	22.68	13.51	99
ABCC index	92.81	4.73	79.15	99.30
Market potential	1105.30	1088.61	61.78	6358.05
Market potential without self-potential	867.95	834.89	53.92	3394.69
Domestic market potential	609.45	513.82	58.52	4293.33
Domestic market potential without self-potential	372.10	235.46	49.66	1160.18
Foreign market potential	495.85	754.83	3.26	2851.61
Labor abundance	405884.50	228507.30	96385	1800638
Population density	47.69	36.91	12.52	233
Educational infrastructure per 1,000 inhabitants	9.14	4.09	3.24	20.28
Infrastructure stock	71.35	46.48	4.69	384.44
Patents per capita	79.67	146.53	0.41	1000.58
Share of manufacturing in total employment	0.12	0.07	0.02	0.52
Share of population over regional population	0.32	0.29	0.06	1
Industrial skilled wage	0.71	0.24	0.33	1.33
Industrial unskilled wage	0.47	0.16	0.20	0.95
Agrarian wage	0.40	0.19	0.13	1.08

Source: See text and Table A1 in the Appendix.

improvements, some scholars have pointed to institutional factors—such as the liberal regulatory framework and varying monetary policy regimes—as key determinants of capital flows in the Spanish context (Betrán and Pons 2022). With regard to migration, Rosés and Sánchez-Alonso (2004) argued that internal migration played a limited role in Spain until the early 20th century, consistent with the broader pattern observed in Southern European, Latin countries. It was not until the 1920s that a significant rise in internal migration occurred, largely driven by economic forces (Silvestre 2005; Pons et al. 2007).[28]

The patents per capita variable, measured as the number of patents per 100,000 inhabitants, is a proxy for the stock of knowledge capital, which is a production factor that directly affects regional performance, usually through productivity gains (Ottaviano and Pinelli 2006). This is a valid instrument for us to use because of the strong external dependence of the Spanish innovation system. Therefore, foreign patents dominated throughout the period, so the impact of changes in national literacy rates should not be significant. Using Spanish patent records from 1882 to 1935, Ortiz-Villajos (2005) finds that Spanish applicants obtained roughly 30% of all the patents granted in 1882 and 54% in 1935. Individual applicants were the most numerous patentees, although companies increased their share from 8.2% of all patents in 1882 to 41.6% in 1935. Of these business patents, a large proportion (68.2%) went to foreign companies. This makes sense given Spain's technological backwardness compared to other European countries: 19th-century Spain had comparatively low levels of education and technology (Ortiz-Villajos 2005).

These two variables are used to instrument the five definitions of market potential. Table 3 also shows some statistics from the first- and second-stage regressions that indicate that our instruments perform well, since the F-test takes high values and all models pass the overidentification test (Hansen J statistic) for any level of significance. The weak instrument hypothesis is rejected using the Stock–Yogo test and the model passes the underidentification test at any significance level in all cases.[29] Furthermore, the main result holds in the IV regressions, given that the estimated coefficients of the market potential variables remain positive and significant, although they are slightly higher than those obtained in the panel FE regressions.

We carry out two robustness checks. First, to consider possible gender differences in education levels, we estimate Equation (10) using the total (including both males and females) instead of just the male literacy rate. Table 4 reports the results. Both the panel FE and the IV estimations yield very similar results to those obtained using the male literacy rate. Again, the effect of market potential on education is positive and consistent across the different definitions. The estimated coefficients are almost identical to those reported in Table 3. Furthermore, results unreported here show that, when considering the female literacy rate as the dependent variable, the coefficients are closely aligned with those reported in Tables 3 and 4.

Second, we consider an alternative measure of basic human capital that potentially captures numeracy levels based on age-heaping methodology, as mentioned earlier (Beltrán Tapia et al. 2022). We then rerun Equation (10) using numeracy (ABCC index for males) instead of literacy rates as the dependent variable. Table 5 shows the results. Again, we find a positive and significant effect of market potential on education that is consistent across the different definitions. The main differences are that, first, the magnitude of the

TABLE 3 | Effects of market potential on male literacy rates.

	Panel FE					IV (2SLS)				
Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Market potential	1.215***					1.306***				
	(0.109)					(0.133)				
Market potential without self-potential		1.206***					1.277***			
		(0.104)					(0.134)			
Domestic market potential			1.272***					1.435***		
			(0.117)					(0.145)		
Domestic market potential without self-potential				1.291***					1.444***	
				(0.114)					(0.147)	
Foreign market potential					1.063***					1.012***
					(0.098)					(0.105)
Labor abundance	−1.462**	−1.364**	−1.252**	−0.959	−0.537	−1.564***	−1.388***	−1.529***	−1.229***	−0.824***
	(0.563)	(0.564)	(0.521)	(0.579)	(0.586)	(0.260)	(0.247)	(0.262)	(0.246)	(0.208)
Population density	1.567***	1.616***	1.714***	1.881***	0.701	1.234**	1.309**	1.293**	1.562***	1.264*
	(0.480)	(0.539)	(0.467)	(0.514)	(0.641)	(0.591)	(0.634)	(0.573)	(0.595)	(0.679)
Educational infrastructure	−0.354	−0.442	−0.140	−0.155	−0.140	−0.483	−0.547	−0.333	−0.335	−0.068
	(0.665)	(0.755)	(0.602)	(0.710)	(0.810)	(0.376)	(0.409)	(0.335)	(0.369)	(0.404)
Provincial fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.958	0.956	0.959	0.957	0.956					
Observations	188	188	188	188	188	188	188	188	188	188
First stage, Shea partial R^2						0.455	0.486	0.402	0.424	0.603
First stage, F-test (p-value)						33.65 (0.000)	37.90 (0.000)	26.09 (0.000)	27.18 (0.000)	62.86 (0.000)
Underidentification test (Kleibergen-Paap rk LM statistic), p-value						0.000	0.000	0.000	0.000	0.000
Weak identification test (Kleibergen-Paap rk Wald F statistic)						33.651	37.904	26.091	27.181	62.860
Hansen J statistic, p-value						0.981	0.896	0.527	0.268	0.641
Uncentered R^2						0.976	0.976	0.977	0.976	0.975

Note: Dependent variable: logistic transformation of the male literacy rate. Coefficient (robust standard errors). Robust standard errors clustered by province in panel FE regressions. All variables in logarithmic scale and all models include a constant. Significant at *10%, **5% and ***1% levels. Instruments in IV regressions: infrastructure stock and knowledge capital, both in logs.

TABLE 4 | Effects of market potential on total literacy rates.

	Panel FE					IV (2SLS)				
Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Market potential	1.174***					1.279***				
	(0.115)					(0.081)				
Market potential without self-potential		1.193***					1.251***			
		(0.116)					(0.078)			
Domestic market potential			1.209***					1.409***		
			(0.118)					(0.089)		
Domestic market potential without self-potential				1.269***					1.421***	
				(0.119)					(0.082)	
Foreign market potential					1.035***					0.994***
					(0.109)					(0.073)
Labor abundance	−1.663***	−1.634***	−1.414***	−1.218***	−0.782*	−1.477***	−1.305***	−1.446***	−1.153***	−0.753***
	(0.464)	(0.450)	(0.429)	(0.428)	(0.446)	(0.257)	(0.240)	(0.250)	(0.220)	(0.236)
Population density	0.898***	0.884***	1.083***	1.161***	0.037	0.213	0.285	0.259	0.513*	0.233
	(0.318)	(0.320)	(0.316)	(0.307)	(0.445)	(0.344)	(0.327)	(0.365)	(0.310)	(0.368)
Educational infrastructure	−0.413	−0.538	−0.182	−0.246	−0.216	−0.562**	−0.625***	−0.419**	−0.425**	−0.157
	(0.383)	(0.386)	(0.344)	(0.346)	(0.410)	(0.227)	(0.228)	(0.211)	(0.200)	(0.223)
Provincial fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.974	0.975	0.973	0.975	0.972					
Observations	188	188	188	188	188	188	188	188	188	188
First stage, Shea partial R^2						0.455	0.486	0.402	0.424	0.603
First stage, F-test (p-value)						33.65 (0.000)	37.90 (0.000)	26.09 (0.000)	27.18 (0.000)	62.86 (0.000)
Underidentification test (Kleibergen-Paap rk LM statistic), p-value						0.000	0.000	0.000	0.000	0.000
Weak identification test (Kleibergen-Paap rk Wald F statistic)						33.651	37.904	26.091	27.181	62.860
Hansen J statistic, p-value						0.479	0.562	0.952	0.444	0.934
Uncentered R^2						0.979	0.980	0.977	0.979	0.978

Note: Dependent variable: logistic transformation of the total literacy rate. Coefficient (robust standard errors). Robust standard errors clustered by province in panel FE regressions. All variables in logarithmic scale and all models include a constant. Significant at *10%, **5% and ***1% levels. Instruments in IV regressions: infrastructure stock and knowledge capital, both in logs.

TABLE 5 | Effects of market potential on numeracy levels (ABCC index, males).

Variables:	Panel FE					IV (2SLS)				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Market potential	0.436***					0.237***				
	(0.070)					(0.062)				
Market potential without self-potential		0.438***					0.232***			
		(0.072)					(0.061)			
Domestic market potential			0.450***					0.264***		
			(0.071)					(0.068)		
Domestic market potential without self-potential				0.466***					0.268***	
				(0.075)					(0.068)	
Foreign market potential					0.340***					0.186***
					(0.061)					(0.050)
Labor abundance	−0.474*	−0.450*	−0.384	−0.299	−0.060	−0.106	−0.074	−0.102	−0.048	0.028
	(0.256)	(0.259)	(0.251)	(0.236)	(0.230)	(0.091)	(0.087)	(0.090)	(0.083)	(0.077)
Population density	0.012	0.019	0.079	0.120	−0.154	0.592**	0.604**	0.593**	0.634***	0.590**
	(0.197)	(0.187)	(0.203)	(0.190)	(0.229)	(0.254)	(0.251)	(0.253)	(0.242)	(0.264)
Educational infrastructure	0.094	0.055	0.179	0.162	0.230	0.376**	0.363**	0.399**	0.396**	0.449***
	(0.230)	(0.237)	(0.226)	(0.220)	(0.237)	(0.159)	(0.163)	(0.155)	(0.157)	(0.157)
Provincial fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.965	0.965	0.965	0.965	0.958					
Observations	188	188	188	188	188	188	188	188	188	188
First stage, Shea partial R^2						0.455	0.486	0.402	0.424	0.603
First stage, F-test (p-value)						33.65 (0.000)	37.90 (0.000)	26.09 (0.000)	27.18 (0.000)	62.86 (0.000)
Underidentification test (Kleibergen-Paap rk LM statistic), p-value						0.000	0.000	0.000	0.000	0.000
Weak identification test (Kleibergen-Paap rk Wald F statistic)						33.651	37.904	26.091	27.181	62.860
Hansen J statistic, p-value						0.378	0.405	0.490	0.621	0.500
Uncentered R^2						0.997	0.997	0.997	0.997	0.997

Note: Dependent variable: logistic transformation of the numeracy levels (ABCC index, males). Coefficient (robust standard errors). Robust standard errors clustered by province in panel FE regressions. All variables in logarithmic scale and all models include a constant. Significant at *10%, **5% and ***1% levels. Instruments in IV regressions: infrastructure stock and knowledge capital, both in logs.

estimated effect is smaller given that the estimated coefficients for the market potential variables are lower than those obtained in Table 3, and second, in this case there are substantial differences between the panel FE and IV estimates, since the IV estimated coefficients (columns 6–10) are approximately half the panel FE estimates (columns 1–5). Nevertheless, although the magnitude of the effect is different, the main result holds, given that we obtain a positive and statistically significant effect of market potential in all cases.

