**DOCUMENTS DE TREBALL
DE LA FACULTAT D'ECONOMIA I EMPRESA**

Col·lecció d'Economia

E10/244

The determinants of industrial location in Spain, 1856-1929

Julio Martínez-Galarraga

Adreça correspondència:

Departament d'Història i Institucions Econòmiques
Facultat d'Economia i Empresa
Universitat de Barcelona
Av. Diagonal, 690
08034 Barcelona (Spain)
Tfn. +34 934037231
Email.- julio.martinez@ub.edu

Acknowledgements

I am very grateful to Daniel A. Tirado for his comments and advice. I would also like to thank Fabian Gouret and the members of the Network in Economics and Public Policies (XREPP) launched by the Generalitat de Catalunya, especially Sergio Espuelas, Jordi Guilera and Alfonso Herranz (who also shared unpublished data with me), for their helpful comments. And finally, to the participants at the 4th Globaleuronet Summer School "Growth and Globalization: hardships, barriers and economic policies" held in Lisbon for their insightful suggestions, especially to Nick Crafts. Remaining errors are my own responsibility. Financial support from ESF (GlobalEuronet), XREPP and the Centre d'Estudis Antoni de Capmany is gratefully acknowledged.

Tots els treballs publicats en els “Documents de Treball de la Facultat d’Economia i Empresa” han passat i superat un procés d’avaluació anònima i externa.

Abstract:

During the 19th century, the Spanish economy went through the early stages of the industrialisation process. This process developed in parallel to the growing market integration of goods and factors as a result of the liberal reforms and the construction of the railway network, with the subsequent fall in transport costs. In that period, there were major changes in the pattern of industrial location across Spain, with an increasing spatial concentration of industrial activities between the 1850s and the Spanish Civil War (1936-39) and a deeper regional specialisation. What were the forces behind these changes? On the theoretical side, the Heckscher-Ohlin model suggests that the spatial distribution of economic activity is determined by comparative advantage due to factor endowments. In turn, NEG models show the existence of a bell-shaped relationship between the process of market integration and the degree of concentration of industrial activity in the territory. This paper examines empirically the determinants of industrial location in Spain between 1856 and 1929 estimating a model that nests both Heckscher-Ohlin and NEG factors and tests the relative strength of these forces, since they are not mutually exclusive and might be at work simultaneously. The analysis of the results shows that both comparative advantage and NEG-type mechanisms were determinant drivers of industrial location in Spain, although their relative strength changed over time.

Keywords: factor endowments, new economic geography, regional economics, market integration, industry location, economic history

JEL Classification: F1, N6, N9, R12, R3

Resum:

Durant el segle XIX, l'economia espanyola va transitar per les primeres etapes de la industrialització. Aquest procés es va donar en paral·lel a la integració del mercat domèstic de béns i factors, en un moment en què les reformes liberals i la construcció de la xarxa ferroviària, entre d'altres, van generar una important caiguda en els costos de transport. Al mateix temps que es donava aquesta progressiva integració del mercat domèstic espanyol, es van produir canvis significatius en la pauta de localització industrial. D'una banda, hi hagué un augment considerable de la concentració espacial de la indústria des de mitjans de segle XIX i fins a la Guerra Civil, i d'altra, un increment de l'especialització regional. Ara bé, quines van ser les forces que van generar aquests canvis? Des d'un punt de vista teòric, el model de Heckscher-Ohlin suggereix que la distribució a l'espai de l'activitat econòmica ve determinada per l'avantatge comparativa dels territoris en funció de la dotació relativa de factors. Al seu torn, els models de Nova Geografia Econòmica (NEG) mostren l'existència d'una relació en forma de campana entre el procés d'integració econòmica i el grau de concentració geogràfica de l'activitat industrial. Aquest article examina empíricament els determinants de la localització industrial a Espanya entre 1856 i 1929, mitjançant l'estimació d'un model que combina els elements de tipus Heckscher-Ohlin i els factors apuntats des de la NEG, amb l'objectiu de contrastar la força relativa dels arguments vinculats a aquestes dues interpretacions a l'hora de modular la localització de la indústria a Espanya. L'anàlisi dels resultats obtinguts mostra que tant la dotació de factors com els mecanismes de tipus NEG van ser elements determinants que expliquen la distribució geogràfica de la indústria des del segle XIX, tot i que la seva força relativa va anar variant amb el temps.

1. Introduction

The Spanish market became increasingly integrated during the 19th century, when the expansion of the railway network and the technological improvements in sea transport triggered a marked reduction in transport costs. Likewise, this gradual integration of the domestic market was also encouraged by the liberal reforms implemented by successive Spanish governments. By the end of the century this integration was completed.

In turn, the integration of the Spanish market brought about dramatic changes in the geographical distribution of manufacturing activities within Spain, in a period when the Spanish economy was going through the first stages of industrialisation. As various authors have stressed (Nadal, 1987; Parejo, 2001; Paluzie *et al.*, 2004), during the second half of the 19th century and the first decades of the 20th century, there was a significant increase in the spatial concentration of industrial production. In those years, manufacturing activities eventually tended to be agglomerated in a limited number of territories, mainly in the peripheral regions of Catalonia and the Basque Country. In addition, there was a process of deeper regional specialisation from the mid-1850s until the end of the century, although this tendency came to a halt in the interwar years (Betrán, 1999; Tirado *et al.*, 2006). But what were the forces driving the location of different industries across the country in that period?

On the theoretical side, two major explanations stand out in the literature. First, the neoclassical trade theory represented by the Heckscher-Ohlin model suggests that the spatial distribution of economic activity is determined by comparative advantage due to factor endowments. Assuming absence of transport costs, commodities being produced with constant returns to scale, and considering markets operating under perfect competition, the Heckscher-Ohlin theorem (HO) predicts that the distribution of economic activity would be determined by the availability of factor endowments in a location relative to the endowments available at alternative locations. Second, New Economic Geography (NEG) models highlight that the interaction between transport costs, increasing returns to scale and the size of market under a monopolistic competition framework may favour the spatial agglomeration of economic activity.

Several empirical studies, focused on different national experiences and historical periods, have attempted to analyse how the location of industry has responded to the forces stressed by these two explanations, taking into account the fact that comparative advantage and NEG mechanisms are not exclusive and therefore may be at

work simultaneously. The pioneering work of Kim (1995, 1999) examined the long term evolution in the location of manufacturing activities and its determinants in the US between 1860 and 1987. He concluded that industrial location in the United States was mainly explained by the relative differences in resource endowments across regions, thus, limiting the role of increasing returns in shaping the spatial distribution of industry. As regards Spain, Tirado *et al.*, (2002) analysed the determinants of industrial location in the second half of the 19th century following the approach suggested by Kim (1995). These authors showed that at the end of the century, in parallel with the advance of the integration of the domestic market, NEG effects were playing an increasing role in shaping the industrial map of Spain. The evolution of the US and Spain was somewhat similar in terms of manufacturing concentration and regional specialisation. From 1860 to 1930, a period in which regional economies in the US were integrating into a national economy, industry became more localized as regions became more specialised, although regional specialisation flattened out in the interwar years. Nevertheless, Marshallian externalities were more present in the US case before World War II, whereas in Spain Betrán (1999) has shown the relevance of externalities *à la Jacobs* in the interwar years.

Accordingly, Midelfart-Knarvik *et al.*, (2002) studied the changes in the geographical distribution of industry within the European Union between the 1970s and the 1990s. Their results showed that the supply of skilled workers and researchers became progressively more important in attracting industries. On the NEG side, the pull of centrality to industries with increasing returns was significant but decreased over time, as industries with high shares of intermediate goods in production moved to regions with good access to markets. Moreover, the standard model of Midelfart-Knarvik *et al.*, (2000, 2002) allowed the use of an alternative empirical strategy to explore the underlying forces that determine industrial location combining both traditional trade theory and economic geography variables. This approach presents several advantages over the exercise applied by Kim (1995)¹. First, the equation to be estimated is directly derived from a theoretical model. Second, a much larger number of variables are considered in the equation including region and industry characteristics, and more importantly, the interaction between them. And third, the role of the market

¹ A critical evaluation of Kim's approach can be found in Combes *et al.*, (2008), chapter 11.

potential in industrial location decisions, a key element in NEG explanations, can be tested directly.

Furthermore, the analysis of the European Union is particularly relevant. The ongoing process of supranational integration in Europe has raised some concerns about the potential effects of this process on the location of industrial activities, since this may imply uneven adjustments for the territories involved. Therefore, looking back at the period of formation of national markets may shed light on the dynamics and the impact of economic integration on industrial location. Some studies in economic history have made important contributions within this analytical framework.

Wolf (2007) made use of this model in his analysis of the determinants of the industrial relocation in interwar Poland after the reunification in 1918. The integration of the domestic market led to important changes in regional specialisation and the spatial concentration of manufacturing, and his results showed that the forces shaping these changes were similar to the forces operating in the European Union in recent times. Likewise, Crafts and Mulatu (2005, 2006) explored the determinants of industrial location in Britain, also estimating an equation based on Midelfart-Knarvik *et al.*, (2002). In the British case, between the 1870s and the beginning of the 20th century, where the geographical concentration of industry and regional specialisation revealed an evolution characterised by stability, the pattern of industrial location would have mainly responded to factor endowments and natural resources, although it was reinforced by NEG forces (*scale effects*).

Following a similar approach, Klein and Crafts (2009) have recently questioned some of the conclusions obtained by Kim (1995) for the US². In this case, the analysis was focused on the years between 1880 and 1920 in an attempt to explain the emergence of the US manufacturing belt. According to their results, NEG-type mechanisms were at the root of the notable concentration of the manufacturing activity in the north-east of the country. In addition, the paper by Klein and Crafts (2009) has made an important contribution in improving, from a methodological point of view, the empirical strategy; by addressing a good number of econometric issues, the robustness of the findings has increased. In this regard, the next pages follow the path breaking work of these authors.

² Kim (1999) reached the same conclusion in an exercise where the Rybszynski theorem was tested.

In this paper, the empirical strategy developed by Midelfart-Knarvik *et al.*, (2002) is applied in order to study the Spanish experience during the second half of the 19th century and up to the outbreak of the Spanish Civil War (1936-1939), a period when the integration of the domestic market was completed and a notable increase in the spatial concentration of manufacturing activities was recorded. Thus, the Spanish case, with its particular features, can be added to other relevant study cases in countries that have been examined within this framework in order to obtain a broader picture. Moreover, the exercise proposed to analyse the forces driving industrial location in Spain covers a larger period than previous studies undertaken for the Spanish case³. In addition, a more thorough empirical approach, which is directly derived from a theoretical model, is used⁴, and so this alternative approach may help to examine whether the new evidence confirms the results reached in earlier works.

The remainder of the paper is structured as follows. In the next section, the integration of the domestic market in Spain and the main changes in industrial location during the second half of the 19th century and the first decades of the 20th century are reviewed. Next, in section 3, the empirical strategy is defined introducing the basic features of the Midelfart-Knarvik *et al.*, (2002) model and presenting the variables included in the exercise. In section 4, the sources used in the construction of the data set are detailed and the general patterns observed in the variables considered are briefly described. In section 5, the estimation results are shown. At this point, different robustness tests are carried out in order to deal with some econometric issues incorporating the methodological improvements recently suggested by Klein and Crafts (2009). Then, the standardised beta coefficients, which allow assessment of the relative strength of Heckscher-Ohlin and NEG forces, are reported. Section 6 is devoted to the discussion of the main results obtained for the Spanish case. Finally, the paper ends with some concluding remarks.

2. Market integration, transport costs and industrial location in Spain.

During the second half of the 19th century the Spanish market became increasingly integrated. The different liberal governments implemented institutional reforms addressed to reduce restrictions that hindered interregional trade. The removal

³ As described in the next section, Rosés (2003) focused on 1861, Tirado *et al.*, (2002) analysed the second half of the 19th century, and Betrán (1999) studied the interwar years.

⁴ In this case, the exception would be Rosés (2003) who followed Davis and Weinstein (1999, 2003).

of internal tariffs and customs, the abolition of the guilds and the Mesta, the disentailment process both for real estate and subsoil and the unification of the system of weights and measures gradually triggered the integration of the regional markets of goods (Madrazo, 1984; Tedde, 1994; Simpson, 1995; Carreras and Tafunell, 2003).

In addition, transport in Spain traditionally had to face serious difficulties derived from the country's complex geography. Compared with its European neighbours, Spain is a mountainous country with an average altitude around 660 m. above sea level, ranking third in Europe after Switzerland (1,300 m.) and Austria (910 m.), and more than doubling the European average (297 m.)⁵. Apart from the ruggedness of the land, the rivers are characterised by their poor and irregular flow, and inland navigation, the cheapest means of transport before the railway era, was almost non-existent. Therefore, prior to the construction of the railways, there were major obstacles to the development of a modern transport system which did not favour the integration of the Spanish market.

The road network was small, poorly designed since the radial structure hampered the connection between markets, and in a poor state of conservation due to the difficulties of the Spanish public finances. The means of transport were obsolete and the supply of transport, made up of farmers who combined this activity with agricultural work, was insufficient. Together, these factors resulted in a slow, expensive and generally inefficient land transport system (Ringrose, 1970; Gómez Mendoza, 1982)⁶. However, the road network improved substantially in the second half of the 19th century. The total length of paved roads, which in the late eighteenth century was about 2,000 km., increased to 4,000 km. by 1830, to 19,815 km. by 1868, and reached 36,300 km. in 1900 (Madrazo, 1984; Gómez Mendoza and Sanromán, 2005).

But the most prominent impulse for market integration came from the construction of the railway network. The Railway Acts of 1848 and 1855 established the legal framework for the construction of railways and its radial structure around Madrid. The first line was finished in 1848, and over the following decades the basic network was completed. By 1901, all provincial capitals were connected by rail (Wais, 1987). The railway network expanded fast between 1855 and 1866 when the railway connected the main economic centres in Spain. In just ten years, the length of the broad

⁵ Small states like Andorra (1,995 m.) and Liechtenstein (1,100 m.) are also among the most mountainous countries.

