Market Integration and Regional Inequality in Spain, 1860-1930

Julio Martínez Galarraga

PROGRAMA DE DOCTORAT EN HISTÒRIA ECONÒMICA (BIENNI 2000-2002)
DEPARTAMENT D'HISTÒRIA I INSTITUCIONS ECONÒMIQUES
UNIVERSITAT DE BARCELONA

MARKET INTEGRATION AND REGIONAL INEQUALITY IN SPAIN, 1860-1930

Tesi doctoral presentada per a l'obtenció del títol de doctor per la Universitat de Barcelona
Barcelona, setembre de 2010

Autor: Julio Martínez Galarraga
Director: Daniel A. Tirado Fabregat
Tutor: Alfonso Herranz Loncán

Contents

Acknowledgements **13**

Chapter 1. Introduction **15**

1.1. Regional inequality: the empirical evidence 15

1.2. Economic theory and regional inequality. An initial examination 19

1.3. Market integration and industrialisation in Spain 22

1.3.1. The integration of the Spanish market 23

1.3.2. The first stages in the process of industrialisation in Spain 30

1.4. Objectives and the development of the thesis 33

Chapter 2. New Economic Geography (NEG): theory and empirics **41**

2.1. Description of the theoretical framework: New Economic Geography 41

2.1.1. Background 41

2.1.2. 'New Economic Geography': economic integration, market access and regional inequality 45

2.1.3. Multi-regional NEG models: cross-country economic integration and the internal geography of countries 53

2.2. A survey of NEG empirics: market potential and agglomeration 57

2.2.1. Market access: relevance and origins 57

2.2.2. Market potential and the empirics of NEG 60

2.2.2.1. Market potential attracts production factors 61

2.2.2.2. Market potential raises the price of the production factors 63

2.2.3. Market potential, NEG and economic history 68

Chapter 3. Market potential in the Spanish provinces, 1867-1930 79

3.1. Introduction 79

3.2. Theoretical foundations underpinning market potential 82

3.3. The construction of provincial market potential in Spain, 1867-1930 86

3.3.1. Definition and selection of provincial and foreign 'nodes' 86

3.3.2. Internal market potential 89

3.3.2.1. The GDP of the Spanish provinces. 89

3.3.2.2. Inter-provincial transport costs. 93

3.3.2.3. The self-potential of each province. 102

3.3.3. Foreign market potential 103

3.3.3.1. Size of foreign markets. 105

3.3.3.2. International distances. 108

3.3.3.3. Tariffs. 109

3.4. Results and initial hypotheses 111

Chapter 4. The determinants of industrial location in Spain, 1856-1929 121

4.1. Introduction 121

4.2. Transport costs and industrial location in Spain 129

4.3. Empirical strategy 133

4.3.1. The model 133

4.3.2. Data 137

4.3.3. Comparative advantage and NEG variables: a description of the patterns 140

4.4. Estimation results 145

4.4.1. Robustness, alternative specifications and standardised coefficients 151

4.5. Discussion 160

Chapter 5. Market potential and economic growth in Spain, 1860-1930 167

5.1. Introduction 167

5.2. Related empirical literature: first nature vs. second nature geography 173

5.3. Empirical strategy 178

5.3.1. The model 178

5.3.2. Data 183

5.4. Estimation results 188

5.5. Discussion 192

Chapter 6. Conclusions 197

Appendix 207

Annex 1. The calculation of internal transport costs between Spanish provinces 214

Annex 2. Transport costs and the calculation of the foreign market potential 223

Bibliography 231

List of Tables

Table 1.1. Migratory flows in Spain. 28

Table 3.1. Distribution of the volume of traded goods by coastal shipping and rail 95

Table 3.2. Mean freight rates charged by the railway companies for the transport of commodities (current pesetas/tkm) 98

Table 3.3. Maritime freight rates 100

Table 3.4. Peseta/pound sterling exchange rate 106

Table 3.5. Estimated peseta/pound sterling exchange rate 107

Table 3.6. Foreign GDP (millions of pesetas at current prices) 108

Table 3.7. Mean tariffs (% respect to imports) 110

Table 4.1. Region and industry characteristics 136

Table 4.2. Regression results. Dependent variable $\ln(s_{it}^{k})$ 149

Table 4.2. (continued) 150

Table 4.3. Robustness Check: Fixed effects. Dependent variable $\ln(s_{it}^{k})$ 151

Table 4.3. (continued) 152

Table 4.4. Estimation results: 2-step GMM/IV. Dependent variable $\ln(s_{it}^{k})$ 154

Table 4.5. Robustness Check II: Alternative specification 156

Table 4.5. (continued) 157

Table 4.6. Standardised beta coefficients. Interaction variables 159

Table 5.1. Growth regressions. Dependent variable: GDP per capita growth 189

Table A1. Geographical distribution of Spanish exports (%) 207

Table A2. Provincial 'nodes' connected to the railway network in 1867 207

Table A3. Agrarian and industrial wages in Spanish provinces, 1860-1930 208

Table A4. Provincial Gross Domestic Product, 1860-1930 (current million pesetas) .. 209

Table A5. Market potential, 1867-1930 (current millions pesetas) 210

Table A6. Market potential relative to Barcelona and to 1867 211

Table A7. Market potential, self-potential excluded (current millions pesetas) 212

Table A8. 'First nature' variables 213

Table A9. Transport costs matrix for 1867 215

Table A10. Inland provinces (Mediterranean) and Mediterranean ports, 1867 225

Table A11. Inland provinces (Mediterranean) and Atlantic ports, 1867 225

Table A12. Inland provinces (Atlantic) and Atlantic ports, 1867 226

Table A13. Inland provinces (Mediterranean) and Mediterranean ports, 1900-1930 .. 228