6 | Exploring the Economic Mechanism: Market Access and the Skill Premium

Redding and Schott (2003) theoretical model shows the existence of a positive relationship between market potential and human capital accumulation, in which the mechanism that would explain that relationship would be the skill premium. This aspect of the Redding and Schott (2003) model is expanded by Fallah et al. (2011), who provide an NEG model to explain wage differentials using market access. Their theoretical framework establishes a structural relationship between nominal skill premium and market access given the location of economic activities. Intuitively, a region's improvement in market access would increase the size of production sectors characterized by monopolistic competition, thus increasing the demand for skilled workers in those sectors. The other sectors of the economy, which operate in a context of perfect competition, cannot release more skilled workers. Thus the wages of skilled workers in the monopolistic competition sectors increase while unskilled workers' wages decrease, bringing about an increase in the skill premium. Considering that both these sectors of the economy are located in different spatial settings and characterized by different economies of scale and transportation costs, the movement of workers from one to another, which according to the model would be caused by a positive shock to market access, ultimately provides a theoretical mechanism to explain the spatial location of economic activity.

While Fallah et al. (2011) tested this theoretical mechanism using a cross-section of US cities in 2001, here we take advantage of our historical dataset to estimate a panel data model, although the sample period (1910–1930) is shorter than in previous sections because of limitations in wage data availability.[30] As explained in the data section, we use data for skilled industrial workers (*obreros cualificados*), unskilled industrial workers (*peones*) and agricultural workers (Ministerio de Trabajo 1927, 1931; Bringas 2000). Figure 7 shows the skill premium between skilled industrial workers and agricultural workers (skill premium 1), the most closely related measure to Redding and Schott's (2003) theoretical model, while Figure 8 shows the skill premium between skilled and unskilled industrial workers (skill premium 2).

At first glance the maps in Figures 7 and 8 present a complex picture. On the one hand, it is difficult to detect a clear geographical pattern in the spatial distribution of the skill premium, while on the other, extracting any conclusions from the changes observed between 1910 and 1930 does not appear straightforward. Therefore, to further examine this regional variation and assess the relationship between market potential and the skill premium, we estimate the following panel data model:

$$\ln(wage\ skilled_{it}) - \ln(wage\ unskilled_{it}) = \alpha + \beta \ln(MP_{it-1}) + \Gamma' X_{it-1} + \Pi' T_{it} + \varphi \eta_i + \varepsilon_{it} \quad (11)$$

where the term on the left-hand side is the wage differential across two different skill groups (skilled and unskilled) measured at provincial level, and the variables on the right-hand side are those used in Equation (10). Again, the main explanatory variable is the lagged market potential (MP_{it-1}), for which we use the five different definitions, and X_{it-1}, η_i and T_{it} are respectively the vector of additional (lagged) control variables, the provincial fixed effects and a set of province-specific time trends.

While Fallah et al. (2011) reported only OLS estimates, endogeneity remains a potential concern in this context. To mitigate this issue, all explanatory variables are lagged by one

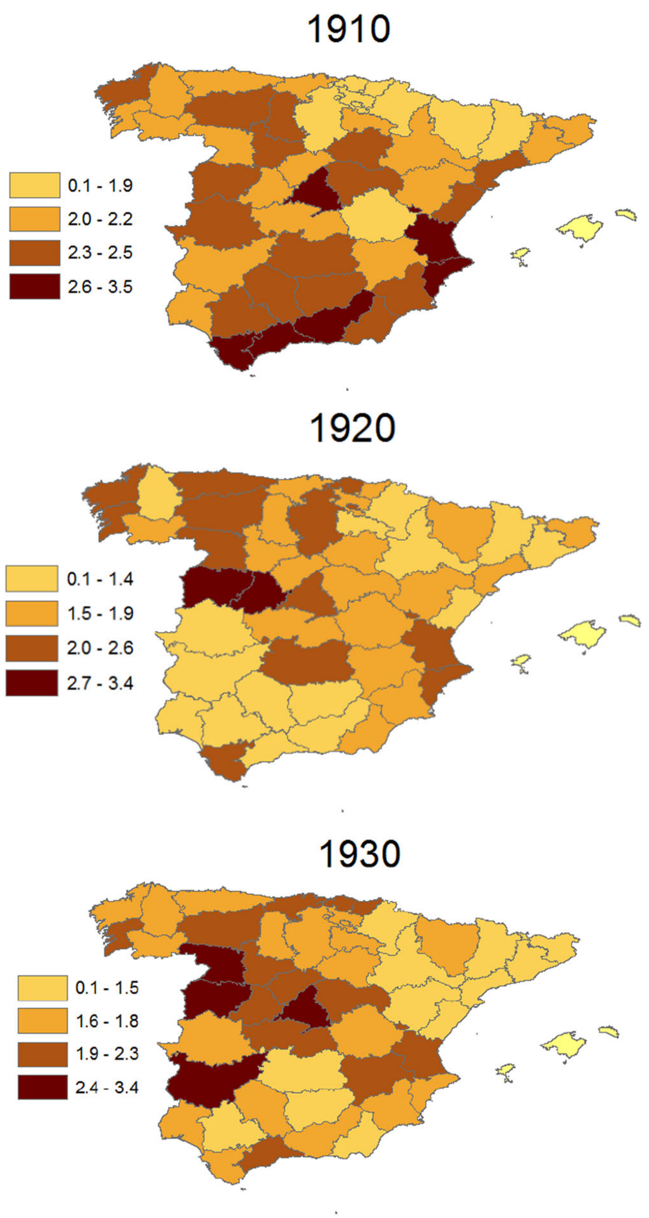


FIGURE 7 | Skill premium 1: skilled industrial workers v. agricultural workers. *Source:* Ministerio de Trabajo (1927, 1931) and Bringas (2000). [Color figure can be viewed at wileyonlinelibrary.com]

period, and we also consider alternative versions of the market potential measure that exclude self-potential. Nevertheless, to more directly address potential endogeneity, we estimate Equation (11) using instrumental variables (2SLS). However, the instruments employed in the previous analysis are not suitable here, as both infrastructure stock and knowledge capital may influence the skill premium via their effects on labor productivity. Consequently, it is necessary to identify instruments that affect market potential but are not themselves influenced by the skill premium. The instruments used for the market potential variable are the share of manufacturing in total employment and the ratio of provincial population to the population of the corresponding NUTS2 region.

Population—whether in relative or absolute terms—is frequently used as a proxy for market potential (Black and Henderson 2003; González-Val et al. 2017). The importance of the manufacturing sector is also clearly related to market potential. However, whether these two variables are independent of the skill premium is open to debate. Two key aspects of Spanish economic history, however, help justify their use as instruments: the limited internal migration affecting population distribution until the 1920s, and the dominance of the agricultural sector, which reduces the potential impact of shifts in manufacturing employment on the skill premium.

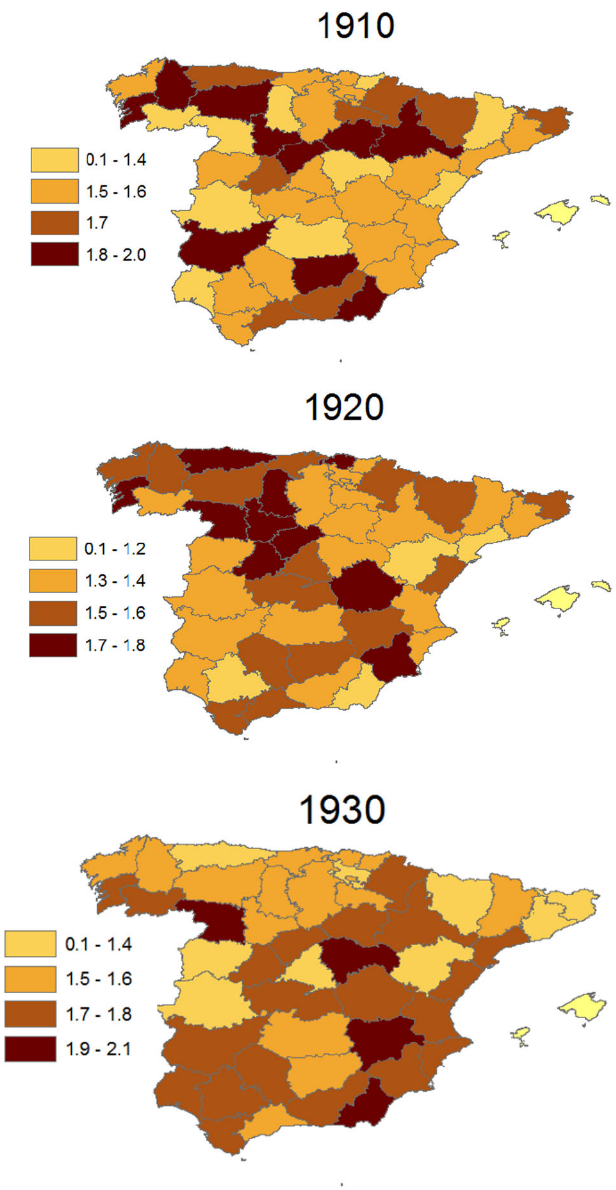


FIGURE 8 | Skill premium 2: skilled industrial workers v. unskilled industrial workers. *Source:* Ministerio de Trabajo (1927, 1931). [Color figure can be viewed at wileyonlinelibrary.com]

As noted earlier, interregional migration rates remained low in Spain until the 1920s. Rural-urban wage differentials were modest, a factor that may have prevented the emergence of higher migration rates. Even after the increase in internal migration during the 1920s, its effect on wage convergence has been shown to be insignificant (Rosés and Sánchez-Alonso 2004; Silvestre 2005). In terms of employment distribution across sectors, agriculture remained dominant well into the 20th century. As late as 1910, approximately two-thirds of the labor force was employed in agriculture (Harrison 1990), while manufacturing accounted for only about 9% of total employment (Tirado and Martinez-Galarraga 2008). Although industrialization gained momentum in the early 20th century, industrial activity became increasingly concentrated in just two regions—Catalonia and the Basque Country—while central and northwestern regions underwent progressive deindustrialization. These two industrial regions alone accounted for over 61% of total employment in key industrial sectors such as metallurgy, engineering, chemicals, and textiles (Tirado et al. 2002; Rosés 2003). As a result, excluding Catalonia and the Basque Country, most regions exhibited low values for our instrumental variable and experienced limited variation over time.