⁶ A more optimistic and dynamic view of the state of road transport before 1850 in Spain has been defended by Frax and Madrazo (2001).

gauge network went from 440 to 5,076 km. In a second stage, spanning from 1873-1896, the railway reached other territories and in the late 19th century over 10,827 km. of track had been constructed (Gómez Mendoza and San Román, 2005; Herranz, 2005).

The main effect of the new infrastructures was the decrease in transport costs. According to the calculations of Herranz (2005), the ratio between the unit price of railway transport for commodities and the alternative mean of transport was around 0.14 in 1878, meaning that the introduction of railways represented an 86% reduction in transport prices. Moreover, coastal navigation, which represented about 20% of the volume of commodities transported by railway over this period (Frax, 1981), also experienced a decrease in freight rates (Gómez Mendoza, 1981; Herranz, 2005). In this case, transport freights were affected by the strong competition between coastal navigation and railways, especially on the routes along the Mediterranean coast. During this period railway companies tried to attract a greater volume of trade at the expense of coastal shipping, resulting in a reduction in railway and maritime freights (Frax, 1981; Pascual, 1990). Furthermore, coastal navigation underwent major changes in the second half of the 19th century; the introduction of iron increased transport capacity and the replacement of sail by steam resulted in considerable time savings. These technological innovations were complemented by the improvement of the major ports to allow the berthing of vessels of greater tonnage (Gómez Mendoza, 1981; Herranz, 2004, 2007). As a result, the total volume of goods transported by coastal trade raised from 690,000 tn. in 1857 to 2.02 million tn. in 1900.

In turn, falling transport costs led to a gradual integration of the domestic market for goods, characterised by a convergence in regional prices (Sánchez-Albornoz, 1975; Peña and Sánchez-Albornoz, 1984; Barquín, 1997; Martínez Vara, 1999; Reher, 2001; Llopis and Sotoca, 2005; Matilla *et al.*, 2008). In parallel, capital and labour markets were also becoming increasingly integrated. In the capital markets, the reforms affected both the monetary and the banking system. Since the late 18th century there had been a large number of different currencies in circulation in Spain, and this situation continued for the first third of the 19th century (Sardà, 1948); in 1864 there were still 84 different currencies in circulation (Martorell, 2001). However, the decree passed in 1868 by the Finance Minister Laureano Figuerola, unified the Spanish monetary system around a single currency: the peseta.

Besides, during the 18th and 19th centuries the Spanish banking system was underdeveloped, as shown by the persistence in the use of bills of exchange. The

convergence and the reduction in the fluctuations of interest rates for bills of exchange between different regional markets has been identified as a signal of a progressive integration of the monetary system during the second half of the 19th century (Castañeda and Tafunell, 1993; Maixé-Altés and Iglesias, 2009). In addition, the new legal framework of the 1840s and the 1850s allowed the creation of private banks as limited liability companies (Tortella, 1973) and, from 1856 onwards, gave provincial banks the right to issue banknotes (Sudrià, 1994)⁷. With the restoration of the monarchy in 1874 new measures were adopted. The Echegaray Decree of 1874 granted the monopoly of issuing money to the Bank of Spain. At the same time, the Bank of Spain gradually established branches in various provincial capitals and in 1885 a system of transfers between accounts at different branches of the Bank was created (Tortella, 1973). Then, the system based on bills of exchange was finally substituted, favouring the integration of the capital market (Castañeda and Tafunell, 1993). In a similar vein, the expansion of the telegraph increased the speed at which information could be transmitted, thus reducing transaction costs for the companies (Calvo, 2001). Between 1855 and 1900, the length of the telegraph lines in Spain increased from 713 to 32,320 km. (Herranz, 2004).

As regards the integration of the labour market, following Boyer and Hatton (1997) Rosés and Sánchez-Alonso (2004) suggested that an increasing integration in the labour market was possible even in a context of a limited mobility of workers, as in the case of Spain. Internal migrations remained at very low levels until the 1920s, when the economic growth of the Spanish economy and industrialisation opened up new opportunities for workers (Erdozáin and Mikelarena, 1996; Silvestre, 2005)⁸. In addition, Silvestre (2007) has emphasized the relevance of temporary internal migrations over the period 1860-1930. International migration also rose notably in the first decade of the 20th century and up to World War I, although it came to a halt in the interwar years with the ‘globalization backlash’ (Sánchez-Alonso, 2000). In this context, Rosés and Sánchez-Alonso (2004) found evidence in favour of the existence of β -convergence in regional real wages from 1860 to 1930, the only exception to this

⁷ Prior to 1856, only two banks (Barcelona and Cádiz) together with the newly created Bank of Spain substituting the *Nuevo Banco Español de San Fernando*, enjoyed the privilege of issuing money. In 1874, there were 20 issuing banks.

⁸ As was the norm in the Mediterranean countries of southern Europe (Hatton and Williamson, 1998; O’Rourke and Williamson, 1999). The reasons for the modest internal mobility appear to have been the late industrialisation process and the slow structural change in these economies. The debate in the Spanish literature can be followed in Silvestre (2005).

trend being found in the period covering the World War I years. As a result, the dispersion of regional real wages (σ -convergence) also decreased. On the basis of this evolution, these authors suggested a high level of integration for the Spanish labour market in that period.

But how did the location of industry respond to falling transport costs and the integration of the domestic market for goods and factors? Did industry become more geographically concentrated? Did the productive structure of Spanish regions converge? Several studies have focused on these questions. Paluzie *et al.*, (2004) analysed the spatial distribution of industry across Spain in the long term, finding that the geographical concentration of industry increased from 1856 to 1929. Then, between 1955 and 1975, under the Franco regime, this trend came to a halt and few significant changes are observed in that period. Finally, from the 1980s onwards industrial production tended to be more spatially dispersed in a framework characterised by a strong restructuring of the industrial sector and the beginning of the process of European integration. Therefore, the Spanish experience shows in the long run a bell-shaped non-monotonic relationship between market integration and the spatial distribution of industry⁹. In particular, it was in the first stages of the industrialisation process, in parallel to the deeper integration of the Spanish domestic market, that industrial production became increasingly concentrated in a limited number of provinces.

Tirado *et al.*, (2006) computed industrial intensity indices in order to identify the provinces in which manufacturing production was concentrating. The results showed that the number of provinces with an industrial specialisation decreased between 1856 (14 provinces) and 1893 (9 provinces), and also from 1913 (8 provinces) to 1929 (7 provinces). Thus, the integration of the Spanish market and the take-off of the industrialisation process led to important changes in the geographical pattern of industrial concentration. During the second half of the 19th century an industrial axis centred in the eastern and north-eastern Mediterranean provinces appeared. Nevertheless, in the first decades of the 20th century, the Mediterranean axis weakened and new locations like Saragossa emerged as industrial centres together with Catalonia, the Basque Country, Madrid and Valencia. As regards regional specialisation in Spain,

⁹ In the case of the US, there was also evidence of a bell-shaped curve in manufacturing concentration in the long term (Kim, 1995). Similarly, a non-monotonic evolution in the geographical concentration of manufacturing was also found for France (Combes *et al.*, 2008).

Tirado *et al.*, (2006) demonstrated that it increased between 1856 and 1893. However, after World War I this trend was reversed and no further differences in the productive structure between regions were observed between 1913 and 1929. The same conclusion was reached by Betrán (1999) in her study of the industrial location in Spain during the interwar years.

Finally, when individual industries were studied, the evidence provided by Paluzie *et al.*, (2004) showed that most of the seven industries considered shared an increase in the geographical concentration between 1856 and 1929. The only exceptions were *wood/furniture* throughout the period and *paper* between 1856 and 1893. At the starting date of their analysis, the industry that was most spatially concentrated was *textiles*, the leading industry in the first stages of the industrialisation process in Spain. In turn, at the end of the period, *metallurgy* and *chemicals*, two sectors linked more to the Second Industrial Revolution, were approaching the levels of concentration observed for *textiles*. Besides, the Gini index in these industries was very close to one, indicating an almost complete concentration of the activities. That would be the case of *textiles*, which increasingly agglomerated around Barcelona. Conversely, the most dispersed industry in the years under study was *foodstuffs*.

Overall, this evidence suggests that the integration of the Spanish market triggered an intense geographical concentration of industrial activities and an increase in regional specialisation, although the latter does not seem to have continued to rise during the interwar years. Several empirical studies have tried to explain the forces driving this notable increase in the geographical concentration of industry. Rosés (2003) argued that the industrialisation of the Spanish regions in mid-19th century was the result of a combination of comparative advantage and increasing returns. Following Davis and Weinstein (1999, 2003), this author tested the existence of a ‘*home market effect*’, concluding that new modern manufacturing industries characterised by increasing returns to scale tended to be concentrated in regions in which home-market effects were larger. Therefore, regions such as Catalonia, where new industries showing increasing returns were established, reinforced its initial comparative advantage in terms of human capital endowments.

In turn, Tirado *et al.*, (2002) studied the forces that shaped the increase in geographical concentration in manufacturing during the second half of the 19th century. Applying an empirical approach similar to the one implemented by Kim (1995) for the US experience, these authors compared two time points, 1856 and 1893. They found

that in 1856 human capital endowment was not significant in explaining the relative industrial intensity of Spanish provinces, but by 1893 this variable became significant, possibly reflecting, as argued by these authors, the importance of technological skills. Furthermore, NEG effects were also relevant. First, the provinces specialised in sectors where economies of scale were important showed a relatively higher industrial intensity. Second, the impact of economies of scale increased. Finally, market size was a relevant factor for regional specialisation in 1856 but at the end of the century, as the Spanish market became more integrated, access to markets turned out to be more important.

The relevance of different explanations behind the location of industry in the interwar period was analysed by Betrán (1999). In this case, the increase in the manufacturing per capita of the Spanish provinces between 1913 and 1929 was explored in terms of the role played by both Marshallian and Jacobs externalities. Her results showed that inter-industrial relations (*à la Jacobs*) were significant and more important than intra-industrial relations (*à la Marshall*). Hence, in a period when the process of further regional specialisation stopped, the diversification of the industrial structure in the Spanish provinces had a positive effect on industrial growth: the less specialised provinces, where intermediate and capital goods were more prominent, experienced a higher rate of growth in their manufacturing per capita.

Within this context, the methodology developed by Midelfart-Knarvik *et al.*, (2002) allows an analysis of the determinants of industrial location in Spain and complements the studies described in the final part of this section. This approach has some advantages: first, it is based on a sound empirical test in the sense that it is directly derived from a theoretical model; second, the exercise covers a much longer period, the years between the mid-19th century and the Spanish Civil War. Therefore, the entire period of increasing spatial concentration in manufacturing activities in Spain can be examined.

3. The location of industry and the interaction between region and industry characteristics.

Traditional trade theory explains the existence of trade flows and international specialisation assuming markets operating under perfect competition and constant returns to scale. In the Ricardian model, countries will specialise in the production of goods according to their comparative advantage and trade will be the result of the differences between countries in terms of technology and productivity. Therefore,

countries will specialise in those goods that are produced more efficiently within their borders. Accordingly, the Heckscher-Ohlin theory assumes, in a two-country, two-sector and two-production goods framework, that countries have similar access to technologies but they differ in factor endowments. In short, taking two production factors, labour and capital, with trade liberalisation the country where labour is relatively abundant will specialise in the production of the good that uses this factor intensively. The contrary holds in a country where capital is relatively abundant, and a capital-abundant location will specialise in the production of capital-intensive goods. The result of this specialisation will be, as the Factor Price Equalisation (FPE) theorem suggests, that the relative price of the production factors will eventually tend to be the same in both countries.

As regards the second explanation, New Economic Geography models stress that the interaction between transport costs, increasing returns to scale and the size of market under a monopolistic competition framework can lead to the agglomeration of economic activity. In this context, the distribution of economic activity in space is shaped by the existence of two types of forces operating in different directions: *agglomeration* or *centripetal forces*, and *dispersion* or *centrifugal forces*. As shown by Krugman (1991), the concentration of economic activity would be a result of the interaction of centripetal forces. First, in order to save on transport costs, firms tend to agglomerate near the locations with better access to markets, that is, close to large centres of demand. Thus, this increase in the market size generates a more than proportional increase in the share of firms in that location¹⁰, pushing nominal wages upward. The increase in the number of firms allows for a greater variety of local goods and consumption can benefit from lower transport costs. Consequently, the lower local price index and the resulting increase in real wages attract new workers to the urban centres. Hence, access to markets, or market potential, has a positive influence in the location decisions made by firms and workers, inducing factor mobility (capital and labour, respectively), and leading to a cumulative process, where agglomeration is reinforced once it has started.

This market access mechanism is amplified by the demand for intermediate goods. Locations with a large number of firms and therefore, a large demand of intermediates, will be more attractive for intermediate producers. Conversely, locations with a large number of intermediate producers would be preferred by firms producing

¹⁰ This property is known as the ‘*home-market effect*’. Helpman and Krugman (1985).

final goods. In this case, the lower transport costs make inputs cheaper. When intermediate goods are considered, the presence of *input-output linkages* emerges as an alternative mechanism favouring agglomeration (Krugman and Venables, 1995; Venables, 1996). Nevertheless, when transport costs reach a sufficiently low level, new *dispersion* or *centrifugal forces* may generate a dispersion of manufacturing production.

This idea of a bell-shaped evolution in the spatial concentration of manufacturing was further developed by Puga (1999)¹¹. This author showed that the relationship between the process of regional integration and the degree of concentration in the economic activity can describe a non-monotonic evolution. When transport costs are high, industry is dispersed across space. When transport costs fall to an ‘intermediate’ level, *centripetal forces* intensify agglomeration when workers are mobile. For low transport costs, a new tendency towards dispersion can emerge. Congestion costs, wage differentials, fragmentation of firms, or non-economic motivations affecting the decision to migrate (*amenities*), act as *centrifugal forces* favouring the dispersion of economic activity.