Table 6 reports the results of the panel IV estimations for both skill premia. Formal overidentification tests (Hansen test) support the validity of the instruments.[31] Furthermore, the Stock–Yogo test rejects the weak instrument hypothesis, and the model consistently passes the underidentification test at all conventional significance levels. We obtain a positive and significant effect of market potential (regardless of its definition) on the wage differential between skilled and unskilled workers. In other words, greater market access was associated with higher wages for skilled workers compared to unskilled workers. Furthermore, the magnitude of this effect is consistent across the different versions of market potential, though it is larger and more statistically significant in the case of skill premium 1.

Since the variables are in logarithmic scale, the estimated coefficients can be interpreted as elasticities. Specifically, a 1% increase in market potential is associated with a 0.375% increase in skill premium 1 (column 1) and a 0.115% in skill premium 2 (column 6). Alternatively, the estimates imply that a one standard deviation increase in market potential is associated with a 0.274 increase in the log wage gap between skilled industrial and agricultural workers (equivalent to a 1.03 standard deviation increase in the dependent variable), and a 0.084 log point increase (or approximately 8.4%) in the wage gap between skilled and unskilled industrial workers (equivalent to a 0.73 standard deviation increase). These results suggest that market potential had a substantially stronger effect on the skill premium in historical Spain than what Fallah et al. (2011) found for modern US metropolitan areas (more than double the

TABLE 6 | Effects of market potential on the skill premium, IV (2SLS) results.

	Wage skilled industrial workers v. wage agricultural workers					Wage skilled v. wage unskilled industrial workers				
Variables:	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**
Market potential	0.375***					0.115**				
	(0.093)					(0.049)				
Market potential without self-potential		0.412***					0.122**			
		(0.101)					(0.054)			
Domestic market potential			0.333***					0.103**		
			(0.089)					(0.044)		
Domestic market potential without self-potential				0.385***					0.115**	
				(0.094)					(0.050)	
Foreign market potential					0.506***					0.140*
					(0.151)					(0.075)
Labor abundance	0.971***	1.022***	0.915**	0.961***	1.192***	0.101	0.115	0.085	0.098	0.158
	(0.350)	(0.350)	(0.359)	(0.353)	(0.383)	(0.192)	(0.193)	(0.192)	(0.193)	(0.205)
Population density	0.406	0.520	0.460	0.570	−0.112	−0.554	−0.519	−0.538	−0.504	−0.694
	(0.765)	(0.766)	(0.778)	(0.773)	(0.787)	(0.488)	(0.490)	(0.493)	(0.497)	(0.503)
Educational infrastructure	0.088	0.066	0.094	0.057	0.055	0.138	0.136	0.138	0.131	0.143
	(0.248)	(0.249)	(0.259)	(0.251)	(0.271)	(0.104)	(0.105)	(0.105)	(0.105)	(0.112)
Provincial fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Province x Time	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	141	141	141	141	141	141	141	141	141	141
First stage, Shea partial R^2	0.793	0.752	0.829	0.757	0.762	0.793	0.752	0.829	0.757	0.762
First stage, F-test (p-value)	75.33 (0.000)	63.68 (0.000)	138.02 (0.000)	81.93 (0.000)	42.76 (0.000)	75.33 (0.000)	63.68 (0.000)	138.02 (0.000)	81.93 (0.000)	42.76 (0.000)
Underidentification test (Kleibergen-Paap rk LM statistic), p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Weak identification test (Kleibergen-Paap rk Wald F statistic)	75.330	63.680	138.016	81.927	42.764	75.330	63.680	138.016	81.927	42.764
Hansen J statistic, p-value	0.657	0.397	0.863	0.459	0.154	0.432	0.324	0.507	0.350	0.178
Uncentered R^2	0.970	0.970	0.969	0.970	0.968	0.981	0.981	0.981	0.981	0.981

Note: Dependent variable: ln(wage skilled industrial workers)—ln(wage agricultural workers) in columns 1–5, ln(wage skilled industrial workers)—ln(wage unskilled industrial workers) in columns 6–10. Coefficient (robust standard errors). Robust standard errors clustered by province in panel FE regressions. All variables in logarithmic scale and all models include a constant. Significant at *10%, **5% and ***1% levels. Instruments in IV regressions: share of manufacturing in total employment and share of the provincial population over regional (NUTS2 regions) population.

relative impact). This likely reflects stronger mobility constraints, higher returns to industrial skills, and more segmented labor markets in early twentieth-century Spain.

These results thus show the existence of a positive effect of market access on the skill premium, as predicted by Redding and Schott's (2003) model and empirically tested by Fallah et al. (2011) in their study at the level of US cities in 2001. However, once the existence of this effect has been proven for Spain in the early stages of economic development, there continues to be a missing link. To close the circle, we still need to analyze whether the higher skill premium in those areas with better access to markets in turn translated into a greater accumulation of human capital. In our case, testing for the existence of this relationship between the skill premium and literacy rates is complicated. On the one hand this is because data availability restricts our period of study to only two decades (1910–1930)—which means we do not have a sufficient number of observations—and on the other, because human capital accumulation is not a rapid process and takes time.[32]

7 | Semiparametric Analysis

Given the difficulty of estimating the relationship between the skill premium and human capital accumulation parametrically, as in previous exercises, on this occasion the study follows another approach based on a semiparametric analysis to examine possible nonlinearities in the relationship between market potential and the skill premium, but also between market potential and the human capital indicators. A parametric specification, such as those in Equations (10) and (11), 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 provinces (the relationship is restricted to being stationary over the entire structure of the distribution of the market potential variable). This assumption can be problematic if the effect of market potential varies along the distribution, a possibility that cannot be ruled out a priori. For example, Fallah et al. (2011) found that in US metropolitan areas, market potential primarily increased the wage premium between medium- and low-skilled workers, with effects concentrated in the lower half of the wage distribution.

To overcome these limitations, Durlauf (2001) suggested the use of semiparametric methods. This approach allows us to tackle the possible nonlinear effect of the market potential on the different dependent variables in a more flexible way. For instance, the standard correlation index and the coefficients from parametric regressions give us only an aggregate average relationship between variables, and this relationship is restricted by the fact that it must remain unchanged through the entire distribution of the market potentials. In contrast, the semiparametric estimate allows the relationship to vary with market potential over the entire distribution, allowing for the linear effects of other conditioning variables.

We perform a semiparametric analysis using the kernel regression estimator (Robinson 1988). The semiparametric specification of the model shown in Equation (10), relating literacy rates to market potential, is:

$$\ln\left(\frac{\textit{literacy rate}}{1-\textit{literacy rate}}\right) = \alpha + f(MP) + \Gamma' X + \Pi' T + \varphi\eta + \varepsilon, \tag{12}$$

in which, for the sake of clarity, we drop the subscripts it. The model contains a linear part including the vector X of (lagged) control variables (the province's total population, its population density, and a measure of the educational infrastructure), provincial fixed effects (η) and a set of province-specific time trends (T). The difference with the parametric model is $f(\bullet)$, a smooth and continuous, possibly nonlinear, unknown function of the market potential. ε is a random error term.

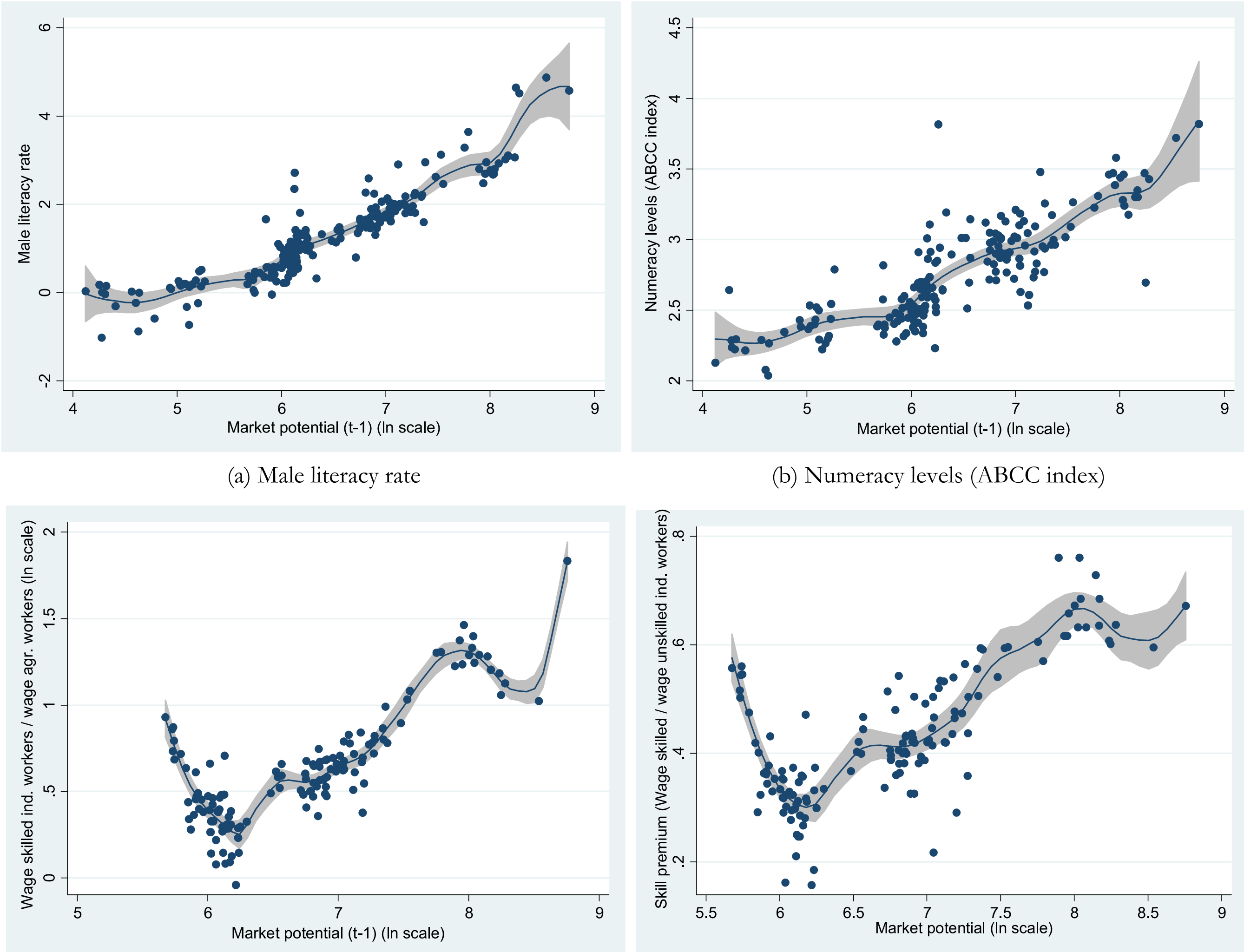
Robinson's approach is a two-step methodology. First, the linear coefficients ($\hat{\Gamma}'$, $\hat{\Pi}'$, and $\hat{\varphi}$) are estimated by applying a procedure similar to that whereby variables can be partialed out of an OLS regression (but using nonparametric regressions). Second, a kernel regression of $ln\left(\frac{\textit{literacy rate}}{1-\textit{literacy rate}}\right) - \hat{\Gamma}'X - \hat{\Pi}'T - \hat{\varphi}\,\eta$ on market potential is carried out. In both stages, a Gaussian kernel-weighted local polynomial fit is used for kernel regressions.