The approach to be implemented in order to study the spatial distribution of industry in Spain is based on the model suggested by Midelfart-Knarvik *et al.*, (2000), which nests both comparative advantage and New Economic Geography variables¹². The equation to be estimated includes a set of region and industry characteristics, and more interestingly, the interaction between them as determinants of the pattern of industrial location. Regions, or provinces as in the Spanish case, are heterogeneous and therefore differ in factor endowments and in their relative access to demand in the presence of transport costs. Hence, some regions have a comparative advantage in the sense that they are relatively abundant in land; others are relatively abundant in labour; and some regions have an advantage in terms of proximity to markets. Similarly, industries also differ in their attributes, for instance, the intensity in the use of agricultural inputs or skilled workers, the size of the plants, or the share of intermediate inputs. The model developed by Midelfart-Knarvik *et al.*, (2000, 2002) provides the equation to be estimated in the empirical exercises, which can be expressed as:

¹¹ Puga (1999) combines the assumptions of interregional mobility of labour as in Krugman (1991) and *input-output linkages* as in Krugman and Venables (1995) and Venables (1996). In addition, his model considers the presence of mobility between sectors. Thus, this is a more suitable framework for regional studies.

¹² Here, the basic features of the original model proposed by these authors are reproduced. For more details, see Midelfart-Knarvik *et al.*, (2000), and Midelfart-Knarvik *et al.*, (2002), p. 243-245.

$$\ln(s_i^k) = c + \alpha \ln(pop_i) + \beta \ln(man_i) + \sum_j (\beta[j]y[j]_i z[j]^k - \beta[j]\gamma[j]z[j]^k - \beta[j]\kappa[j]y[j]_i) \quad (1)$$

where s_i^k is the share of industry k in region i . The first two variables (pop_i , man_i) take into account size effects meaning that, all other things being equal, larger regions are expected to have a larger population and a larger industrial share in any given industry: pop_i is the share of Spain's population living in region i , and man_i is the share of the total of Spain's manufacturing located in region i . In turn, $y[j]_i$ is the level of the j th region characteristic in region i , $z[j]^k$ is the industry k value of industry characteristic paired with region characteristic j , and therefore, $y[j]_i z[j]^k$, is the interaction between region and industry characteristics. Finally, c is the constant term, and α , β , $\beta[j]$, $\gamma[j]$ and $\kappa[j]$ are the coefficients. Among them, the estimated coefficients $-\beta[j]\kappa[j]$ correspond to the region characteristics $y[j]_i$, and the coefficients $-\beta[j]\gamma[j]$ are associated with the industry characteristics $z[j]^k$. However, the most relevant information comes from the estimated coefficients $\beta[j]$ for the interactions between region and industry characteristics since these interactions determine industrial location¹³.

The analysis of the determinants of industrial location in Spain between 1856 and 1929 takes into account the region and industry characteristics displayed in table 1. The first three variables in the column where region characteristics are included capture relative factor endowments according to the Heckscher-Ohlin model. The variable agricultural production as a percentage of GDP is considered to be a measure of 'agriculture abundance', that is, the relative abundance of land across provinces¹⁴. The second factor endowment variable refers to the abundance of labour in each province measured by total active population per square kilometre. Educated population, the third Heckscher-Ohlin variable, measures the relative endowment of human capital proxied

¹³ This approach has, nonetheless, some limitations. In particular, it is assumed that all sectors are perfectly competitive: "...we make this assumption in order to have a precise and tractable link between the theory and econometrics, whereas adding imperfect competition would raise a number of further issues which go beyond the scope of this paper. For example, in such an environment there is, in general, a multiplicity of equilibria, and hence no unique mapping from underlying characteristics of countries and industries to industrial location". Midelfart-Knarvik *et al.*, (2000), p. 3.

¹⁴ "We take agricultural production as an exogenous measure of 'agriculture abundance' (rather than going back to an underlying endowment such as land)". Midelfart-Knarvik *et al.*, (2002), p. 243. As usual in this type of exercise, capital is not included among the variables considered based on the assumption of capital mobility across regions.

by the literacy rate. Finally, the fourth region characteristic, market potential, is a NEG variable which captures the advantage of a province in terms of access to markets. As regards the six industry characteristics in table 1, the first three variables are again based on Heckscher-Ohlin theory, and they reflect the intensity in the use of production factors (land, labour and skilled workers). The last three variables are related to NEG forces: the presence of economies of scale, the share of intermediate inputs used, and sales to industry.

[HERE TABLE 1]

How do these region and industry characteristics interact to determine the location of industry? In the present study, a total of six interactions can be examined. The first three interactions are based on comparative advantage theory. The Heckscher-Ohlin model predicts that industries using intensively a production factor (agricultural inputs, labour, and skilled labour) will tend to be established in locations with a better relative endowment of these factors.

The next three interactions consider NEG-type mechanisms. The interaction between *market potential x size of establishment* captures the idea in Krugman (1991) that industries with higher economies of scale or increasing returns tend to concentrate near the demand, in regions characterised by good access to markets. When transport costs are sufficiently low (reaching ‘intermediate’ levels) firms operating at a large scale are induced to locate close to big markets and then supply the demand from these locations. If transport costs were either very high or very low, then the pull of market potential may be weakened. The last two NEG interactions take into account the presence of intermediate goods (Krugman and Venables, 1995; Venables, 1996). First, *market potential x intermediate goods* is based on the NEG account that industries that use a high proportion of intermediate inputs are more likely to be spatially concentrated. These *forward linkages* arise when firms that buy inputs from other producers as intermediate goods locate, in order to minimise transport costs, near central areas where a higher number of suppliers will be concentrated. Thus, a bigger market implies better access to suppliers. On the other hand, firms may also produce goods for other industrial users. The interaction *market potential x sales to industry* is related to *backward linkages*. Industries that sell a high proportion of their output to other firms would prefer to locate close to other producers. In other words, firms prefer to be close to their industrial customers in order to minimise transport costs.

4. Comparative advantage and NEG variables: data and description of the patterns.

The analysis of the determinants of the spatial distribution of industry in Spain is carried out for a total of 43 provinces (NUTS3)¹⁵, seven industrial sectors (*foodstuffs, textiles, metal, chemicals, paper, ceramics/glass, and wood/cork*) and four benchmark years: 1856, 1893, 1913 and 1929. In order to explain the patterns of industrial location, the endogenous variable considered, (s_i^k) , is the share of a certain industry k in the total manufacturing activity of region (province) i , defined as:

$$s_i^k(t) = \frac{x_i^k(t)}{\sum_k x_i^k(t)}, \quad (2)$$

where $x_i^k(t)$ is the level of industrial activity k at location i and time t . This indicator is constructed based on the fiscal information provided by the *Estadística Administrativa de la Contribución Industrial y de Comercio (EACI)*. This publication compiles the total tax paid for the industrial activity by province and industry. Established in 1845, the tax consisted of “a system of fixed rates per active unit of the main production means in each of the branches or productive processes established by the legislator”¹⁶. The share that a certain industry represents in each province is calculated through the aggregation of the tax paid in that industry over the total amount paid in the province. Information for the years 1856 and 1893 has been collected from this source.

[HERE FIGURE 1]

However, the treatment of the data provided in the EACI has to face two main problems: first, the exclusion of the Basque Country (containing the provinces of Álava, Guipúzcoa and Vizcaya) and Navarre, exempt from the payment of the tax because they

¹⁵ Seven provinces are excluded from the sample. The three Basque provinces and Navarre are absent due to the lack of information to construct the variable (s_i^k) . The Balearic Islands and the two provinces within the Canary Islands are not included since estimates of market potential are not available for these insular territories.

¹⁶ Nadal and Tafunell (1992), p. 256. Own translation.

had a special fiscal regime¹⁷. Second, this tax underwent major changes in 1907. Therefore, for the last two years in this study, 1913 and 1929, after the change of legislation, the EACI is not a fully satisfactory source. This problem has been dealt with and overcome by Betrán (1995, 1999) in her in-depth analysis of the industrial localisation in Spain during the first decades of the 20th century, where she reconstructed the industrial taxes paid in each province in 1913 and 1929, based on the two taxes existing at that time: *Estadística Administrativa de la Contribución Industrial y de Comercio (EACI)* and *Estadística de la Contribución de Utilidades (ECU)*. Hence, data for 1913 and 1929 are obtained from Betrán (1995, 1999).

As regards the size variables, the share of Spain's total manufacturing located in province *i* comes from Martínez-Galarraga (2007), where regional GDP estimates were estimated based on Geary and Stark's (2002) methodology. The database covers NUTS3 provinces and the three main economic sectors: agriculture, industry and services. The second size variable, the share of Spain's population living in province *i*, is obtained from the Censuses of Population (1860, 1900, 1910 and 1930).

Industry characteristics are derived from various sources. A good number of the variables reported in table 1 are based on input-output relationships. Unfortunately, the first input-output table for the Spanish economy was published in 1958. Thus, industry characteristics for that year are applied retrospectively to the period considered in this exercise, assuming that the intensity in the use of factors and other inputs is representative for previous periods. Nevertheless, this assumption is customary in relevant studies in Spanish economic history. Carreras (1983) made use of this source in preparing the historical industrial production index for Spain, and Prados de la Escosura (2003) employed this information in his reconstruction of historical series for the Spanish GDP starting in 1850. Hence, the variables agricultural input as a percentage of total costs, labour input as a percentage of GVA, and intermediate goods and sales to industry both as a percentage of output, are obtained by taking the information provided by the input-output table of 1958.

In addition, data on the share of white-collar workers over total workers in each one of the seven industries considered comes from the Industrial Statistics for 1958 (*Estadística Industrial de 1958*). Finally, for the variable size of establishment, usually

¹⁷ The absence of the Basque Country is a real drawback, given the development of a strong metal industry from the 1870s onwards. Between the 1870s and the Spanish Civil War this region became one of the main industrial centres in Spain, alongside Catalonia.

taken as a proxy for economies of scale, the source consulted is again the *Estadística Administrativa de la Contribución Industrial y de Comercio (EACI)*. This publication includes information on the tax quota paid in one particular industry and the total number of contributors. In this case, the data collected from EACI corresponds to the two first benchmark years, 1856 and 1893. Then, the values for 1893 are applied to the years 1913 and 1929¹⁸. Therefore, for the set of industry characteristics considered, the information is not time-varying; the only partially time-varying industry characteristic is the size of establishment¹⁹.

Region (or province) characteristics, in which time-varying information is used, include agricultural abundance as a proxy of land endowment, labour and skilled labour abundance and market potential. Land endowment is calculated taking the agricultural production (Gross Value Added at factor cost) as a percentage of GDP in each Spanish province. As in the case of the manufacturing size variable, this dataset has been constructed in Martínez-Galarraga (2007). Labour abundance is measured through the total active population in each province per square kilometre. In this instance, information on the total active population by province has been compiled from the Censuses of Population and the provincial area in square kilometres from the Statistics National Institute of Spain (*Instituto Nacional de Estadística-INE*). Third, skill labour endowment is proxied by the literacy rates offered by Núñez (1992).

Lastly, estimates of market potential for the continental Spanish provinces come from Martínez-Galarraga (2009), following the recent work by Crafts (2005) for Britain from 1871 to 1931²⁰. For the Spanish case, market potential estimates are provided for 47 NUTS3 provinces (insular territories excluded) in the 1860s, 1900, 1914 and 1930. Regional accessibility is measured through the Harris (1954) market potential equation, which can be defined as:

$$MP_r \equiv \sum_s \frac{M_s}{d_{rs}} \quad (3)$$

¹⁸ The reason lies again in the changes produced in this tax after 1907. Nadal and Tafunell (1992).

¹⁹ This is quite usual in this type of exercise, even in those focused on recent periods: “...getting data on industry characteristics is not simple, so we use information on intensities that is not time-varying”. Midelfart-Knarvik *et al.*, (2002), p. 245.

²⁰ See also Keeble *et al.*, (1982). For historical exercises, see Crafts (2005) and Schulze (2007). In the case of Wolf (2007), estimates of market potential (both for *market access* and *supplier access*) for the Polish regions in the interwar period were calculated, following Redding and Venables (2004), using data on regional bilateral flows. Unfortunately, for Spain this information is not available.

where M_s is a measure of the size of province s (usually GDP) and d_{rs} is the distance, or as in this case, bilateral transport costs between r and s . Following the previous studies, market potential can be divided into two main components: internal and external market potential. First, the internal market potential includes the Spanish provinces, considering the *self-potential* of province r , where the internal distance is calculated as:

$$d_{rr} = 0.333 \sqrt{\frac{(\text{area of province}_r)}{\pi}} \quad (4)$$

The size of provincial markets (M_s) is measured by aggregate income. Data on GDP at a NUTS3 level are obtained from Martínez-Galarraga (2007). For d_{rs} , transport costs are considered. In this case, information on distances and average transport rates for commodities are needed. Internal transport is assumed to be by railway and coastal shipping. For railway distances, the sources consulted are Wais (1987) and Ministerio de Obras Públicas (1902). However, in the first benchmark year, in 1867, only 32 out of the 47 continental provinces were connected to the railway network. For that reason, road transport was also included in the internal market potential estimates in that year²¹. Distances by road in the 1860s are taken from Dirección General de Obras Públicas (1861). For distances between maritime ports, electronic atlases provide information on the length of sea journeys. As regards transport costs, data on railway rates come from Herranz (2005)²² and coastal shipping rates in 1865 have been obtained from Nadal (1975). In order to consider the reduction in sea transport costs, data have been corrected with the freight rate indices calculated by Mohammed and Williamson (2004). For road transport prices in the 1860s, the information in Barquín (1999) has been used. Finally, the relative weight of each transport mode (railways and coastal shipping) in the coastal provinces is obtained from Frax (1981).

Second, foreign markets have to be added to the internal market potential. The construction of the external market potential is based on the elasticities obtained in the gravity equations for international trade estimated by Estevadeordal *et al.*, (2003). In a simple version, gravity equations consider that the intensity in trade flows between two countries is positively related with the size of the economies involved and inversely

²¹ In 1930, the role of motorized road transport remained negligible (Herranz, 2005) and therefore is not considered here.