The analogous semiparametric model to assess the relationship between market potential and the skill premium is:

$$\ln(\textit{wage skilled}) - \ln(\textit{wage unskilled}) = \alpha + f(MP) + \Gamma' X + \Pi' T + \varphi\eta + \varepsilon. \tag{13}$$

Graphs (a) to (d) in Figure 9 show the nonparametric part of the estimation, including the 95% confidence bands[33]. These graphs show the nonlinear relationship between total market potential and different dependent variables considered in our previous analysis: male literacy rates, male ABCC index, and the two skill premia[34]. Graphs (a) and (b) confirm the positive relationship between market potential and the human capital measures, thus supporting the results obtained in the parametric analysis reported in Tables 3–5. The larger the market potential, the higher it is its positive impact on literacy rates and ABBC index. Although this relationship does not seem very different from a straight line, the Hardle and Mammen's (1993) specification test, whose null hypothesis is that parametric and nonparametric fits are not different from a statistical perspective, confirms the significant nonlinear relationship in all cases.

Regarding the skill premium, graphs (c) and (d) in Figure 9 show a U-shaped pattern. While the relationship is increasing for most of the market potential distribution, in the left-side distribution, for the smallest values of market potential, the relationship between market potential and the skill premium is negative. These observations correspond to landlocked interior provinces but which had a relatively high skill premium, although no clear pattern emerges as to whether this is due to high skilled wages or low unskilled wages. However, the figures confirm the positive relationship between market potential and the skill premium for most of the market potential distribution.



(a) Male literacy rate

(b) Numeracy levels (ABCC index)

(c) Wage skilled industrial workers v. wage agricultural workers

(d) Wage skilled industrial workers v. wage unskilled industrial workers

FIGURE 9 | Semiparametric relationship between market potential and the different dependent variables. (a) Male literacy rate. (b) Numeracy levels (ABCC index). (c) Wage skilled industrial workers v. wage agricultural workers. (d) Wage skilled industrial workers v. wage unskilled industrial workers. Notes: Semiparametric estimates and 95% confidence bands. Dependent variable: (a) logistic transformation of the male literacy rate, (b) logistic transformation of the male numeracy levels (ABCC index), (c) ln(*wage skilled industrial* workers)—ln(*wage agricultural* workers), and (d) ln(*wage skilled industrial* workers)—ln(*wage unskilled industrial* workers). The linear part of the model includes control variables (labor abundance, population density, and educational infrastructure, all in logarithmic scale), provincial fixed effects and province-specific time trends. Robust standard errors clustered by province. [Color figure can be viewed at wileyonlinelibrary.com]

That is, for medium and large market potentials, the larger the market potential, the higher it is its positive impact on the skill premium in either of our two definitions.

Overall, once these links are established, the exercises carried out show that the different levels of market access in Spanish provinces would explain the variation observed in the skill premium in the period before the Civil War, and on the accumulation of basic human capital in the form of higher literacy rates.

8 | Concluding Remarks

The existence of notable differences between regions is one of the most striking elements of the economic growth processes in developed and developing economies. The NEG is a conceptual framework that has made it possible to obtain a well-founded explanation for the generation of regional inequalities in parallel with the advance of technological change and market integration, which are the drivers of economic development in societies. However, making the connection between the generation of regional inequalities in the early stages of the economic development processes and their persistence over time requires us to establish links between NEG literature and economic growth theory.

While endogenous growth theory indicates that differences in the rates of human capital accumulation could be at the root of the different growth paths followed by countries or regions, the work of Redding and Schott (2003) bridges these two strands of the literature. These authors show that, because of the higher trade costs its companies have to bear, countries on the periphery of the world economy cannot adequately remunerate the domestic factors of production, in particular skilled workers. The theoretical model they present therefore suggests that remoteness has a negative effect on the skill premium (it does not substantially increase), which discourages the accumulation of human capital. The result is that countries with lower market access have lower levels of educational attainment, and this hinders the economic development of peripheral countries.

To advance in the study of this hypothesis, in this paper we use the case of Spain during the early stages of the economic development process to analyze whether or not a relationship exists between regional variability in market access and education levels. This, in turn, would act as an explanatory element for the consolidation of diverse growth paths between regions. The results obtained allow us to confirm not only the hypothesis put forward by Redding and Schott (2003) but also the presence of a mechanism that—along the lines proposed by Fallah et al. (2011)—relates the provinces' market size to their production specialization and human capital accumulation, this mechanism being regional differences in the returns on investment in education.

Thus our findings show that in the case of Spain between 1860 and 1930, those provinces with the greatest market potential would have recorded greater specialization in those production sectors that call for more intensive use of human capital. In a context of reduced worker mobility between regions, they would also have offered greater remuneration to qualified workers, thus enabling a higher skill premium. Given these conditions, the presence of increasing returns on economic activities requiring intensive use of human capital in their production processes would make it possible to connect market access and the unequal distribution of activity in territories (as suggested by the NEG) to the appearance of significant gaps in human capital accumulation rates and in the growth trajectories of the Spanish regions (as suggested by endogenous growth theory).

Acknowledgments

The authors would like to thank the Bank of Spain and the FBBVA for their financial support through the "Research Grant Programme in Macroeconomics, Monetary, Financial and Banking Economics, and Economic History" and the "Grant Programme for Scientific Research Teams in Digital Humanities," respectively. Financial support from the Spanish Ministry of Science and Innovation, projects PGC2018-095821-B-I00, PID2020-114354RA-I00, PID2021-127636NB-I00 and PID2023-147822NB-I00 (MCIU/AEI/FEDER, UE) and contract FPU18/01210; the Gobierno de Aragón (SR39_23R, ADETRE Research Group); and the Generalitat Valenciana (AICO/2020/2017, PROMETEU/2020/83 and CIPROM/2023/39) is also gratefully acknowledged. Earlier versions of this study were presented at the XIV Labour Economics Meeting (online), XLVII International Conference on Regional Science (Granada), 25th INFER Conference (València), Iecon Economic History Seminar (Montevideo), International Meeting on Education, Development, Human Capital between the 19th and 20th Century: a European Perspective (Sassari), 70th North American Meetings of the Regional Science Association International (San Diego), XXXII Meeting of the Economics of Education Association (València), 94th Annual Meeting of the Southern Economics Association (Washington DC) and the 32nd APDR Congress (Porto). The authors would like to thank the participants for their comments, as well as the anonymous referees and the editor for their suggestions.

Endnotes

[1] See Rosés et al. (2010, pp.245–246) for a detailed account of the integration of the Spanish market between 1860 and 1930.

[2] In particular, it fits the dynamics behind the model proposed by Fujita et al. (1999), on which the work by Redding and Schott (2003) is based.

[3] In the Spanish case, interprovincial migrations only reached noteworthy size in the 1920s and were mainly associated with movement towards Madrid and Barcelona (Pons et al. 2007).

[4] In the case of wages, the information is restricted to the years 1910, 1920, and 1930, as will be explained later.

[5] According to calculations made by Herranz (2005a), haulage costs in 1878 fell by up to 86% thanks to the introduction of the railway.

[6] This methodology has repeatedly been used in economic history: Wolf (2007) analyzed Poland after WWI, Crafts and Mulatu (2005, 2006) studied Britain, and Klein and Crafts (2012) applied the approach to examine the manufacturing belt in the US between 1880 and 1920. Betrán (2011) also analyzed the Spanish case in the long term at regional NUTS2 level.

[7] Spain is divided into 50 provinces. However, in constructing the market potential variable we excluded the three non-mainland provinces from the analysis, and thus we work with a total of only 47 provinces.

[8] On the issue on female age heaping, see also Földvári et al. (2012) and Blum et al. (2017). A critique of the use of age heaping as an indicator of numeracy in A'Hearn et al. (2022).

[9] "Despite the fact that we bring theory to empirical implementation in a structural way, the 'correct' measure of market potential actually underperforms the atheoretical Harris (1954) measure" (Head and Mayer 2004).

[10] See Martinez-Galarraga (2014) for a detailed description.

[11] This is why the first year for which we have information for this variable is 1867. This has a historical logic due to the fact that the first basic stage in the construction of the country's rail network had been completed by 1866, joining up the major cities and centers of economic activity. We take the 1867 market potential calculations for the 1860 benchmark.

[12] In this case, due to data availability, the data set for wages begins in 1910.

[13] For 1910 we take the data corresponding to 1914.

[14] We have decided not to use the category of apprentices because, due to their nature, it is likely that in many cases these were contractual relationships with a strong training-oriented component. Thus, it is likely that reported wages did not really reflect the real remuneration of work. In addition, it is important to bear in mind that women's participation in the labor market at the time was very limited and was circumscribed to very specific sectors and under a very marked tutelage of the male gender. These occupations generally required little training, which is why their salaries are systematically lower than those of men (Sarasúa 1997; Vilar 2004). For this reason, our analysis considers exclusively the data relating to male categories.

[15] See Ministerio de Trabajo (1927: xviii, 1931: xxviii) for the procedure used to obtain average wages by industry and province.

[16] These daily wages are then converted to hourly wages assuming 10 h of work, as is usual for this period in the literature (Prados de la Escosura 2017). The absence of data on agricultural wages in 1920

for the provinces of Pontevedra and León has been offset by interpolating the values provided in Bringas (2000) for 1919 and 1921 in these provinces.

[17]*Anuario Estadístico de España* of 1860, the *Anuario Estadístico de Instrucción Primaria* 1900-1901, the *Estadística Escolar* of 1908, the *Estadística de Escuelas Nacionales* of 1923 and the *Anuario Estadístico de España* of 1930 (Díez-Minguela et al. 2023).

[18]Successive changes were nevertheless introduced during the 113 years that the PIA remained in force. Compulsory education, for example, initially from ages 6 to 9, was extended to age 12 in 1909 and age 14 in 1964.

[19]The compulsory nature of education was not absolute, since pupils could ask to be excused when they were "*sufficiently provided with this type of education in their homes or in a private establishment*" (Art. 7). And to obtain free elementary education, it was necessary to provide a "*certificate issued by the relevant parish priest and endorsed by the town mayor*" (Art. 9).

[20]Female literacy, on the other hand, was only 11.9%. Taken as a whole, of a total population of almost 16 million in Spain it is estimated that only 26.5% could read and write in 1860.

[21]The provincial distribution of female and total literacy rates between 1860 and 1930 can be seen in Figures A.1 and A.2 respectively, in the Appendix.

[22]In 1930 the female literacy rate was 65.2% and total literacy was 73.3%.

[23]In Figures A.3 and A.4 in the Appendix we include information on the geographical distribution of domestic and foreign market potential respectively.

[24]If we consider the ABCC index as the human capital indicator, the results are analogous to those obtained with the male literacy rates (see Figure A.5 in the Appendix). Graphs using total and female literacy rates are also similar. These are available from the authors on request.

[25]If instead of using lagged values we take contemporaneous values of market potential—as Redding and Schott (2003) did—we obtain similar results. These are available from the authors on request.

[26]In this respect, we have carried out the exercises using different measures for educational infrastructure. Thus, we have considered only the number of public schools, the number of schools per capita and, finally, the number of pupils per school, an indicator that might be closer to approximating the quality of education. In all cases, the results obtained are analogous.

[27]To adjust the time trends to the structure of our panel, we have defined them as 1860:1, 1900:5, 1910:6, 1920:7, 1930:8.

[28]In an alternative specification we also considered migrations. We therefore re-estimated all the models, including the migration rate by province (Silvestre 2005) as a control variable. However, this variable was not significant in any case and for this reason we decided not to include it in our main specification. Given that the coefficients for the other variables barely change when the migration rate is included and we pass all the tests for the IV regressions, we do not think that there is a selection bias problem. We are grateful to Javier Silvestre for kindly sharing migration data with us.

[29]Scatter plots of the residuals from the non-instrumented regression versus each instrument may complement formal overidentification tests, as emphasized in Angrist and Pischke (2009). In our case, they show no systematic relationship between the residuals and the instruments, as evidenced by approximately horizontal trends in all cases, supporting the fulfillment of the exclusion restriction. These residual plots are available from the authors on request.

[30]We therefore focus on the years 1910, 1920, and 1930.

[31]Again, non-reported plots of the residuals from the non-instrumented regression versus each instrument show no systematic relationship between the residuals and the instruments. These plots are available from the authors on request.

[32]This problem could be solved by extending the study period beyond 1930 so as to try to capture the positive effect of the skill premium in 1910 on the literacy rate several decades later. However, taking literacy beyond 1930 into account is not without problems in the Spanish case. First, because 1940 is an exceptional year given the impact of the Civil War (1936–39), and second, because from 1950 onwards the variation in literacy rates is considerably smaller, especially in the northern part of the peninsula, an area that was very close to universal literacy at that time.

[33]Estimations of the linear part of the model are not shown. These linear coefficients are consistent with those reported in Tables 3–6, although there are slight differences. These results are available from the authors on request.

[34]The results obtained considering the alternative definitions of market potential are similar. These results are available from the authors on request.

References

A'Hearn, B., J. Baten, and D. Crayen. 2009. "Quantifying Quantitative Literacy: Age Heaping and the History of Human Capital." *Journal of Economic History* 69, no. 3: 783–808.

A'Hearn, B., A. Delfino, and A. Nuvolari. 2022. "Rethinking Age Heaping: A Cautionary Tale From Nineteenth-Century Italy." *Economic History Review* 75, no. 1: 111–137.

Abraham, K. G., and J. Mallatt. 2022. "Measuring Human Capital." *Journal of Economic Perspectives* 36, no. 3: 103–130.

Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Barquín, R. 1999. "El Ahorro Social: Una Perspectiva Dinámica." In *Siglo y Medio del Ferrocarril en España, 1848-1998*, edited by M. Muñoz, J. Sanz, and J. Vidal, 337–354. Fundación de los Ferrocarriles Españoles.

Barro, R. J. 2013. "Education and Economic Growth." *Annals of Economics and Finance* 14, no. 2: 301–328.

Barro, R. J., and J. W. Lee. 2013. "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics* 104: 184–198.

Beltrán Tapia, F. J., A. Díez-Minguela, J. Martinez-Galarraga, and D. A. Tirado-Fabregat. 2021. "The Uneven Transition Towards Universal Literacy in Spain, 1860-1930." *History of Education* 50, no. 5: 605–627.

Beltrán Tapia, F. J., A. Díez-Minguela, J. Martinez-Galarraga, and D. A. Tirado-Fabregat. 2022. "Two Stories, One Fate: Age-Heaping and Literacy in Spain, 1877-1930." *Revista de Historia Económica/Journal of Iberian and Latin American Economic History* 40, no. 3: 405–438.

Beltrán Tapia, F. J., and J. Martinez-Galarraga. 2018. "Inequality and Education in Pre-Industrial Economies: Evidence From Spain." *Explorations in Economic History* 69, no. 3: 81–101.

Betrán, C. 2011. "Regional Specialisation and Industry Location in the Long Run: Spain in the US Mirror (1856–2002)." *Cliometrica* 5, no. 3: 259–290.

Betrán, C., and M. A. Pons. 2022. "Capital Flow Bonanzas as a Fundamental Ingredient in Spain's Financial Crises, 1850-2015." *Revista de Historia Económica/Journal of Iberian and Latin American Economic History* 40, no. 1: 135–169.

Black, D., and V. Henderson. 2003. "Urban Evolution in the USA." *Journal of Economic Geography* 3, no. 4: 343–372.

Blum, M., C. L. Colvin, L. McAtackney, and E. McLaughlin. 2017. "Women of An Uncertain Age: Quantifying Human Capital Accumulation in Rural Ireland in the Nineteenth Century." *Economic History Review* 70, no. 1: 187–223.

Brakman, S., H. Garretsen, and M. Schramm. 2004. "The Spatial Distribution of Wages: Estimating the Helpman-Hanson Model for Germany." *Journal of Regional Science* 44, no. 3: 437–466.

Breinlich, H. 2006. "The Spatial Income Structure in the European Union: What Role for Economic Geography?" *Journal of Economic Geography* 6: 593–617.

Bringas, M. A. 2000. *La Productividad de los Factores en la Agricultura Española (1752-1935)*. Banco de España.

Brülhart, M., and F. Sbergami. 2009. "Agglomeration and Growth: Cross-Country Evidence." *Journal of Urban Economics* 65, no. 1: 48–63.

Castañeda, L., and X. Tafunell. 1993. "Un Nuevo Indicador Para la Historia Financiera Española: La Cotizacion de las Letras de Cambio a Corto Plazo." *Revista de Historia Económica/Journal of Iberian and Latin American Economic History* 11, no. 2: 367–383.

Ciccone, A. 2002. "Agglomeration Effects in Europe." *European Economic Review* 46, no. 2: 213–227.

Ciccone, A., and R. E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review* 86, no. 1: 54–70.

Clark, C., F. Wilson, and J. Bradley. 1969. "Industrial Location and Economic Potential in Western Europe." *Regional Studies* 3, no. 2: 197–212.

Combes, P. P., and L. Gobillon. 2015. "The Empirics of Agglomeration Economies." In *Handbook of Urban and Regional Economics*, edited by G. Duranton, J. V. Henderson, and W. Strange, 5, 247–348. Elsevier.

Combes, P. P., T. Mayer, and J. F. Thisse. 2008. *Economic Geography: The Integration of Regions and Nations*. Princeton University Press.

Crafts, N. 2005. "Market Potential in British Regions, 1871-1931." *Regional Studies* 39, no. 9: 1159–1166.

Crafts, N., and A. Mulatu. 2005. "What Explains the Location of Industry in Britain, 1871-1931?" *Journal of Economic Geography* 5, no. 4: 499–518.

Crafts, N., and A. Mulatu. 2006. "How Did the Location of Industry Respond to Falling Transport Costs in Britain Before World War I?" *Journal of Economic History* 66, no. 3: 575–607.

Crafts, N., and A. J. Venables. 2003. "Globalization in History: A Geographical Perspective." In Globalization in Historical Perspective*, Chapter 7*, edited by M. D. Bordo, A. M. Taylor, and J. G. Williamson. University of Chicago Press.

Crozet, M. 2004. "Do Migrants Follow Market Potentials? An Estimation of a New Economic Geography Model." *Journal of Economic Geography* 4, no. 4: 439–458.

Crozet, M., T. Mayer, and J. L. Mucchielli. 2004. "How Do Firms Agglomerate? A Study of FDI in France." *Regional Science and Urban Economics* 34, no. 1: 27–54.

Davis, D. R., and D. E. Weinstein. 1999. "Economic Geography and Regional Production Structure: An Empirical Investigation." *European Economic Review* 43, no. 2: 379–407.

Davis, D. R., and D. E. Weinstein. 2003. "Market Access, Economic Geography and Comparative Advantage: An Empirical Test." *Journal of International Economics* 59, no. 1: 1–23.

Diebolt, C., and R. Hippe. 2018. "Remoteness Equals Backwardness? Human Capital and Market Access in the European Regions: Insights From the Long Run." *Education Economics* 26, no. 3: 285–304.

Díez-Minguela, A., P. Insa-Sánchez, and J. Martinez-Galarraga 2023. "Only a Public School? Education and Growth in a Developing Economy: Spain, 1860-1930," Paper Presented at Iberometrics 2023, 25-26 May, Lisbon.

Díez-Minguela, A., J. Martinez-Galarraga, and D. A. Tirado. 2018. *Regional Inequality in Spain, 1860-2015*. Palgrave-Macmillan.

Díez-Minguela, A., J. Martinez-Galarraga, and D. A. Tirado-Fabregat. 2016. "Why Did Spanish Regions not Converge Before the Civil War? Agglomeration Economies and (Regional) Growth Revisited." *Revista de Historia Económica/Journal of Iberian and Latin American Economic History* 34, no. 3: 417–448.

Dirección General de Obras Públicas. 1861. *Memoria Sobre el Progreso de las Obras Públicas en España, Años 1859 y 1860*. Edited by Ministerio de Fomento.

Durlauf, S. N. 2001. "Manifesto for a Growth Econometrics." *Journal of Econometrics* 100, no. 1: 65–69.

Estevadeordal, A., B. Frantz, and A. M. Taylor. 2003. "The Rise and Fall of World Trade, 1870-1939." *Quarterly Journal of Economics* 118: 359–407.