²² Unpublished data kindly provided by the author.

related with distance and tariffs. Taking this reduced specification of a gravity equation, the external market potential is calculated as:

$$\varphi_{rs} = GDP_s (distance)_{rs}^{\delta} (tariffs)_s^{\gamma} \quad (5)$$

where GDP of the main trading partners for Spain is obtained from Crafts (2005) based on the estimates of Prados de la Escosura (2000); maritime distances are once again taken from electronic atlases; tariffs come from O'Rourke (2000) and Mitchell (1998a, 1998b); and finally, the elasticities for the distance and tariffs, δ and γ , respectively, from Estevadeordal *et al.*, (2003).

Before implementing the empirical strategy, the main patterns and evolution of the variables considered in the exercise are briefly described. Following the order shown in table 1, region characteristics are examined first. As regards comparative advantage, the relative abundance of land across provinces is measured by agricultural production. In 1860 the picture that emerges is one of a general agrarian specialisation in Spanish provinces where the more agricultural provinces were mainly located in the north-west and also in some provinces in the north-east (Lleida, Castellón, Teruel and Huesca). Throughout the period covered in this study, the north-western provinces stand out in terms of agriculture abundance (particularly Ourense and Lugo) and towards the end of the period some Castilian provinces like Cuenca also showed a high agrarian specialisation. Provinces with the lowest levels of agricultural production were Madrid, Barcelona, the northern coast and western Andalusia.

Labour abundance is measured by the total active population per square km. In this case, a clear geographical pattern emerges in which the coastal provinces show in general a higher abundance of labour than inland provinces. The labour abundant provinces are spread across the northern and north-western coasts (Vizcaya, Guipúzcoa, Coruña and Pontevedra), the Mediterranean coast (Barcelona, Valencia and Alicante) and Madrid, the only inland province, which showed a notable increase after 1900.

The spatial distribution of educated population in Spain showed an increasing polarisation over the period considered (Núñez, 1992). In 1860, literacy rates were higher in the area going from Madrid, through Castile-Leon (Burgos, Palencia, Segovia, Valladolid and Soria) and northern Spain (Cantabria and Álava). Interestingly, although it was the main industrial province, Barcelona was not among the high literacy

provinces. However, by 1900, literacy rates in Catalonia and Aragon had risen, leading to a marked division of the country between the north and the south in terms of education. The only exceptions were Galicia in the north-west with a lower literacy rate than the rest of the northern provinces, and south-western Andalusia (Cádiz, Huelva and Seville) where literacy rates were the highest among the southern provinces. In 1930, the transition to universal literacy was almost completed in the north, and the south had started to close the gap with the northern provinces although illiteracy was still an important issue on the eve of the Spanish Civil War.

Market potential is the NEG region characteristic considered in this study. This variable captures the advantage of a province in terms of its accessibility to markets. Did the reduction in transport costs change the relative market potential of Spanish provinces? During the second half of the 19th century, there were major variations in the geographical pattern of market potential as a result of falling transport costs (figure 2). In the 1860s, three different groups of provinces can be found: coastal provinces characterised by a high market potential, inland provinces connected to the railway network, and inland provinces with no access to the railway. In 1900, the picture had changed shifting to a more polarised spatial structure: a first group containing coastal provinces with a higher relative market potential (with the addition of Madrid) and a second group consisting of inland provinces with a lower relative market potential. Once established, this structure became persistent in the first three decades of the 20th century. Thus, the major changes in the market potential of Spanish provinces occurred during the second half of the 19th century when the Spanish market was integrating, the basic railway network was constructed, and transport costs were falling fast.

[HERE FIGURE 2]

Looking at the different industries and their attributes, several features are worth noting. Among the comparative advantage variables, *foodstuffs* stand out for the high intensity in the use of agricultural inputs. *Wood/cork* is the most labour intensive industry and *chemicals* the least. In turn, *chemicals* and *metal* are skilled labour intensive industries, while *wood/cork* and *ceramics/glass* use intensively unskilled labour. NEG variables show that economies of scale are most important for *metal* and *paper* industries; the use of intermediates is higher in *foodstuffs* and *textiles*; and finally, sales to industry are particularly relevant in *ceramics/glass*.

5. Estimation results.

Equation (1) is estimated by OLS. Since there are two potential sources of heteroskedasticity, both across provinces and across industries, White heteroskedasticity-robust standard errors have been used²³. However, Klein and Crafts (2009) have identified several additional econometric issues that may arise in the estimation of this standard equation. Among them, the presence of a clustered sample, the poolability of data across time and endogeneity, stand out. These problems are addressed in this section following the methodological improvements suggested by these authors.

On the one hand, the observations included in the regressions are not spatially independent since seven industries for each province are included, that is, seven observations corresponding to the same province. This may generate problems with the standard errors obtained in the estimation²⁴. The within group or intra-cluster dependence can be addressed using cluster-robust standard errors (Arellano, 1987), in this case, at a provincial level.

First, a pooled OLS estimation is carried out pooling over the seven industries and the four years considered, giving a total of 1204 observations, using both White heteroskedasticity-robust standard errors and cluster-robust standard errors. Table 2 shows these first estimation results, including the constant term, the coefficients of the two size variables (pop_i, man_i), the four region characteristics, $y[j]_i$, the six industry characteristics, $z[j]^k$, and the six interaction variables. The $\beta[j]$ coefficients associated with the interaction variables, which capture the combined effect of region (province) and industry characteristics on the location of industry, contain the most relevant information. The results for the pooled sample (1856-1929) are reported in column 1. When the whole period is considered, two of the six interactions included are statistically significant and among the Heckscher-Ohlin variables, agriculture is significant and has the correct sign. As regards NEG variables, the significant interaction is the one relating market potential and size of establishment, showing the

²³ The Breusch-Pagan/Godfrey test displayed in table 2 confirms the presence of heteroskedasticity since the null hypothesis of homoskedasticity is rejected at the 1% significance level (5% in 1893).

²⁴ “Research that ignores potential correlation between respondents sharing the same cluster may draw distorted inferences”. Pepper (2002), p. 342.

expected positive sign. Therefore, these results suggest that both HO and NEG factors had a significant impact on the location of industry in Spain.

[HERE TABLE 2]

However, pooling data across years may generate problems since it implicitly assumes that the parameters of the equation are constant over time²⁵. This concern is important in this case since the time-span considered in the study of the Spanish industry is quite long, covering more than 70 years. Thus, the assumption of constant coefficients across time needs to be tested. In order to decide whether it is appropriate or not to pool the data, a Chow test is performed, and the null hypothesis that there is no structural break in this period is rejected²⁶. Hence, equation (1) is again estimated by OLS, with heteroskedasticity-robust standard errors using White's method and cluster-robust standard errors, pooling across industries, for each particular year: 1856, 1893, 1913 and 1929.

The results are displayed in columns 2-5 in table 2. In this case, the total number of observations is, depending on the year, around 300. The goodness of the fit in terms of the adjusted- R^2 , ranging from 0.58 to 0.68, is acceptable when compared to other similar exercises. The first three interactions are based on comparative advantage considerations. The coefficient for the interaction relating agricultural abundance and agricultural input use has the right positive sign, and it is statistically significant in the first and the last year of study. This result shows that, for those particular dates, manufacturing industries that made intensive use of agricultural inputs were located in regions with a relatively good endowment for agricultural production. Changes in regional specialisation might explain this evolution, as it will be argued in the discussion. The second Heckscher-Ohlin interaction refers to labour abundance. In this case, the coefficient is significantly different from zero in 1893 and 1913, showing a positive sign as expected. Thus, those regions with a relatively larger abundance of labour attracted industries which were intensively reliant on this production factor.

²⁵ “...there are three potential sources of variation in the underlying system –the characteristics that define the reference country can change [...], those defining the reference industry can change [...], or industries can become more or less responsive to country and industry characteristics, so $\beta[j]$ changes”. Midelfart-Knarvik *et al.*, (2000), p. 16.

²⁶ The Chow test for structural change uses an F-test to determine whether the coefficients in a regression are the same in two sub-samples (in this case, 1856-1893 and 1893-1929). The calculated value of the F-statistic $F(19, 1119)$ is 12.75423, and therefore the null hypothesis is rejected at the 1% significance level.

However, in 1856 and 1929, this effect vanishes since this interaction is not significant. Finally, the interaction for skilled labour is significant in 1856, although it shows a negative sign, suggesting that literacy rates did not have an impact on industrial location at the start of the industrialisation process²⁷. Overall, there seems to be evidence in favour of traditional factor endowment effects on the location of industry in Spain throughout the period, although these forces varied in the different years under study.

The last three interactions capture NEG-type mechanisms. In 1856, the variable that relates market potential and the size of establishment is insignificant. However, from 1893 to 1929, this interaction becomes significant and exhibits a positive sign. Thus, at that time, industries with increasing returns to scale tended to be located near the central areas in terms of higher market potential. The magnitude of the coefficient associated to this interaction increased between 1893 and 1913 but then, decreased in the interwar years. Conversely, no evidence of *linkage effects* is found. The interactions between market potential and the share of intermediate inputs used and between market potential and sales to industry are always statistically insignificant and in some years have the wrong sign. Hence, *backward linkages* were not important determinants of industrial location in this period, i.e., industries which sell a large share of their output to industry did not tend to locate in provinces with a high market potential. The same result is obtained for *forward linkages*: industries which are heavily dependent on intermediate goods did not tend to locate in high market potential provinces with good access to intermediate inputs. According to these results, NEG forces were not at work in mid-19th century Spain, although they become significant afterwards through the interaction between market potential and economies of scale.

However, these estimation results can be biased in the presence of endogeneity, that is, if some explanatory variables and the residuals of the regression are correlated. In NEG empirical studies endogeneity often arises as a result of the self-reinforcing nature of the process described in the theoretical models, leading to reverse causality. In the current specification the potentially endogenous variable is market potential and, consequently, the interactions capturing NEG effects. A location with good access to markets will attract industrial activities and this, in turn, will increase the market potential of this location (through the domestic component of the market potential equation). Therefore, an endogeneity test on the potentially endogenous regressors is

²⁷ Note that this interaction is significant in column ‘a’ for 1913, but not when the estimation is corrected taking into account cluster-effects.

performed. According to the Durbin-Wu-Hausman endogeneity test the market potential variable and its interactions are endogenous in 1893 and 1913 but exogenous in 1929²⁸.

When some regressors are endogenous, OLS estimation generally results in inconsistent estimators and so the method of instrumental variables (IV) provides a solution to the problem of endogeneity (Wooldridge, 2002; Cameron and Trivedi, 2005). In this context, the problem is to find an observable variable which is correlated with the endogenous explanatory variables but not with the residuals. As in previous exercises (Klein and Crafts, 2009), lagged variables for the market potential is the instrument selected for the estimation. However, in this case, the complexity of obtaining homogeneous market potential estimates for the Spanish provinces before the 1860s restricts the estimation in this study to the years 1893, 1913 and 1929. In addition, IV estimates can be more inconsistent and less efficient than OLS estimators if weak instruments are used. Thus, in order to test whether the instrument applied is valid a weak instrument test is conducted. The test suggested by Stock and Yogo (2005) allows rejecting the null hypothesis of weak instruments²⁹.

Lastly, instead of using IV, equation (1) is estimated using 2-step GMM (Hansen, 1982), a more general case, which is preferred to the former in the presence of heteroskedasticity³⁰. The results of these tests are reported in table 3 together with the 2-step GMM estimation for 1893, 1913 and 1929. The significance and the magnitude of the coefficients associated with the interaction variables are similar to the ones obtained with OLS (table 2), confirming the validity of the previous results³¹.

[HERE TABLE 3]

²⁸ The statistic test follows a chi-squared distribution with degrees of freedom equal to the number of explanatory variables tested for endogeneity. See table 3 for the values obtained.

²⁹ The Stock and Yogo (2005) test is based on the Cragg-Donald F-statistic and the critical values in the tables provided by these authors. See table 3.

³⁰ “The conventional IV estimator (though consistent) is, however, inefficient in the presence of heteroskedasticity. The usual approach today when facing heteroskedasticity of unknown form is to use the Generalized Method of Moments (GMM). [...] The advantages of GMM over IV are clear: if heteroskedasticity is present, the GMM estimator is more efficient than the simple IV estimator. [...] If in fact the error is homoskedastic, IV would be preferable to efficient GMM. For this reason a test for the presence of heteroskedasticity when one or more regressors is endogenous may be useful in deciding whether IV or GMM is called for”. Baum, Schaffer and Stillman (2003), p. 2 and 11. Such a test for heteroskedasticity was previously implemented (see table 2).

³¹ The only change can be found in the interaction capturing labour intensity in 1893, which is now significant at 10%.

Once the econometric issues recently highlighted by Klein and Crafts (2009) in their work on the United States have been addressed, two alternative robustness tests are carried out in order to validate the previous findings, since some statistical problems may still be potentially affecting the results obtained. First, the equation estimated includes both region and industry characteristics. However, it is possible that the variables capturing these characteristics may not be adequately measured for one particular industry or region. To test this potential measurement error, an alternative specification including region (in our case, provinces) and industry fixed effects is estimated (Midelfart-Knarvik *et al.*, 2000). In this case, the four region and the six industry characteristics are replaced by a set of dummy variables and then, the equation is re-estimated. The significance and the coefficients of the interaction variables are virtually unchanged (table 4) and therefore the results previously shown in table 2 are robust when region and industry fixed effects are included.

[HERE TABLE 4]

In addition, an alternative specification is estimated. Spain has traditionally been well endowed with a good amount of mineral resources. Mining activities experienced a boost in the last decades of the 19th century with the political changes and the new legislation that followed the Revolution of 1868 (Nadal, 1975; Tortella, 1994; Chastagnaret, 2000). In addition, the increase in the international demand for some Spanish mining products also contributed to the progress of this sector. Major reserves of iron ore were located in the provinces of Vizcaya and Santander in the north, and in Málaga in the south³²; lead ore mines existed in southern Spain (Murcia, Jaén, Almería, Córdoba, Granada, Badajoz and Ciudad Real); copper was abundant in the south (Huelva)³³; there were mercury mines in Almadén (Ciudad Real), in the south western area of Castile, next to Andalusia; and finally, coal was mainly concentrated in the north (Asturias and León) and in some southern provinces (Ciudad Real and Córdoba)³⁴.

³² The absence of the Basque Country in the sample is an even greater loss when dealing with the impact of mining resources on industry location. In the case of Vizcaya, the availability of non-phosphorous iron ore (at the time when the Bessemer converter was developed to produce steel) led to the emergence of a strong iron and steel industry in this province.

³³ Copper ore (or chalcopyrite) was employed to make components in the new electricity sector. Moreover, copper pyrites in Huelva contained sulphur that was used as an input in chemicals.

³⁴ Spain's coal reserves were small, poor quality and difficult to extract. However, domestic production did benefit from tariffs on coal imports since 1891.

[HERE TABLE 5]

Hence, a new interaction variable (*mineral resources abundance x mineral resources intensity*) is included in the equation³⁵. Mining production by province is obtained from the Spanish Statistical Yearbooks (*Anuarios Estadísticos de España - AEE*) for the years 1860, 1915 and 1931³⁶. The intensity in the use of mineral resources in each of the seven industries considered in the exercise is again calculated from the input-output table of 1958. The results (table 5) are interesting on two respects. First, with the new specification the significance of the interaction variables and the magnitude of the coefficients are not qualitatively altered³⁷. Therefore, the results are robust to this alternative specification. Second, the estimation adds a role for another comparative advantage variable. In 1913, the interaction capturing the abundance of mineral resources is significant and shows a positive sign. Thus, industries that used intensively mineral resources were responsive to the provincial endowment of these resources. It can be hypothesised that the protectionist turn of the Spanish trade policy in the 1890s and its intensification in the following decades could explain this result. As regards coal, the implementation of a new tariff on coal imports could have changed the comparative advantage of the provinces favouring northern territories where coalmines were located³⁸. Nevertheless, in 1929, this effect is no longer present.

To conclude this section, the relative strength of the forces shaping the location of industry in Spain can be examined. So far, the analysis has focused on the significance and the signs of the interaction variables capturing Heckscher-Ohlin and NEG mechanisms. The natural extension of this exercise is to quantify and directly compare the relative importance of these two alternative explanations. To do so, standardised beta coefficients are constructed (table 6). Beta coefficients express the number of standard deviations the dependent variable increases or decreases with a one deviation increase in the independent variable, and therefore, all the parameters are

³⁵ A total of six interactions can be considered, since seven industries are taken into account. The comparative advantage variable replaced, skilled labour, is selected on the basis of the previous results.

³⁶ For 1900, productivity in the mining sector in 1920 has been applied to the active provincial population enrolled in the mining sector in 1900 according to the Census of Population of that year. The choice of 1920 is based on the better quality in the registration of mining activities in that Census of Population. This procedure is based on Geary and Stark (2002).

³⁷ The only changes are found in the level of significance in the interactions for agriculture in 1856 (from 1% to 10%) and economies of scale in 1929 (from 5% to 10%).

³⁸ However, coal is only a part of the total mineral resources in Spain.

expressed in the same unit (standard deviations)³⁹. The higher value of the beta coefficient for an independent variable indicates a higher effect of this variable on the dependent variable in that year. In addition, the comparison of the different years allows quantification of the possible changes in the relative impact over time.

[HERE TABLE 6]

In 1856, comparative advantage drove the spatial distribution of industry across Spain since no significant NEG effects are found. However, in the rest of the years considered, when both HO and NEG mechanisms were at work, the effects captured by the NEG interactions exceeded those of the HO interactions in all cases. Therefore, as market integration progressed, NEG-type mechanisms relating increasing returns and market access became more relevant. The difference in the magnitude of the coefficients for HO and NEG variables was already notable in 1893, when NEG effects doubled the contribution of the HO significant interaction. Then, *scale effects* increased in the turn of the century and decreased in the interwar years. In 1929, the last benchmark in this study, the relative impact of NEG forces was still above (and doubling) that of comparative advantage.

6. Discussion: from comparative advantage to market access.

The results presented so far indicate that both Heckscher-Ohlin and NEG factors were driving the spatial distribution of industry in Spain in the second half of the 19th century and the first decades of the 20th century. In the mid-19th century, when Spain was going through the first stages of the industrialisation process and the domestic market was not fully integrated, industry was spatially dispersed across the country. With the major exception of textiles in Catalonia, the structure of manufacturing production was mainly dominated by foodstuffs⁴⁰, and thus, the spatial distribution of industry was determined by comparative advantage. According to the results, at that

³⁹ Beta coefficients are calculated as follows: $\text{Beta}(i) = [s(x_i)/s(y)] * b(x_i)$, where $b(x_i)$ is the unstandardised coefficient of x_i , $s(x_i)$ is the standard deviation of the independent variable x_i and $s(y)$ is the standard deviation of the dependent variable y (Klein and Crafts, 2009). These coefficients are calculated from the regressions in table 2. This specification has been selected on the basis of the higher goodness of fit and the Schwarz and Akaike Info criterions, which show lower values and therefore favour this selection.

⁴⁰ At an aggregate level, the data provided by Prados de la Escosura (2003) show that in 1856 *foodstuffs* represented 48.7% of the GVA in total Spanish manufacturing.

time, the location of industry was driven by the relative endowment of land, measured through agricultural production.

However, during the second half of the 19th century, when industry concentrated in a very limited number of regions and a process of regional specialisation was taking place, evidence of NEG effects is found. The significance and magnitude of the interaction between market potential and economies of scale shows that *scale effects* were in operation. The domestic market gradually became integrated, industrialisation progressed at the same time as technological advances were progressively incorporated, and large scale production achieved higher development within the industrial sector. In this regard, the theory predicts that the forces pulling increasing returns industries into central locations are strongest at ‘intermediate’ levels of transport costs. Therefore, the expansion of the railway network in the second half of the 19th century and the subsequent fall in transport costs played a key role in the process in a country where geographical conditions had traditionally imposed heavy costs on communications. Thus, it can be argued that, as stated by NEG models, the interaction of increasing returns, transport costs and the size of market may have favoured the emergence of agglomeration forces, explaining the remarkable geographical concentration of industry observed in Spain in that period.

In addition, at the end of the century, factor endowments were still relevant, but in this case, through the interaction considering labour abundance. Then, labour intensive industries were attracted to regions where labour was relatively abundant. Throughout the second half of the 19th century, Spanish manufacturing was mostly oriented towards the production of consumption goods, which are generally labour intensive products⁴¹. As a result, such specialisation would have exerted a pull on industries to locate in labour abundant provinces, in line with the traditional factor endowments explanation. From the late 1800s until World War I, there was an intensification in the pattern where primarily *scale effects* and labour abundance were the forces shaping the location of industry in Spain.

In the interwar years, some changes are observed. The relative endowment of labour was no longer significant. In that period there was an expansion of industries linked to the Second Industrial Revolution and the production of capital goods experienced notable development in Spain (Betrán, 1999). In turn, in 1929, traditional

⁴¹ The production of capital goods (iron and metal industry) developed strongly in the Basque Country, but as already mentioned, this region is absent from the sample due to statistical restrictions.

factor endowments once again affected the location of industry through agriculture abundance. The significance of the interaction between agricultural abundance and the intensity in the use of agricultural inputs could be related to the changes that took place in the agricultural sector in the first decades of the 20th century. Agricultural specialisation has traditionally differed across Spanish regions due to (among other things) the differences in climate and the quality of land (Jiménez Blanco, 1986). Throughout the 19th century, the production of cereals, a staple food, was spread across the country, especially in the inland provinces, whereas the Mediterranean regions specialised in vegetables, fruits and vineyards. However, agricultural specialisation deepened between 1900 and 1930 (Tirado *et al.*, 2006) in a context where the crop of cereals was declining and the Mediterranean regions gradually increased their production⁴².

Besides, the strength of the *scale effects* in the interwar years declined slightly in relation to the previous period. This could be linked to the further reduction in transport costs, which may be close to leave the intermediate levels category suggested in the NEG theoretical models. However, the interaction between market potential and size of establishment was still the major force driving the location of industry in Spain and explaining the high degree of spatial concentration prior to the Spanish Civil War. Nevertheless, although the results show that the reduction in transport costs experienced in Spain in the 19th century encouraged industries with economies of scale to move to locations with high market potential, transport costs were not low enough to exert a pull of centrality on industries with *linkage effects*⁴³.

In the analysis of the processes of industrialisation in historical perspective the literature has suggested a potential link between industrial development and the endowment of skilled workers. Yet, the availability of educated population was not a significant determinant in Spain in the period under study, and thus, industry did not tend to concentrate in the provinces where literacy rates were relatively higher. Therefore, education, measured through the literacy rate, was not an important factor during the first stages of the industrialisation process. As described above, the geographical pattern of literacy in Spain shows that it was higher in provinces where

⁴² At that time, the regions that achieved a deeper specialisation in their agricultural sector experienced a higher increase in productivity. Simpson (1995).

⁴³ This result is similar to the one obtained by Crafts and Mulatu (2005, 2006) for the British case. These authors suggested that “...the apparent unimportance of market-potential interactions involving linkage effects may mean that it was not until the motor-transport era that these became relevant for location decisions”. Crafts and Mulatu (2006), p. 600.

manufacturing activities were not predominant, as has been noted in the literature (Núñez, 1992)⁴⁴. Likewise, industrial activities were not attracted to the provinces where mineral resources were abundant. These results can shed some light on a more general debate. Sachs and Warner (1995, 2001) argued that developing countries in the second half of the 20th century did not benefit from the abundance of natural resources (the so-called curse of natural resources). However, a positive relationship between industrialisation and natural resources has traditionally been emphasised in the studies of the 19th and early 20th centuries. Within our analytical framework, Crafts and Mulatu (2005, 2006) confirmed the importance of factor endowments in the location of industry in Victorian Britain, and in particular, for coal abundance, showing that regions endowed with coal mines attracted industries that made intensive use of steam power. The US experience is, nonetheless, not so clear⁴⁵: Klein and Crafts (2009) did not find evidence of this relationship in the period 1880-1920⁴⁶. As regards Spain, the benefits of being endowed with mineral resources have been questioned, in view of the low linkage effects that mining activities produced on the industrial sector, with the exception of iron ore in Vizcaya⁴⁷. In this regard, the results show that over the period considered, with the sole exception of 1913, industries that made intensive use of mineral resources did not tend to locate in mineral resource abundant provinces. The positive relationship between the endowment of mineral resources and industry location in 1913 might be reflecting the change in comparative advantage that the protection on coal imports after 1891 conferred to the provinces where this mineral was produced.

At this point, the new evidence can be compared with previous studies that have focused on the determinants of industrial location in Spain. Rosés (2003) found that both comparative advantage and NEG factors were important already in 1861. Following Davis and Weinstein (1999, 2003), this author provided evidence of a ‘home

⁴⁴ For the US experience, “... the work of Goldin and Katz (1998) suggests that it is not surprising that the educated population-white collar workers interaction is insignificant. They convincingly argue that in this ‘factory-production’ phase of manufacturing, physical capital was a substitute for skill and technological advance was downgrading the role of skilled labour”. Klein and Crafts (2009), p. 26. An alternative, less satisfactory explanation, is that the literacy rate is not a good proxy for human capital in this period, and that it might be more appropriate to consider technical education. For the Spanish case, see Lozano (2007).

⁴⁵ In fact, returning to the British case, when the variable ‘steam power use’ was replaced by ‘coal use’, the interaction became statistically insignificant. Crafts and Mulatu (2005).

⁴⁶ “Both coal and skilled-labor interactions change signs and are insignificant for most of the time”. Klein and Crafts (2009), p. 20.

⁴⁷ The debate in the Spanish historiography about the positive or negative effects of mining activities on the economic performance can be followed in Escudero (1996). For a recent empirical exercise, see Domenech (2008).

market effect' and concluded that increasing returns, which were relevant in the modern manufacturing industries, together with the endowment of artisans and capital, could explain the diverse fortunes of particular regions in the Spanish industrialisation process. However, in the exercise carried out in this paper, the analysis of the interactions suggests that NEG forces were not strong enough in 1856 to exert a pull on industrial location⁴⁸.

In turn, Tirado *et al.*, (2002) studied the second half of the 19th century in an exercise based on the proposal of Kim (1995). These authors also concluded that market size was significant in 1856, stressing the role of economies of scale. Then, they showed that in 1893 the relevance of market size had increased and access to markets and economies of scale became the main variables explaining the relative industrial intensity of provinces as the Spanish market gradually integrated⁴⁹. Therefore, a growing role for increasing returns as a driving force of industrial location was suggested to be behind of the geographical concentration of industry in Spain. This picture has been confirmed and extended in this paper until the 1930s through the interaction between market potential and economies of scale⁵⁰.

The historical experience of Spain reinforces the conclusions reached for other countries that both Heckscher-Ohlin and NEG forces interact simultaneously and are ultimately responsible for the spatial distribution of industry. Several studies have applied the Midelfart-Knarvik *et al.*, (2002) model in historical perspective, and therefore, the Spanish case can be incorporated in the international discussion and can help to complete a more global picture of the determinants of industrial location.

However, the results do not point to the existence of a unique pattern in the location of industry, since the relative strength of these explanations and the significance of the variables considered differ in these studies. When compared to other experiences, some aspects can be highlighted. First, Crafts and Mulatu (2005, 2006) concluded that traditional factor endowments (agriculture, coal abundance and skilled labour) were the most important elements for explaining the location of industry in

⁴⁸ However, Rosés (2003) applied an alternative empirical strategy and a different regional aggregation, where historical Spanish regions were considered. Nevertheless, in this paper it is suggested that the lack of complete integration of the domestic market at mid-1800s might be the reason for the absence of NEG-type mechanisms operating in Spanish manufacturing.