Faíña, J. A., and J. López-Rodríguez. 2006. "Market Access and Human Capital Accumulation: The European Union case." *Applied Economics Letters* 13, no. 9: 563–567.

Fallah, B. N., M. D. Partridge, and M. R. Olfert. 2011. "New Economic Geography and US Metropolitan Wage Inequality." *Journal of Economic Geography* 11, no. 5: 865–895.

Földvári, P., B. Van Leeuwen, and J. Van Leeuwen-li. 2012. "How Did Women Count? A Note on Gender-Specific Age Heaping Differences in the Sixteenth to Nineteenth Centuries." *Economic History Review* 65, no. 1: 304–313.

Frago, A. V. 1990. "The History of Literacy in Spain: Evolution, Traits, and Questions." *History of Education Quarterly* 30, no. 4: 573–599.

Frax, E. 1981. *Puertos y Comercio de Cabotaje en España, 1857-1934*. Banco de España.

Fujita, M., P. Krugman, and A. J. Venables. 1999. *The Spatial Economy: Cities, Regions, and International Trade*. MIT press.

Galor, O. 2011. "Inequality, Human Capital Formation and the Process of Development." In *Handbook of the Economics of Education*, edited by E. A. Hanushek, S. Machin, and L. Woessmann, 4, 441–493. Elsevier.

Gennaioli, N., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer. 2013. "Human Capital and Regional Development." *The Quarterly Journal of Economics* 128, no. 1: 105–164.

Goldin, C. 2016. "Human Capital." In Handbook of Cliometrics*, Chapter 3*, edited by C. Diebolt and M. Haupert, 55–86. Springer.

González-Val, R., D. A. Tirado-Fabregat, and E. Viladecans-Marsal. 2017. "Market Potential and City Growth: Spain 1860-1960." *Cliometrica* 11, no. 1: 31–61.

Hanson, G. H. 1998. "Market Potential, Increasing Returns, and Geographic Concentration." NBER Working Paper 6429.

Hanson, G. H. 2005. "Market Potential, Increasing Returns, and Geographic Concentration." *Journal of International Economics* 67, no. 1: 1–24.

Hanushek, E. A., and L. Woessmann. 2012. "Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation." *Journal of Economic Growth* 17: 267–321.

Hardle, W., and E. Mammen. 1993. "Comparing Nonparametric Versus Parametric Regression Fits." *Annals of Statistics* 21: 1926–1947.

Harris, C. D. 1954. "The Market as a Factor in the Localization of Industry in Theunited States." *Annals of the Association of American Geographers* 64: 315–348.

Harrison, J. 1990. "The Economic History of Spain Since 1800." *Economic History Review* 43, no. 1: 79–89.

Head, K., and T. Mayer. 2011. "Gravity, Market Potential and Economic Development." *Journal of Economic Geography* 11: 281–294.

Head, K., and T. Mayer. 2004. "The Empirics of Agglomeration and Trade." In *Handbook of Regional and Urban Economics 4: Cities and Geography*, edited by J. V. Henderson and J. F. Thisse, 2609–2669. Elsevier.

Herranz, A. 2005a. “La Reducción de los Costes de Transporte en España (1800-1936).” *Cuadernos Económicos del ICE* 70: 183–203.

Herranz, A. 2005b. “The Spanish Infrastructure Stock, 1845-1935.” *Research in Economic History* 23: 85–129.

Insa-Sánchez, P. 2024. “Spatial Inequality of Opportunity in Access to Secondary Education in Nineteenth-Century Spain.” *European Review of Economic History* 28, no. 1: 67–90.

Karahasan, B. C., and E. López-Bazo. 2013. “The Spatial Distribution of Human Capital: Can it Really be Explained by Regional Differences in Market Access?” *International Regional Science Review* 36, no. 4: 451–480.

Keeble, D., P. L. Owens, and C. Thompson. 1982. “Regional Accessibility and Economic Potential in the European Community.” *Regional Studies* 16, no. 6: 419–432.

Kim, S. 1995. “Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987.” *Quarterly Journal of Economics* 110, no. 4: 881–908.

Klein, A., and N. Crafts. 2012. “Making Sense of the Manufacturing Belt: Determinants of U.S. Industrial Location, 1880-1920.” *Journal of Economic Geography* 12, no. 4: 775–807.

Krugman, P. 1991. “Increasing Returns and Economic Geography.” *Journal of Political Economy* 99, no. 3: 483–499.

Krugman, P., and A. J. Venables. 1995. “Globalization and the Inequality of Nations.” *Quarterly Journal of Economics* 110, no. 4: 857–880.

Kunst, D., R. B. Freeman, and R. Oostendorp. 2022. “Occupational Skill Premia Around the World: New Data, Patterns and Drivers.” *Labour Economics* 79: 102255.

Liu, D., and C. M. Meissner. 2015. “Market Potential and the Rise of US Productivity Leadership.” *Journal of International Economics* 96, no. 1: 72–87.

López-Rodríguez, J., J. A. Faíña, and J. López-Rodríguez. 2007. “Human Capital Accumulation and Geography: Empirical Evidence From the European Union.” *Regional Studies* 41, no. 2: 217–234.

Lucas, R. E. 1988. “On the Mechanics of Economic Development.” *Journal of Monetary Economics* 22, no. 1: 3–42.

Martinez-Galarraga, J. 2012. “The Determinants of Industrial Location in Spain, 1856-1929.” *Explorations in Economic History* 49, no. 2: 255–275.

Martinez-Galarraga, J. 2014. “Market Potential Estimates in History: A Survey of Methods and an Application to Spain, 1867–1930.″ European Historical Economics Society (EHES) Working Papers in Economic History 51.

Martinez-Galarraga, J., E. Paluzie, J. Pons, J. Silvestre, and D. A. Tirado. 2021. “New Economic Geography and Economic History: A Survey of Recent Contributions Through the Lens of the Spanish Industrialization Process.” *Cliometrica* 15, no. 3: 719–751.

Martinez-Galarraga, J., E. Paluzie, J. Pons, and D. A. Tirado-Fabregat. 2008. “Agglomeration and Labour Productivity in Spain Over the Long Term.” *Cliometrica* 2, no. 3: 195–212.

Martinez-Galarraga, J., D. A. Tirado, and R. González-Val. 2015. “Market Potential and Regional Economic Growth in Spain (1860-1930).” *European Review of Economic History* 19, no. 4: 335–358.

Matas, A., J. L. Raymond, and J. L. Roig. 2015. “How Market Access Shapes Human Capital Accumulation in a Peripheral Country: The Case of Spain.” *Applied Economics* 47, no. 11: 1118–1132.

Midelfart-Knarvik, K. H., H. Overman, S. Redding, and A. J. Venables. 2002. “The Location of European Industry.” In European Economy. European Integration and the Functioning of Product Markets*, 2/2002*, edited by European Commission, 213–269. Directorate-General for Economic and Financial Affairs.

Ministerio de Obras Públicas. 1902. *Estadísticas de Obras Públicas, años 1899 y 1900*. Dirección General de Obras Públicas.

Ministerio de Trabajo. 1927. *Estadística de Salarios y Jornadas de Trabajo Referida al Período 1914–1925*. Dirección General de Trabajo y Acción Social.

Ministerio de Trabajo y Previsión. 1931. *Estadística de Salarios y Jornadas de Trabajo Referida al Período 1914–1930*. Dirección General de Trabajo.

Mitchell, B. R. 1998a. *International Historical Statistics. Europe 1750–1993*. Macmillan.

Mitchell, B. R. 1998b. *International Historical Statistics. The Americas 1750–1993*. Macmillan.

Mohammed, S. I. S., and J. G. Williamson. 2004. “Freight Rates and Productivity Gains in British Tramp Shipping 1869–1950.” *Explorations in Economic History* 41, no. 2: 172–203.

Nadal, J. 1975. *El Fracaso de la Revolución Industrial en España, 1814-1913*. Ariel.

Nadal, J. 1987. “La Industria Fabril Española en 1900: Una Aproximación.” In *La Economía Española en el Siglo XX: Una Perspectiva Histórica*, edited by C. Sudrià, J. Nadal, and A. Carreras, 23–61. Ariel.

Niebuhr, A. 2006. “Market Access and Regional Disparities: New Economic Geography in Europe.” *The Annals of Regional Science* 40, no. 2: 313–334.

Nogués-Marco, P., A. Herranz-Loncán, and N. Aslanidis. 2019. “The Making of a National Currency: Spatial Transaction Costs and Money Market Integration in Spain (1825-1874).” *The Journal of Economic History* 79, no. 4: 1094–1128.

Núñez, C. E. 1991. “El Gasto Público en Educación Entre 1860 y 1935.” *Hacienda Pública Española* 1: 121–146.

Núñez, C. E. 1992. *La Fuente de la Riqueza. Educación y Desarrollo Económico en la España Contemporánea*. Alianza.

Núñez, C. E. 2003. “Literacy, Schooling and Economic Modernization: A Historian's Approach.” *Paedagogica Historica* 39, no. 5: 535–558.

O'Rourke, K. H. 2000. “Tariffs and Growth in the Late 19th Century.” *The Economic Journal* 110: 456–483.

Ortiz-Villajos, J. M. 2005. “Spain's Low Technological Level: An Explanation.” In *Technology and Human Capital in Historical Perspective*, edited by J. Ljungberg and J. P. Smits, 182–204. Palgrave Macmillan.

Ottaviano, G. I. P., and D. Pinelli. 2006. “Market Potential and Productivity: Evidence From Finnish Regions.” *Regional Science and Urban Economics* 36, no. 5: 636–657.

Paluzie, E., J. Pons, and D. A. Tirado. 2004. “The Geographical Concentration of Industry Across Spanish Regions, 1856–1995.” *Jahrbuch Für Regionalwissenschaft* 24, no. 2: 143–160.

Peña, D., and N. Sánchez-Albornoz. 1983. *Dependencia Dinámica Entre Precios Agrícolas. El Trigo en España, 1857–1890. Un Estudio Empírico*. Servicio de Estudios de Historia Económica, Banco de España.

Pérez Moreda, V. 1997. “El Proceso de Alfabetización y la Formación de Capital Humano en España.” *Papeles de Economía Española* 73: 243–253.

Pérez Moreda, V., D. S. Reher, and A. Sanz Gimeno. 2015. *La Conquista de la Salud. Mortalidad y Modernización en la España Contemporánea*. Marcial Pons.

Pons, J., E. Paluzie, J. Silvestre, and D. A. Tirado. 2007. “Testing the New Economic Geography: Migrations and Industrial Agglomerations in Spain.” *Journal of Regional Science* 47, no. 2: 289–313.

Prados de la Escosura, L. 1982. *Comercio Exterior y Crecimiento Económico en España, 1826-1913: Tendencias a Largo Plazo*. Servicio de Estudios de Historia Económica, Banco de España.

Prados de la Escosura, L. 2017. *Spanish Economic Growth, 1850-2015*. Palgrave-Macmillan.

Prados de la Escosura, L. 2000. "International Comparisons of Real Product, 1820–1990: An Alternative Data Set." *Explorations in Economic History* 37, no. 1: 1–41.

Prados de la Escosura, L., and J. R. Rosés. 2010. "Capital Accumulation in the Long Run: The Case of Spain, 1850–2000." In *Research in Economic History*, edited by A. J. Field, 27, 141–200. Emerald Group Publishing Limited.

Puga, D. 1999. "The Rise and Fall of Regional Inequalities." *European Economic Review* 43, no. 2: 303–334.

Redding, S., and P. K. Schott. 2003. "Distance, Skill Deepening and Development: Will Peripheral Countries Ever Get Rich?" *Journal of Development Economics* 72, no. 2: 515–541.

Redding, S., and A. J. Venables. 2004. "Economic Geography and International Inequality." *Journal of International Economics* 62, no. 1: 53–82.

Reher, D. S. 1997. "La Teoría del Capital Humano y las Realidades de la Historia." *Papeles de Economía Española* 73: 254–261.

Reher, D. S. 2023. "Patterns of Literacy in Historic Spain: An Interpretation." *Revista de Historia Económica/ Journal of Iberian and Latin American Economic History* 41, no. 1: 83–117.

Robinson, P. M. 1988. "Root-N Consistent Semiparametric Regression." *Econometrica* 56, no. 4: 931–954.

Romer, P. M. 1986. "Increasing Returns and Long-Run Growth." *Journal of Political Economy* 94, no. 5: 1002–1037.

Roos, M. 2001. "Wages and Market Potential in Germany." *Jahrbuch Für Regionalwissenschaft* 21, no. 2: 171–195.

Rosés, J. R. 2003. "Why Isn't the Whole of Spain Industrialized? New Economic Geography and Early Industrialization, 1797-1910." *Journal of Economic History* 63, no. 4: 995–1022.

Rosés, J. R., J. Martinez-Galarraga, and D. A. Tirado. 2010. "The Upswing of Regional Income Inequality in Spain (1860-1930)." *Explorations in Economic History* 47, no. 2: 244–257.

Rosés, J. R., and B. Sánchez-Alonso. 2004. "Regional Wage Convergence in Spain 1850–1930." *Explorations in Economic History* 41, no. 4: 404–425.

Sáiz, P. 2005. "Investigación y Desarrollo: Patentes." In *Estadísticas Históricas de España, Siglos XIX-XX*, edited by A. Carreras and X. Tafunell, 835–872. FBBVA.

Sarasúa, C. 1997. "The Role of the State in Shaping Women's and Men's Entrance Into the Labour Market: Spain in the Eighteenth and Nineteenth Centuries." *Continuity and Change* 12, no. 3: 347–371.

Silvestre, J. 2005. "Internal Migrations in Spain, 1877-1930." *European Review of Economic History* 9, no. 2: 233–265.

Tirado, D. A., and J. Martinez-Galarraga. 2008. "Una Nova Estimació Retrospectiva del VAB Regional Industrial. Espanya (1860-1930)." *UB Economics Working Papers E08/192*.

Tirado, D. A., E. Paluzie, and J. Pons. 2002. "Economic Integration and Industrial Location: The Case of Spain Before World War I." *Journal of Economic Geography* 2, no. 3: 343–363.

Tirado, D. A., J. Pons, E. Paluzie, and J. Martinez-Galarraga. 2013. "Trade Policy and Wage Gradients: Evidence From a Protectionist Turn." *Cliometrica* 7, no. 3: 295–318.

Tollnek, F., and J. Baten. 2016. "Age-Heaping-Based Human Capital Estimates." In *Handbook of Cliometrics*, edited by C. Diebolt and M. Haupert, 131–154. Springer.

Vilar, M. 2004. "La Ruptura Posbélica a Través del Comportamiento de los Salarios Industriales: Nueva Evidencia Cuantitativa (1908-1963)." *Revista de Historia Industrial* 25: 81–121.

Wais, F. 1987. *Historia de los Ferrocarriles Españoles*. Fundación de los Ferrocarriles Españoles.

Wolf, N. 2007. "Endowments vs. Market Potential: What Explains the Relocation of Industry After the Polish Reunification in 1918?" *Explorations in Economic History* 44, no. 1: 22–42.

Appendix A

Table A1, Figure A1, Figure A2, Figure A3, Figure A4, Figure A5.

TABLE A1 | Variables and sources used.

Variable	Secondary source	Primary source
Male literacy rate	Beltrán Tapia et al. (2022)	Population Censuses (1860, 1900, 1910, 1920, 1930)
Total literacy rate	Beltrán Tapia et al. (2022)	Population Censuses (1860, 1900, 1910, 1920, 1930)
ABCC index	Beltrán Tapia et al. (2022)	Population Censuses (1860, 1900, 1910, 1920, 1930)
Market potential	Martinez-Galarraga (2014)	See text
Labor abundance	—	Population Censuses (1860, 1900, 1910, 1920, 1930)
Population density	—	INE and Population Censuses (1860, 1900, 1910, 1920, 1930)
Educational infrastructure	Díez-Minguela et al. (2023)	*Anuario Estadístico de España* (1860), *Anuario Estadístico de Instrucción Primaria* (1900-1901), *Estadística Escolar* (1908), *Estadística de Escuelas Nacionales* (1923), *Anuario Estadístico de España* (1930)
Infrastructure stock	Herranz (2005b)	—
Patents per capita	Sáiz (2005)	—
Share of manufacturing in total employment	Tirado and Martinez-Galarraga (2008)	Population Censuses (1860, 1900, 1910, 1920, 1930)
Share of population over regional population	—	Population Censuses (1860, 1900, 1910, 1920, 1930)
Industrial skilled wage	—	Ministerio de Trabajo (1927, 1931)
Industrial unskilled wage	—	Ministerio de Trabajo (1927, 1931)
Agrarian wage	Bringas (2000)	—

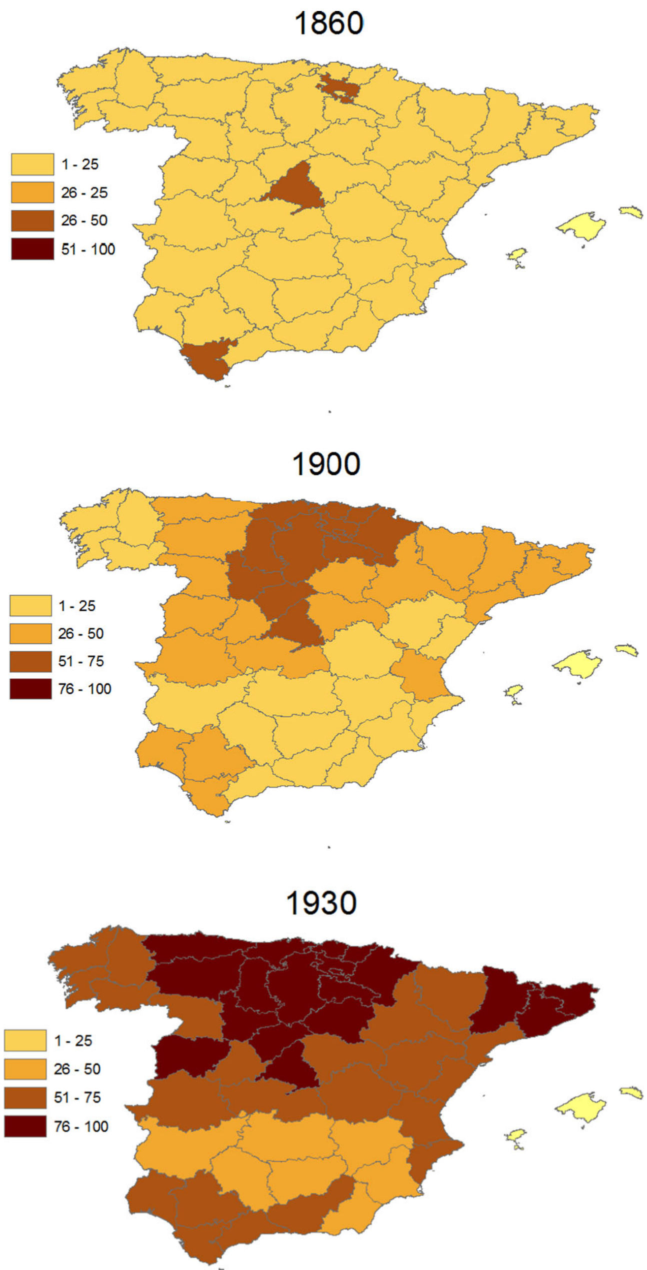


FIGURE A1 | Female literacy rates, 1860–1930. *Source:* Population censuses. [Color figure can be viewed at wileyonlinelibrary.com]

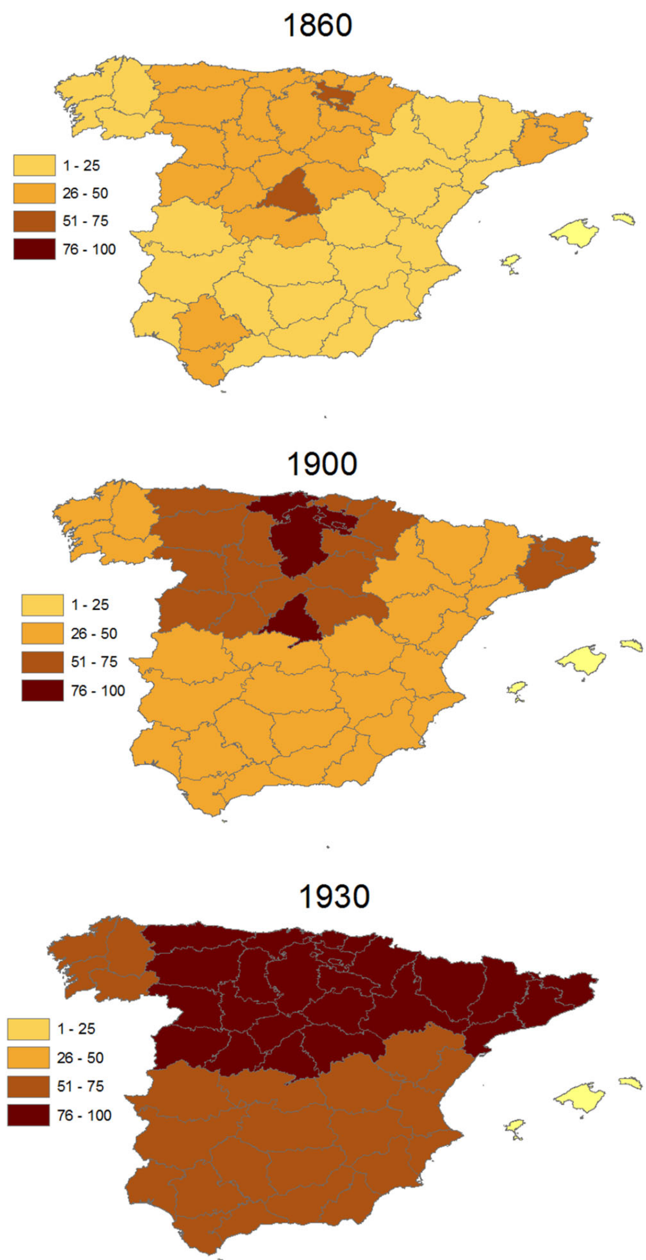


FIGURE A2 | Total literacy rates, 1860–1930. *Source:* Population censuses. [Color figure can be viewed at wileyonlinelibrary.com]