⁴⁹ For the comparative advantage variable considered, human capital endowment, these authors found that literacy rate had a positive impact on regional industrial intensity in 1893, a result which is at odds with the current exercise.

⁵⁰ As regards the work of Betrán (1999) for the interwar years, the new evidence cannot complete her results since the scope of the Marshallian and Jacobs externalities she analysed was internal to provinces.

Victorian Britain, although NEG effects, measured by the interaction between market potential and economies of scale, also played a role. The results for Spain differ in two aspects: first, skilled labour did not seem to be relevant in location decisions; and second, NEG *scale effects* had a stronger impact on industrial location in Spain than comparative advantage. This could be linked to the fact that changes in the regional specialisation and the concentration of manufacturing in Spain (1856-1929) were much more profound than in Britain (1871-1931) where the industrialisation process was at a more mature stage. Nevertheless, the evolution of NEG-type mechanisms was similar in the two economies: *scale effects*, although significant, decreased in the interwar period, and no evidence of *linkage effects* is found before the 1930s when the analyses conclude.

On the other hand, Wolf (2007) stressed the relevance of skilled labour abundance, innovative activities and *forward linkages* as drivers of industrial location in interwar Poland (1926-1934). His results showed that the forces in operation after the reunification in 1918 when the domestic market was being created were, to a large extent, similar to the mechanisms present in the manufacturing sector in the EU in recent times (Midelfart-Knarvik *et al.*, 2002). However, although Poland and Spain were two similar-size economies in the periphery of Europe, there is no evidence that these forces were important drivers of industrial location in Spain at that time⁵¹.

Finally, the comparison with the US can also be illustrative. In their attempt to explain the emergence of the manufacturing belt between 1880 and 1920, Klein and Crafts (2009) emphasised some elements which resemble the Spanish experience. First, in two countries characterised by a remarkable increase in the concentration of manufacturing activities, most of the explanation falls on NEG-type mechanisms. As regards comparative advantage variables, agriculture was important, at least until 1900. Moreover, they did not find a role for coal abundance or skilled labour, and when this latter variable was significant (in 1900) it showed a negative sign. For NEG variables, the interaction capturing *scale effects* is significant and the main driver for a good part of the years considered, although it decreases over time. The main difference when compared to Spain, however, is that in the case of the US, *linkage effects* eventually became the main determinants of industrial location and in 1920 their impact exceeded that of *scale effects*.

⁵¹ Innovation activities have not been considered in the Spanish case.

7. Concluding remarks

The integration process in the European Union has raised concerns about the effects of this process on the spatial distribution of economic activities. Looking back to the creation of the national economies from a historical perspective can shed light on the impact that economic integration may generate on the geographical changes in manufacturing activities. The Spanish experience in the second half of the 19th and early 20th centuries constitutes an interesting study case. At that time, the integration of the domestic market was completed while the Spanish economy was going through the first stages of industrialisation. The liberal reforms implemented by successive governments, the integration of the goods and factor markets and the fall in transport costs that followed the expansion of the railway network, came together with a remarkable increase in the concentration of manufacturing activities in a limited number of provinces and with an increase in regional specialisation.

In order to study the determinants of these notable changes in industrial location in Spain during this period, the standard model developed by Midelfart-Knarvik *et al.*, (2002) is applied. This approach has proved to be very useful to explore the determinants of industrial location in different historical cases like interwar Poland (Wolf, 2007), Victorian Britain (Crafts and Mulatu, 2005, 2006) and the US between 1880 and 1920 (Klein and Crafts, 2009). In the latter exercise, the authors have, in addition, improved the robustness of the empirical results by addressing a good number of econometric issues that might be affecting the estimation of the standard equation derived from the model proposed by Midelfart-knarvik *et al.*, (2002). Therefore, the analysis of the Spanish case takes into account the methodological improvements suggested by Klein and Crafts (2009). Besides, one of the interesting features of this model is its capacity to combine both Heckscher-Ohlin factor endowments theory and NEG-type mechanisms in order to explore the forces behind the spatial evolution of manufacturing activities over time. The relative relevance of these forces is captured by a series of interactions regarding region and industry characteristics.

In this paper, the analysis has focused on Spanish NUTS3 provinces, considering seven industrial sectors in four years: 1856, 1893, 1913 and 1929. The main findings of the exercise can be summarized as follows. First, market potential was, as suggested by NEG models, a relevant variable to explain the location of industry in Spain in a period of falling transport costs and a strong increase in the geographical concentration of

manufacturing. Second, there was a role for both comparative advantage and economic geography in driving the changes in the industrial map of Spain. The evidence suggests that both Heckscher-Ohlin and NEG forces were simultaneously at work, although the relative strength of these two explanations changed over time. In 1856, before the integration of the Spanish market was completed, comparative advantage (agriculture abundance) explained the spatial distribution of industry in Spain, which at that time showed a high level of dispersion. As the domestic market became integrated, and industrialisation progressed, the impact of NEG forces increased and in 1893 NEG *scale effects* were already the main determinants of industrial location in Spain, although a role for labour abundance is also found. The influence of increasing returns rose steadily up to the eve of the World War I and then weakened in the interwar years although it remained the main force driving industrial location. The presence of this agglomeration force (Krugman, 1991) could, therefore, explain the notable spatial concentration of industrial activity in Spain between the second half of the 19th century and the outbreak of the Spanish Civil War.

References

- Arellano, M., 1987. Computing robust standard errors for within-group estimators. *Oxford Bulletin of Economics and Statistics* 49, 431-434.
- Barquín, R., 1997. Transporte y precio del trigo en el siglo XIX: Creación y reordenación de un mercado nacional. *Revista de Historia Económica* 15, 17-48.
- Barquín, R., 1999. El ahorro social: una perspectiva dinámica. In: Muñoz, M., Sanz, J., Vidal, J. (Eds.), *Siglo y medio del ferrocarril en España, 1848-1998*. Fundación de los Ferrocarriles Españoles, Madrid, 337-354.
- Baum, C. F., Schaffer, M. E., Stillman, S., 2003. Instrumental variables and GMM. *Stata Journal* 3, 1-31.
- Betrán, C., 1995. Industria y crecimiento económico en el primer tercio del siglo XX. España, 1913-1929. PhD dissertation (unpublished), Universitat de València.
- Betrán, C., 1999. Difusión y localización industrial en España durante el primer tercio del siglo XX. *Revista de Historia Económica* 17, 663-696.
- Boyer, G. R., Hatton, T. J., 1994. Regional labour market integration in England and Wales, 1850-1913. In: Grantham, G., MacKinnon, M. (Eds.), *Labour market evolution. The economic history of market integration, wage flexibility and the employment relation*. Routledge, London, 84-106.
- Calvo, A., 2001. Los inicios de las telecomunicaciones en España: El telégrafo. *Revista de Historia Económica* 19, 613-636.
- Cameron, A. C., Trivedi, P. K., 2005. *Microeconometrics: Methods and applications*. Cambridge University Press, New York.

- Carreras, A., 1983. La producció industrial espanyola i italiana des de mitjan segle XIX fins a l'actualitat. PhD dissertation (unpublished), Universitat Autònoma de Barcelona.
- Carreras, A., 1984. La producción industrial española, 1842-1981: Construcción de un índice anual. *Revista de Historia Económica* 1, 127-157.
- Carreras, A., Tafunell, X., 2003. Historia económica de la España contemporánea. Crítica, Barcelona.
- Castañeda, L., Tafunell, X., 1993. Un nuevo indicador para la historia financiera española: La cotización de las letras de cambio a corto plazo. *Revista de Historia Económica* 11, 367-383.
- Chastagnaret, G., 2000. L'Espagne, puissance minière dans l'Europe du XIXe siècle. Casa de Velázquez, Madrid.
- Combes, P.-P., Lafourcade, M., Thisse, J., Toutain, J., 2008. The rise and fall of spatial inequalities in France: A long-run perspective. Working Paper 2008-54, Paris School of Economics.
- Combes, P.-P., Mayer, T., Thisse, J., 2008. Economic geography. Princeton University Press, Princeton (NJ).
- Crafts, N., 2005. Market potential in British regions, 1871-1931. *Regional Studies* 39, 1159-1166.
- Crafts, N., Mulatu, A., 2005. What explains the location of industry in Britain, 1871-1931? *Journal of Economic Geography* 5, 499-518.
- Crafts, N., Mulatu, A., 2006. How did the location of industry respond to falling transport costs in Britain before World War I? *The Journal of Economic History* 66, 575-607.
- Davis, D. R., Weinstein, D. E., 1999. Economic geography and regional production structure: An empirical investigation. *European Economic Review* 43, 379-407.
- Davis, D. R., Weinstein, D. E., 2003. Market access, economic geography and comparative advantage: An empirical test. *Journal of International Economics* 59, 1-23.
- Dirección General de Obras Públicas, 1861. Memoria sobre el progreso de las obras públicas en España, en los años 1859 y 1860, Madrid.
- Domenech, J., 2008. Mineral resource abundance and regional growth in Spain, 1860-2000. *Journal of International Development* 20, 1122-1135.
- Erdozain, P., Mikelarena, F., 1996. Algunas consideraciones acerca de la evolución de la población rural en España en el siglo XIX. *Noticiario de Historia Agraria* 12, 91-118.
- Escudero, A., 1996. Pesimistas y optimistas ante el "boom" minero. *Revista de Historia Industrial* 10, 69-91.
- Estevadeordal, A., Frantz, B., Taylor, A. M., 2003. The rise and fall of world trade, 1870-1939. *The Quarterly Journal of Economics* 118, 359-407.
- Frax, E., 1981. Puertos y comercio de cabotaje en España, 1857-1934, Banco de España, Madrid.
- Frax, E., Madrazo, S., 2001. El transporte por carretera, siglos XVIII-XX. *Transportes, Servicios y Telecomunicaciones* 1, 31-53.
- Geary, F., Stark, T., 2002. Examining Ireland's post-famine economic growth performance. *The Economic Journal* 112, 919-935.
- Goldin, C., Katz, L. F., 1998. The origins of technology-skill complementarity. *The Quarterly Journal of Economics*, 693-732.
- Gómez Mendoza, A., 1981. Railways and Spanish economic growth in the late 19th century. PhD dissertation (unpublished).

- Gómez Mendoza, A., 1982. Ferrocarriles y cambio económico en España 1855-1913. Alianza, Madrid.
- Gómez Mendoza, A., San Román, E., 2005. Transportes y comunicaciones. In: Carreras, A., Tafunell, X. (Eds.), Estadísticas históricas de España siglos XIX-XX, vol. II. Fundación BBVA, Bilbao, 509-572.
- Hansen, L., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1054
- Harris, C. D., 1954. The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers* 44, 315-348.
- Hatton, T. J., Williamson, J. G., 1998. The age of mass migration: Causes and economic impact. Oxford University Press, New York.
- Helpman, E., Krugman, P., 1985. Market structure and foreign trade. The MIT Press, Cambridge (MA).
- Herranz, A., 2004. La dotación de infraestructuras en España, 1844-1935. *Estudios de Historia Económica*, Banco de España.
- Herranz, A., 2005. La reducción de los costes de transporte en España (1800-1936). *Cuadernos Económicos ICE* 70, 183-203.
- Herranz, A., 2007. Infrastructure investment and Spanish economic growth, 1850–1935. *Explorations in Economic History* 44, 452-468.
- Jiménez Blanco, J. I., 1986. Introducción. In: Garrabou, R., Barciela, C., Jiménez Blanco, J. I. (Eds.), *Historia agraria de la España contemporánea*, vol.3. Crítica, Barcelona, 9-141.
- Keeble, D., Owens, P., Thompson, C., 1982. Regional accessibility and economic potential in the European community. *Regional Studies* 16, 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. *The Quarterly Journal of Economics* 110, 881-908.
- Kim, S., 1999. Regions, resources, and economic geography: Sources of U.S. regional comparative advantage, 1880–1987. *Regional Science and Urban Economics* 29, 1-32.
- Klein, A., Crafts, N., 2009. Making sense of the manufacturing belt: Determinants of U.S. industrial location, 1880-1920. University of Warwick.
- Krugman, P., 1991. Increasing returns and economic geography. *Journal of Political Economy* 99, 483-499.
- Krugman, P., Venables, A. J., 1995. Globalization and the inequality of nations. *The Quarterly Journal of Economics* 110, 857-880.
- Llopis, E., Sotoca, S., 2005. Antes, bastante antes: La primera fase de la integración del mercado español de trigo, 1725-1808. *Historia Agraria* 36, 225-262.
- Lozano, C., 2007. Ideología, política y realidad económica en la formación profesional industrial española (1857-1936). Milenio, Lleida.
- Madrazo, S., 1984. El sistema de transportes en España, 1750-1850. Turner, Madrid.
- Maixé-Altés, J. C., Iglesias, E. M., 2009. Domestic monetary transfers and the inland bill of exchange markets in Spain (1775–1885). *Journal of International Money and Finance* 28, 496-521.
- Martínez Vara, T., 1999. La integración del mercado del trigo en el norte y en la castilla del Duero, 1800-1860. Algunas reflexiones. *Historia Agraria* 19, 43-73.
- Martínez-Galarraga, J., 2007. New estimates of regional GDP in Spain, 1860-1930. Working Paper E07/177, Universitat de Barcelona.
- Martínez-Galarraga, J., 2009. El potencial de mercado de las provincias españolas, 1867-1930. Universitat de Barcelona. *Mimeo*.