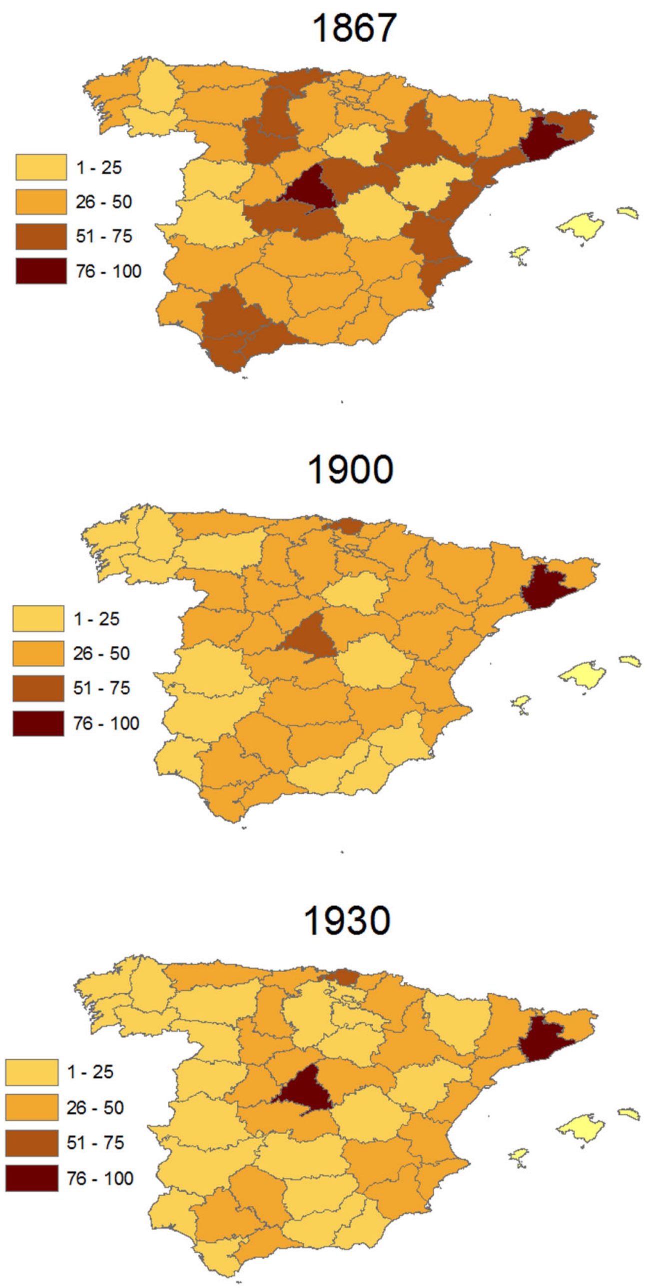


FIGURE A3 | Domestic market potential, 1867–1930 (Barcelona = 100). *Source:* Martinez-Galarraga (2014). [Color figure can be viewed at wileyonlinelibrary.com]

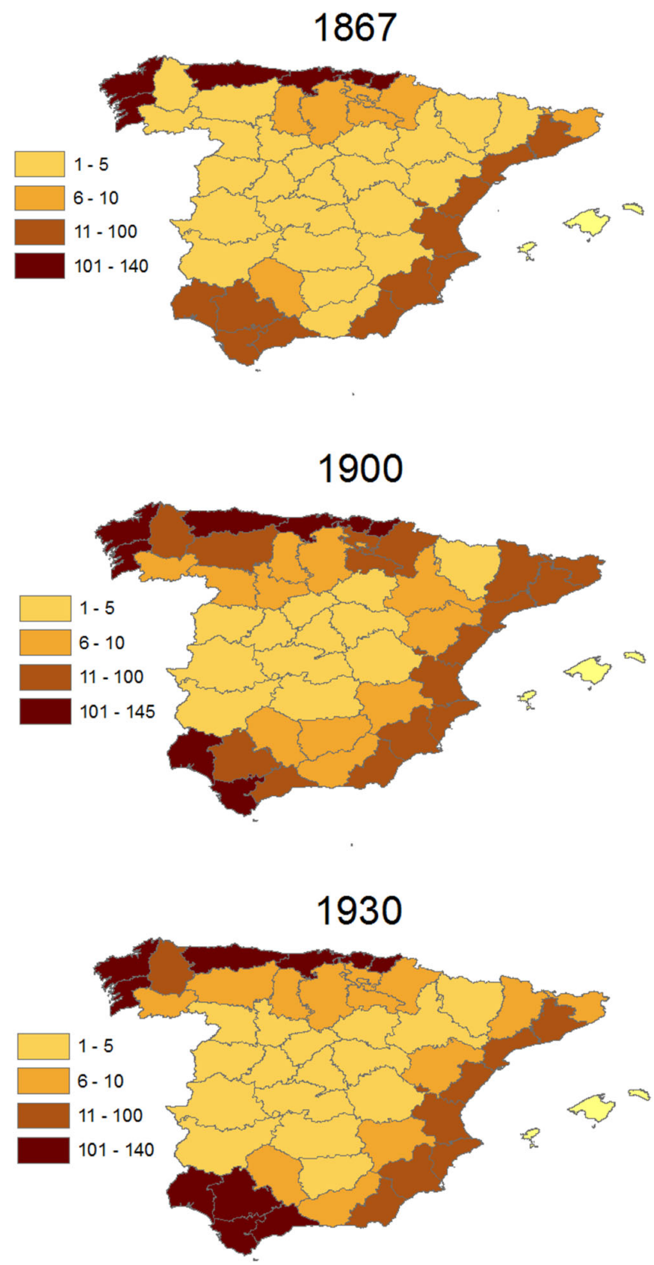


FIGURE A4 | Foreign market potential, 1867–1930 (Barcelona = 100). *Source:* Martinez-Galarraga (2014). [Color figure can be viewed at wileyonlinelibrary.com]

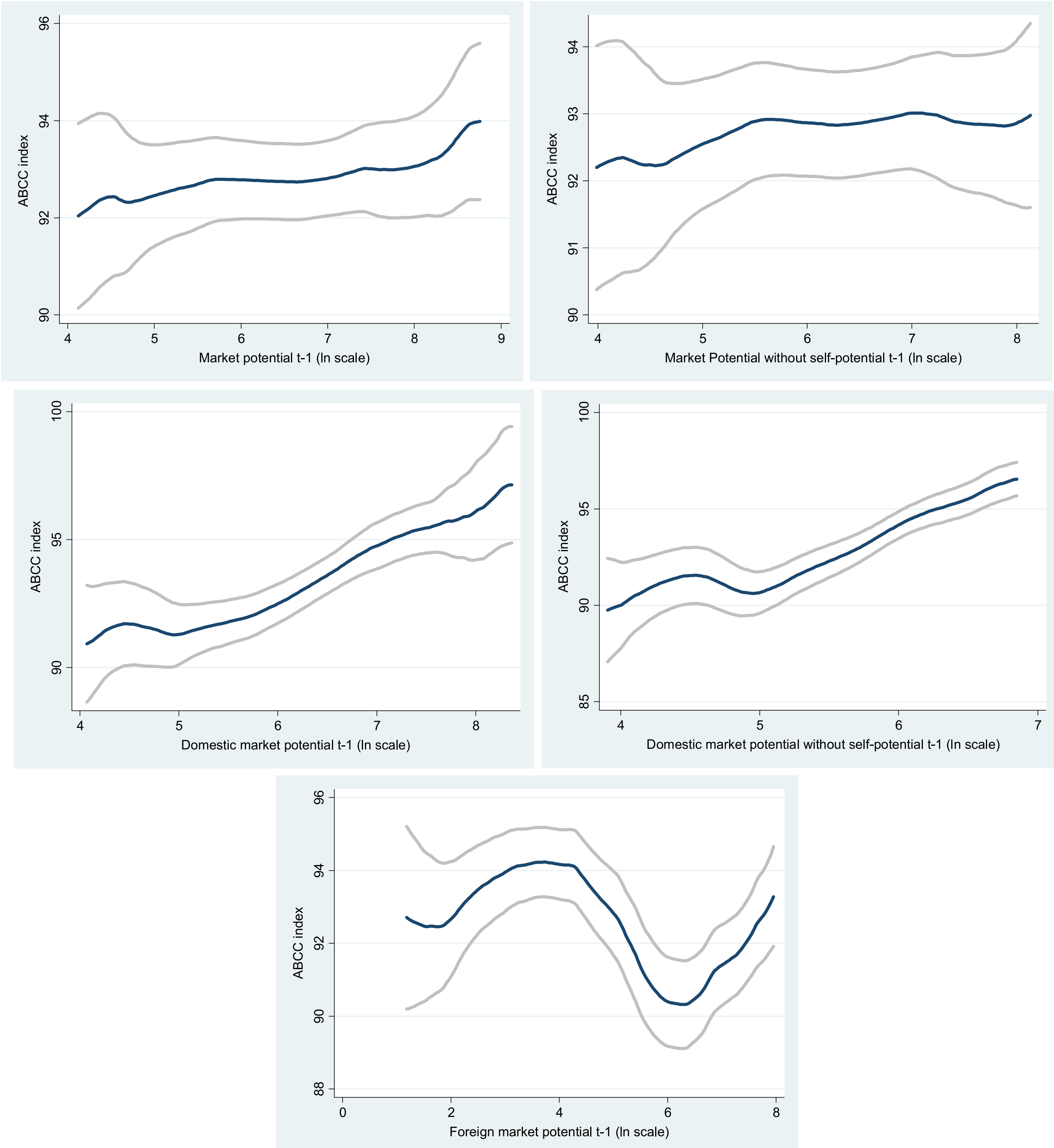


FIGURE A5 | Non-parametric relationship between the male ABCC index and different measures of lagged market potential (1860–1930). *Source:* see text. [Color figure can be viewed at wileyonlinelibrary.com]