- Martorell, M., 2001. Historia de la peseta: La España contemporánea a través de su moneda. Planeta, Barcelona.
- Matilla, M., Pérez, P., Sanz, B., 2008. La integración del mercado español a finales del siglo XIX: Los precios del trigo entre 1891 y 1905. FUNCAS Working Paper 391/2008.
- Midelfart-Knarvik, K. H., Overman, H., Redding, S., Venables, A. J., 2002. The location of European industry. In: European Commission (Ed.), European economy. European integration and the functioning of product markets. Directorate-General for economic and financial affairs, Brussels, 213-269.
- Midelfart-Knarvik, K. H., Overman, H., Venables, A. J., 2000. Comparative advantage and economic geography: Estimating the location of production in the EU. CEPR Discussion Paper, 2618.
- Ministerio de Obras Públicas, 1902. Estadísticas de obras públicas, años 1899 y 1900. Dirección General de Obras Públicas, Madrid.
- Mitchell, B. R., 1998a. International historical statistics. Europe 1750-1993. Macmillan, London.
- Mitchell, B. R., 1998b. International historical statistics. The Americas 1750-1993. Macmillan, London.
- Mohammed, S. I. S., Williamson, J. G., 2004. Freight rates and productivity gains in British tramp shipping 1869–1950. Explorations in Economic History 41, 172-203.
- Nadal, J., 1975. El fracaso de la revolución industrial en España, 1814-1913. Ariel, Barcelona.
- Nadal, J., 1987. La industria fabril española en 1900. Una aproximación. In: Nadal, J., Carreras, A., Sudrià, C. (Eds.), La economía española en el siglo XX. Una perspectiva histórica, Ariel, Barcelona, 23-61.
- Nadal, J., Tafunell, X., 1992. Sant Martí de Provençals, pulmó industrial de Barcelona (1847-1992). Columna, Barcelona.
- Núñez, C. E., 1992. La fuente de la riqueza: Educación y desarrollo económico en la España contemporánea. Alianza, Madrid.
- O'Rourke, K. H., 2000. Tariffs and growth in the late 19th century. The Economic Journal 110, 456-483.
- O'Rourke, K., Williamson, J. G., 1999. Globalization and history: The evolution of a nineteenth-century Atlantic economy. The MIT Press, Cambridge (MA).
- Paluzie, E., Pons, J., Tirado, D. A., 2004. The geographical concentration of industry across Spanish regions, 1856-1995. Jahrbuch für Regionalwissenschaft 24, 143-160.
- Parejo, A., 2001. Industrialización, desindustrialización y nueva industrialización de las regiones españolas (1950-2000). Un enfoque desde la historia económica. Revista de Historia Industrial 19-20, 15-75.
- Pascual, P., 1990. Agricultura i industrialització a la Catalunya del segle XIX. Formació i desestructuració d'un sistema econòmic. Crítica, Barcelona.
- Peña, D., Sánchez-Albornoz, N., 1984. Wheat prices in Spain, 1857-1890: An application of the Box-Jenkins methodology. Journal of European Economic History 2, 353-373.
- Pepper, J. V., 2002. Robust inferences from random clustered samples: An application using data from the panel study of income dynamics. Economic Letters 75, 341-345.
- Prados de la Escosura, L., 2000. International comparisons of real product, 1820–1990: An alternative data set. Explorations in Economic History 37, 1-41.

- Prados de la Escosura, L., 2003. El progreso económico de España 1850-2000. Fundación BBVA, Bilbao.
- Puga, D., 1999. The rise and fall of regional inequalities. *European Economic Review* 43, 303-334.
- Redding, S., Venables, A. J., 2004. Economic geography and international inequality. *Journal of International Economics* 62, 53-82.
- Reher, D., 2001. Producción, precios e integración de los mercados regionales de grano en la España preindustrial. *Revista de Historia Económica* 19, 539-572.
- Ringrose, D. R., 1970. Transportation and economic stagnation in Spain: 1750- 1850. Duke University Press, Durham (NC).
- Rosés, J. R., 2003. Why isn't the whole of Spain industrialized? New economic geography and early industrialization, 1797-1910. *The Journal of Economic History* 63, 995-1022.
- Rosés, J. R., Sánchez-Alonso, B., 2004. Regional wage convergence in Spain 1850–1930. *Explorations in Economic History* 41, 404-425.
- Sachs, J. D., Warner, A. M., 1995. Natural resource abundance and economic growth. NBER Working Paper, 5398.
- Sachs, J. D., Warner, A. M., 2001. The curse of natural resources. *European Economic Review* 45, 827-838.
- Sánchez-Albornoz, N., 1975. Los precios agrícolas durante la segunda mitad del siglo XIX. Banco de España, Madrid.
- Sánchez-Alonso, B., 2000. Those who left and those who stayed behind: Explaining emigration from the regions of Spain, 1880-1914. *Journal of Economic History* 60, 730-755.
- Sardà, J., 1948. La política monetaria y las fluctuaciones de la economía española en el siglo XIX. CSIC, Madrid.
- Schulze, M., 2007. Regional income dispersion and market potential in the late nineteenth century Hapsburg Empire. LSE Working Paper, 106/07.
- Silvestre, J., 2005. Internal migrations in Spain, 1877-1930. *European Review of Economic History* 9, 233-265.
- Silvestre, J., 2007. Temporary internal migrations in Spain, 1860-1930. *Social Science History* 31, 539-574.
- Simpson, J., 1995. Spanish agriculture: The long siesta, 1765-1965. Cambridge University Press, Cambridge (MA).
- Stock, J. H., Yogo, M., 2005. Testing for weak instruments in linear IV regression. In: Andrews, D., Stock, J. H. (Eds.), *Identification and inference for econometric models. Essays in honour of Thomas Rothenberg*, Cambridge University Press, New York, 80-108.
- Sudrià, C., 1994. Los bancos de emisión provinciales en la España del siglo XIX. In: Tedde, P., Marichal, C. (Eds.), *La formación de los bancos centrales en España y América Latina (siglos XIX y XX)*, Banco de España, Madrid, 81-107.
- Tedde, P., 1994. Cambio institucional y cambio económico en la España del siglo XIX. *Revista de Historia Económica* 3, 525-538.
- Tirado, D. A., Paluzie, E., Pons, J., 2002. Economic integration and industrial location: The case of Spain before World War I. *Journal of Economic Geography* 2, 343-363.
- Tirado, D. A., Pons, J., Paluzie, E., 2006. Los cambios en la localización de la actividad industrial en España, 1850-1936. Un análisis desde la nueva geografía económica. *Revista de Historia Industrial* 31, 41-63.

- Tortella, G., 1973. Los orígenes del capitalismo en España. Banca, industria y ferrocarriles en el siglo XIX. Tecnos, Madrid.
- Tortella, G., 1994. El desarrollo de la España contemporánea. Alianza, Madrid.
- Venables, A. J., 1996. Equilibrium locations of vertically linked industries. *International Economic Review* 37, 341-359.
- Wais, F., 1987. Historia de los ferrocarriles españoles. Fundación de los Ferrocarriles Españoles, Madrid.
- Wolf, N., 2007. Endowments vs. market potential: What explains the relocation of industry after the Polish reunification in 1918? *Explorations in Economic History* 44, 22-42.
- Wooldridge, J. M., 2002. Econometric analysis of cross section and panel data. The MIT Press, Cambridge (MA).

Appendix

Figure 1. Spanish provinces.

Figure 2. Market potential, Spanish provinces (Barcelona=100).

Source: Martinez-Galarraga (2009). Note: The maps are expressed in relative terms, with Barcelona, the highest market potential province, being equal to 100.

Table 1. Region and industry characteristics.

<i>Region characteristics (provinces)</i>	<i>Industry characteristics</i>
Agricultural production, % GDP	Agricultural input, % total costs
Total active population per land	Labour input, % of Gross Value Added
Educated population	Skilled workers intensity
Market potential	Size of establishment
	Intermediate goods, % of output
	Sales to industry, % of output

Table 2. Regression results. Dependent variable $\ln(s_{it}^k)$

	Pooled 1856-1929		1856		1893	
	a	b	a	b	a	b
<i>Constant</i>	82.155 (53.954)	82.155 (75.518)	42.866 (106.783)	42.866 (134.033)	-267.011 (183.823)	-267.011 (230.675)
<i>Population</i>	0.192 (0.126)	0.192 (0.139)	0.069 (0.343)	0.069 (0.272)	-0.019 (0.219)	-0.019 (0.189)
<i>Manufacturing</i>	-0.106 (0.092)	-0.106 (0.088)	-0.109 (0.231)	-0.109 (0.173)	-0.061 (0.164)	-0.061 (0.139)
<i>Share of agricultural GVA</i>	-0.099 (0.112)	-0.099 (0.121)	-0.829** (0.396)	-0.829*** (0.265)	-0.180 (0.228)	-0.180 (0.186)
<i>Labour abundance</i>	-0.604 (0.537)	-0.604 (0.762)	-0.149 (1.307)	-0.149 (1.360)	-2.139** (1.010)	-2.139** (1.007)
<i>% Educated population</i>	0.755** (0.374)	0.755* (0.389)	2.246** (0.888)	2.246** (1.012)	-0.196 (0.786)	-0.196 (0.870)
<i>Market potential</i>	-1.016 (8.836)	-1.016 (13.081)	-2.130 (20.405)	-2.130 (25.881)	14.508 (30.311)	14.508 (38.407)
<i>Agricultural input use</i>	-0.741*** (0.107)	-0.741*** (0.148)	-2.094*** (0.322)	-2.094*** (0.320)	-1.155*** (0.217)	-1.155*** (0.267)
<i>Share of labour in GVA</i>	-6.423*** (0.943)	-6.423*** (1.153)	-15.182*** (2.633)	-15.182*** (2.807)	-8.181*** (1.516)	-8.181*** (1.491)
<i>% White-collar workers</i>	-2.814*** (0.719)	-2.814*** (0.997)	-6.508*** (1.807)	-6.508*** (1.943)	-2.070 (1.460)	-2.070 (1.608)
<i>Size of establishment</i>	-1.734*** (0.284)	-1.734*** (0.269)	-1.256 (1.021)	-1.256 (0.881)	-6.733*** (1.201)	-6.733*** (1.233)
<i>Intermediate input use</i>	-4.855 (8.827)	-4.855 (12.322)	8.953 (17.354)	8.953 (21.283)	56.059* (30.585)	56.059 (38.510)
<i>Sales to industry</i>	-6.005 (4.649)	-6.005 (6.442)	0.918 (9.250)	0.918 (11.643)	27.210* (15.931)	27.210 (19.604)
<i>Share of agricultural GVA</i>	0.116*** (0.023)	0.116*** (0.285)	0.212*** (0.076)	0.212*** (0.066)	0.066 (0.047)	0.066 (0.056)
<i>* Agricultural input use</i>						
<i>Labour abundance</i>	0.144 (0.146)	0.144 (0.210)	0.046 (0.353)	0.046 (0.370)	0.572** (0.277)	0.572** (0.279)
<i>* Share of labour in GVA</i>						
<i>Educated population</i>	-0.178 (0.152)	-0.178 (0.158)	-0.935** (0.373)	-0.935** (0.417)	0.117 (0.317)	0.117 (0.360)
<i>* White-collar workers</i>						
<i>Market potential</i>	0.223*** (0.047)	0.223*** (0.048)	-0.051 (0.185)	-0.051 (0.156)	0.602*** (0.183)	0.602*** (0.187)
<i>* Size of establishment</i>						
<i>Market potential</i>	-0.099 (1.455)	-0.099 (2.155)	0.385 (3.339)	0.385 (4.169)	-2.643 (5.023)	-2.643 (6.392)
<i>* Intermediate input use</i>						
<i>Market potential</i>	0.164 (0.760)	0.164 (1.116)	0.183 (1.768)	0.183 (2.276)	-1.617 (2.606)	-1.617 (3.244)
<i>* Sales to industry</i>						
<i>Number of observations</i>	1157	1157	261	261	295	295
<i>Adjusted R²</i>	0.51	0.51	0.66	0.66	0.63	0.63
<i>BPG test chi2(18)</i>	124.225***		52.662***		33.576**	
<i>F-test joint significance</i>	230.92***	241.25***	93.06***	160.54***	103.72***	83.93***

Note: (a) White heteroskedasticity-robust standard error in brackets; (b) Cluster-robust standard error.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

BPG: Breusch-Pagan/Godfrey heteroskedasticity test.

Table 2 (continued)

	1913		1929	
	a	b	a	b
<i>Constant</i>	-171.518 (190.523)	-171.518 (226.947)	90.443 (164.906)	90.443 (182.720)
<i>Population</i>	0.132 (0.254)	0.132 (0.204)	0.052 (0.248)	0.052 (0.156)
<i>Manufacturing</i>	0.013 (0.201)	0.013 (0.160)	-0.052 (0.180)	-0.052 (0.958)
<i>Share of agricultural GVA</i>	0.344 (0.273)	0.344 (0.267)	-0.008 (0.133)	-0.008 (0.905)
<i>Labour abundance</i>	-3.455*** (0.940)	-3.455*** (0.994)	-0.298 (0.924)	-0.298 (0.112)
<i>% Educated population</i>	1.499** (0.717)	1.499* (0.886)	1.737 (1.142)	1.737 (1.458)
<i>Market potential</i>	31.460 (29.330)	31.460 (35.890)	-6.642 (23.129)	-6.642 (26.664)
<i>Agricultural input use</i>	-0.525** (0.251)	-0.525** (0.253)	-0.162 (0.160)	-0.162 (0.162)
<i>Share of labour in GVA</i>	-7.489*** (1.321)	-7.489*** (1.111)	2.397** (1.211)	2.397** (1.107)
<i>% White-collar workers</i>	-0.548 (1.319)	-0.548 (1.666)	4.221** (2.065)	4.221 (2.541)
<i>Size of establishment</i>	-6.060*** (1.103)	-6.060*** (1.084)	-2.932*** (0.958)	-2.932*** (0.893)
<i>Intermediate input use</i>	37.315 (31.809)	37.315 (37.628)	-13.141 (27.363)	-13.141 (30.082)
<i>Sales to industry</i>	19.238 (16.254)	19.238 (19.396)	-10.716 (13.993)	-10.716 (15.689)
<i>Share of agricultural GVA</i> <i>* Agricultural input use</i>	0.063 (0.062)	0.063 (0.062)	0.106*** (0.032)	0.106*** (0.030)
<i>Labour abundance</i> <i>* Share of labour in GVA</i>	0.960*** (0.257)	0.960*** (0.272)	0.118 (0.252)	0.118 (0.300)
<i>Educated population</i> <i>* White-collar workers</i>	-0.505* (0.289)	-0.505 (0.359)	-0.504 (0.467)	-0.504 (0.569)
<i>Market potential</i> <i>* Size of establishment</i>	0.826*** (0.157)	0.826*** (0.153)	0.299** (0.123)	0.299** (0.118)
<i>Market potential</i> <i>* Intermediate input use</i>	-5.558 (4.864)	-5.558 (5.927)	0.698 (3.825)	0.698 (4.380)
<i>Market potential</i> <i>* Sales to industry</i>	-3.179 (2.484)	-3.179 (3.064)	0.603 (1.952)	0.603 (2.278)
<i>Number of observations</i>	300	300	301	301
<i>Adjusted R²</i>	0.56	0.56	0.56	0.56
<i>BPG test chi2(18)</i>	55.187***		45.400***	
<i>F-test joint significance</i>	89.03***	91.43***	51.88***	53.71***

Note: (a) White heteroskedasticity-robust standard error in brackets; (b) Cluster-robust standard error.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

BPG: Breusch-Pagan/Godfrey heteroskedasticity test.

Table 3. Estimation results: 2-step GMM/IV. Dependent variable $\ln(s_{it}^k)$

	1893	1913	1929
<i>Constant</i>	-355.315* (-1.823)	-139.919 (-0.724)	120.155 (0.721)
<i>Population</i>	-0.002 (-0.008)	0.145 (0.569)	0.032 (0.132)
<i>Manufacturing</i>	-0.077 (-0.464)	0.002 (0.008)	-0.038 (-0.215)
<i>Share of agricultural GVA</i>	-0.179 (0.787)	0.339 (1.237)	-0.007 (-0.052)
<i>Labour abundance</i>	-2.097** (-2.024)	-3.424*** (-3.664)	-0.251 (-0.270)
<i>% Educated population</i>	-0.214 (-0.273)	1.497** (2.085)	1.737 (1.518)
<i>Market potential</i>	28.453 (0.896)	26.683 (0.903)	-10.733 (-0.457)
<i>Agricultural input use</i>	-1.133*** (-5.248)	-0.529** (-2.089)	-0.170 (-1.058)
<i>Share of labour in GVA</i>	-8.126*** (-5.355)	-7.455*** (-5.645)	2.419** (1.994)
<i>% White-collar workers</i>	-2.093 (-1.431)	-0.547 (-0.413)	4.240** (2.052)
<i>Size of establishment</i>	-6.670*** (-5.223)	-6.053*** (-5.431)	-2.882*** (-2.918)
<i>Intermediate input use</i>	70.683** (2.174)	32.240 (0.998)	-18.007 (-0.651)
<i>Sales to industry</i>	34.195** (2.027)	16.438 (0.999)	-13.233 (-0.934)
<i>Share of agricultural GVA *</i> <i>Agricultural input use</i>	0.060 (1.289)	0.065 (1.032)	0.108*** (3.428)
<i>Labour abundance</i> <i>* Share of labour in GVA</i>	0.554* (1.967)	0.948*** (3.710)	0.110 (0.434)
<i>Educated population</i> <i>* White-collar workers</i>	0.124 (0.390)	-0.505* (-1.741)	-0.508 (-1.088)
<i>Market potential</i> <i>* Size of establishment</i>	0.592*** (3.029)	0.825*** (5.195)	0.292** (2.286)
<i>Market potential</i> <i>* Intermediate input use</i>	-4.955 (0.939)	-4.793 (-0.977)	1.365 (0.352)
<i>Market potential</i> <i>* Sales to industry</i>	-2.721 (0.997)	-2.757 (-1.103)	0.948 (0.478)
<i>Number of observations</i>	295	300	301
<i>Adjusted R²</i>	0.63	0.56	0.56
<i>DWH Endogeneity Test: Chi-square (4)</i>	7.868*	8.607*	5.932
<i>Cragg-Donald F-statistic</i>	359.281	4209.759	4818.758

Note: White heteroskedasticity-robust standard error in brackets.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4. Robustness Check: Fixed effects. Dependent variable $\ln(s_{it}^k)$

	1856		1893	
	a	b	a	b
Share of agricultural GVA	0.215***	0.215***	0.067	0.067
* Agricultural input use	(0.081)	(0.076)	(0.047)	(0.060)
Labour abundance	0.015	0.015	0.584**	0.584*
* Share of labour in GVA	(0.350)	(0.388)	(0.277)	(0.307)
Educated population	-0.931**	-0.931*	0.104	0.104
* White-collar workers	(0.370)	(0.467)	(0.310)	(0.382)
Market potential	-0.040	-0.040	0.593***	0.593***
* Size of establishment	(0.186)	(0.171)	(0.176)	(0.198)
Market potential	0.545	0.545	-2.350	-2.350
* Intermediate input use	(3.440)	(4.637)	(5.144)	(6.837)
Market potential	0.282	0.282	-1.465	-1.465
* Sales to industry	(1.822)	(2.519)	(2.664)	(3.467)
Province dummies	Yes	Yes	yes	Yes
Industry dummies	Yes	Yes	yes	Yes
Number of observations	261	261	295	295
Adjusted R^2	0.66	0.66	0.62	0.62
F-test joint significance	17.36***	17.36***	19.88***	19.88***

Note: (a) White heteroskedasticity-robust standard error in brackets; (b) Cluster-robust standard error.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4 (continued)

	1913		1929	
	a	b	a	b
Share of agricultural GVA	0.063	0.063	0.106***	0.106***
* Agricultural input use	(0.060)	(0.067)	(0.032)	(0.032)
Labour abundance	0.959***	0.959***	0.118	0.118
* Share of labour in GVA	(0.267)	(0.292)	(0.254)	(0.321)
Educated population	-0.503*	-0.503	-0.504	-0.504
* White-collar workers	(0.295)	(0.384)	(0.477)	(0.610)
Market potential	0.825***	0.825***	0.299**	0.299**
* Size of establishment	(0.162)	(0.164)	(0.124)	(0.126)
Market potential	-5.542	-5.542	0.698	0.698
* Intermediate input use	(4.939)	(6.345)	(3.865)	(4.690)
Market potential	-3.169	-3.169	0.603	0.603
* Sales to industry	(2.548)	(3.280)	(1.992)	(2.439)
Province dummies	yes	yes	yes	yes
Industry dummies	yes	yes	yes	yes
Number of observations	300	300	301	301
Adjusted R^2	0.57	0.57	0.55	0.55
F-test joint significance	20.87***	20.87***	14.71***	14.71***

Note: (a) White heteroskedasticity-robust standard error in brackets; (b) Cluster-robust standard error.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5. Robustness Check II: Alternative specification. Dependent variable $\ln(s_{it}^k)$

	1856		1893	
	a	b	a	b
<i>Constant</i>	-202.203* (114.858)	-202.203 (138.389)	-238.636 (191.766)	-238.636 (233.786)
<i>Population</i>	0.007 (0.358)	0.007 (0.288)	-0.017 (0.218)	-0.017 (0.201)
<i>Manufacturing</i>	-0.056 (0.232)	-0.056 (0.185)	-0.128 (0.184)	-0.128 (0.174)
<i>Share of agricultural GVA</i>	-0.748** (0.357)	-0.748** (0.280)	-0.266 (0.229)	-0.266 (0.217)
<i>Mineral resources abundance</i>	-0.045 (0.045)	-0.045 (0.040)	-0.0004 (0.023)	-0.0004 (0.022)
<i>Labour abundance</i>	-0.248 (1.426)	-0.248 (1.473)	-2.165** (1.094)	-2.165** (1.117)
<i>Market potential</i>	-1.315 (20.845)	-1.315 (26.809)	14.501 (31.188)	14.501 (38.413)
<i>Agricultural input use</i>	-0.190 (0.314)	-0.190 (0.244)	-0.467 (0.315)	-0.467 (0.287)
<i>Mineral resources intensity</i>	1.613*** (0.266)	1.613*** (0.255)	0.525** (0.233)	0.525** (0.220)
<i>Share of labour in GVA</i>	4.014*** (1.129)	4.014*** (1.171)	-2.796* (1.552)	-2.796* (1.570)
<i>Size of establishment</i>	-0.162 (1.075)	-0.162 (0.940)	-5.213*** (1.432)	-5.213*** (1.494)
<i>Intermediate input use</i>	36.504* (18.641)	36.504 (22.266)	46.507 (31.984)	46.507 (31.916)
<i>Sales to industry</i>	10.661 (9.795)	10.661 (12.098)	21.671 (16.811)	21.671 (20.056)
<i>Share of agricultural GVA</i> * <i>Agricultural input use</i>	0.159* (0.084)	0.159** (0.072)	0.077 (0.047)	0.077 (0.055)
<i>Mineral resources abundance</i> * <i>Mineral resources intensity</i>	0.012 (0.020)	0.012 (0.020)	0.007 (0.011)	0.007 (0.014)
<i>Labour abundance</i> * <i>Share of labour in GVA</i>	0.065 (0.385)	0.065 (0.400)	0.588** (0.299)	0.588* (0.308)
<i>Market potential</i> * <i>Size of establishment</i>	-0.051 (0.189)	-0.051 (0.157)	0.583*** (0.191)	0.583*** (0.192)
<i>Market potential</i> * <i>Intermediate input use</i>	0.256 (3.401)	0.256 (4.312)	-2.594 (5.164)	-2.594 (6.375)
<i>Market potential</i> * <i>Sales to industry</i>	0.116 (1.813)	0.116 (2.359)	-1.643 (2.695)	-1.643 (3.262)
<i>Number of observations</i>	241	241	274	274
<i>Adjusted R²</i>	0.64	0.64	0.61	0.61
<i>F-test joint significance</i>	77.77***	120.48***	95.54***	80.19***

Note: (a) White heteroskedasticity-robust standard error in brackets; (b) Cluster-robust standard error.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5 (continued)

	1913		1929	
	a	b	a	b
<i>Constant</i>	-101.289 (202.800)	-101.289 (229.298)	167.797 (181.548)	167.797 (205.651)
<i>Population</i>	0.141 (0.270)	0.141 (0.249)	-0.139 (0.228)	-0.139 (0.140)
<i>Manufacturing</i>	-0.054 (0.212)	-0.054 (0.199)	-0.014 (0.188)	-0.014 (0.115)
<i>Share of agricultural GVA</i>	0.229 (0.303)	0.229 (0.330)	-0.129 (0.130)	-0.129 (0.090)
<i>Mineral resources abundance</i>	-0.0002 (0.022)	-0.0002 (0.024)	0.026 (0.038)	0.026 (0.038)
<i>Labour abundance</i>	-3.319*** (0.933)	-3.319*** (1.025)	0.164 (0.912)	0.164 (1.020)
<i>Market potential</i>	26.601 (30.563)	26.601 (36.089)	-19.440 (25.183)	-19.440 (29.585)
<i>Agricultural input use</i>	0.481 (0.339)	0.481 (0.328)	-1.022*** (0.230)	-1.022*** (0.191)
<i>Mineral resources intensity</i>	0.744*** (0.194)	0.744*** (0.171)	-0.435* (0.222)	-0.435** (0.185)
<i>Share of labour in GVA</i>	0.540 (1.351)	0.540 (1.158)	-3.791*** (1.300)	-3.791*** (1.095)
<i>Size of establishment</i>	-3.983*** (1.226)	-3.983*** (1.024)	-3.949*** (1.139)	-3.949*** (1.003)
<i>Intermediate input use</i>	19.723 (33.845)	19.723 (37.703)	-19.574 (30.253)	-19.574 (33.995)
<i>Sales to industry</i>	9.342 (17.268)	9.342 (19.603)	-12.930 (15.338)	-12.930 (17.552)
<i>Share of agricultural GVA</i> * <i>Agricultural input use</i>	0.066 (0.066)	0.066 (0.068)	0.117*** (0.028)	0.117*** (0.025)
<i>Mineral resources abundance</i> * <i>Mineral resources intensity</i>	0.032*** (0.012)	0.032** (0.014)	0.022 (0.019)	0.022 (0.021)
<i>Labour abundance</i> * <i>Share of labour in GVA</i>	0.925*** (0.255)	0.925*** (0.282)	0.009 (0.249)	0.009 (0.270)
<i>Market potential</i> * <i>Size of establishment</i>	0.800*** (0.159)	0.800*** (0.154)	0.235* (0.131)	0.235* (0.121)
<i>Market potential</i> * <i>Intermediate input use</i>	-4.676 (5.072)	-4.676 (5.935)	2.823 (4.182)	2.823 (4.875)
<i>Market potential</i> * <i>Sales to industry</i>	-2.830 (2.588)	-2.830 (3.097)	1.702 (2.114)	1.702 (2.505)
<i>Number of observations</i>	259	259	280	280
<i>Adjusted R²</i>	0.57	0.57	0.55	0.55
<i>F-test joint significance</i>	74.92***	74.08***	57.80***	57.16***

Note: (a) White heteroskedasticity-robust standard error in brackets; (b) Cluster-robust standard error.

* Significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6. Standardised beta coefficients. Interaction variables

	1856	1893	1913	1929
<i>Share of agricultural GVA</i>	0.0784	0.0271	0.0264	0.0461
<i>* Agricultural input use</i>				
<i>Labour abundance</i>	0.0048	0.0716	0.1231	0.0175
<i>* Share of labour in GVA</i>				
<i>Educated population</i>	-0.0709	0.0114	-0.0523	-0.0589
<i>* White-collar workers</i>				
<i>Market potential</i>	-0.0096	0.1495	0.2214	0.0962
<i>* Size of Establishment</i>				
<i>Market potential</i>	0.0495	-0.3036	-0.6931	0.1033
<i>* Intermediate input use</i>				
<i>Market potential</i>	0.0252	-0.2220	-0.4724	0.1069
<i>* Sales to industry</i>				

Source: see text. Statistically significant variables in bold.