Table A14. Inland provinces (Mediterranean) and Atlantic ports, 1900-1930 228

Table A15. Inland provinces (Atlantic) and Atlantic ports, 1900-1930 229

List of Figures

Figure 3.1. Peseta/pound sterling exchange rate, 1860-1931 107

Figure 3.2. Convergence in market potential of the Spanish provinces, 1867-1900 114

Figure 3.3. Convergence in market potential of the Spanish provinces, 1900-1930 116

Figure 3.4. GDP per capita and market potential, 1867-1930 118

Figure 3.5. GDP per capita and market potential, excluding self-potential.................. 118

Figure 5.1. Long term regional per capita GDP inequality. Spain's NUTS3 168

Figure 5.2. The geographical equilibrium.. 182

List of Maps

Maps 3.1. Expansion of the railway network in Spain (1855-1923) 96

Maps 3.2. Market potential in Spain's provinces, 1867-1930 (Barcelona=100) 111

Maps 4.1. Gross Value Added in Agriculture as a % of GDP, Spanish provinces 142

Maps 4.2. Total active population per square km, Spanish provinces.......................... 143

Maps 4.3. Literacy rate, Spanish provinces 144

Maps 4.4. Market potential, Spanish provinces (Barcelona=100)................................ 145

Maps 5.1. 'First nature' variables 187

Acknowledgements

This research project has been developed in the Economic History Department at the University of Barcelona. I would like to acknowledge its members for their assistance and continuous support throughout these years. Writing a thesis is a long journey. However, the human and academic environment I found at the department definitely made things easier.

I am personally and intellectually indebted to my supervisor, Daniel A. Tirado. It is difficult to find words to express how thankful I am not only for his research guidance but also for his generosity. I am also particularly grateful to all the members of the *Grup de recerca en globalització, desigualtat econòmica i polítiques públiques amb perspectiva històrica* in the *Xarxa de Referència d'R+D+I en Economia i Polítiques Públiques* (XREPP) launched by the Generalitat de Catalunya. This group of excellent human beings has represented a permanent stimulus for the research and a source of encouragement. Among them, the suggestions and help provided at different stages of the research by Marc Badia, Sergio Espuelas, Jordi Guilera, Alfonso Herranz and Marc Prat have contributed to improve the final result. Similarly, I would like to thank Adoración Álvaro, Laura Calosci, Anna Castañer, Tomàs Fernández de Sevilla, Fabian Gouret, Celia Lozano, Pilar Nogués, Javier San Julián, José Miguel Sanjuán, Raül Santaeulalia-Llopis, Daniel Vázquez and Henry Willebald for their academic and personal support.

A good number of institutions have contributed to this thesis with financial support. First, the Department of Economic History offered me the possibility of taking part in a project aimed at the reconstruction of teaching materials for the World Economic History course. I am grateful to Enric Tello and Ramon Ramon for their confidence and for being so understanding. During these years, I did benefit from the resources supplied by the *Ministerio de Educación y Ciencia* to the research projects BEC 2002-00423 and SEJ2005-02498/ECON, directed by Jordi Catalan, the research project ECO2009-1331-C02-02/ECON directed by Alfonso Herranz, and the ESF-Globaleuronet project "Historical Economic Geography in Europe, 1900-2000" organised by Joan Ramon Rosés and Nikolaus Wolf. Moreover, I enjoyed economic support from various grants provided by the *Banco de España*, the *Institut d'Economia i Empresa Ignasi Villalonga*, and the *Fundació Ernest Lluch*, and I also had access to the financial resources of the *Centre d'Estudis Antoni de Capmany*.

Different parts of this thesis were presented in the summer schools organised by the ESF-Globaleuronet. I really appreciate the insightful suggestions I received from the participants, especially those by Nick Crafts and Herman de Jong. I would also like to thank the participants in the Workshops Iberometrics III and IV held in Valencia and Lisbon, respectively, and in the Economic History Seminar at the University of Zaragoza. Likewise, I could benefit from the advice and comments on my research project made by some of the most influential economic historians that during these years have visited the University of Barcelona, like Peter Lindert, John Murray, Leandro Prados de la Escosura, Max Schulze and Jeffrey Williamson. Lastly, I would also like to acknowledge Bartolomé Yun for his hospitality during my visit at the European University Institute (EUI) in Florence.

On the personal side, over the years, I received the support from my family and friends. To all of you: thanks. And especially to my sisters, and to Anna, who shared this journey with me. Finally, I wish to express my deepest gratitude to my parents. This thesis is dedicated to them.

Chapter 1. Introduction

1.1. Regional inequality: the empirical evidence

Regional inequality is one of the main features characterising the Spanish economy today. In recent years, Spain has experienced strong economic growth, which has given rise to a marked process of convergence with the European Union. According to Eurostat figures, in 1999 Spain's Gross Domestic Product (GDP) per capita stood at 83.5% of the EU average (100) as calculated for the fifteen states that at that time made up the union (EU-15). By 2009, the most recently available data for this measure, Spain had recovered ten percentage points, reaching 93.7% of European GDP per capita[1].

However, despite the convergence of the Spanish economy with Europe's, regional disparities within Spain persist. Referring once again to figures published by Eurostat, in 2007 the Spanish regions with the highest GDP per capita were Madrid and the Basque Country at 137 points above the average of 100 for EU-27. In that year, the per capita income in these regions almost doubled Extremadura's, the Spanish region with the lowest GDP per capita (72%)[2]. Between these two extremes a complex regional structure has developed over time reflecting a range of historical events.

The evolution in inequality in Spain has been well documented since 1955, the year in which the Banco Bilbao Vizcaya (BBV, 1999) began publishing income series data which included GDP per capita statistics broken down by region and province (which correspond to NUTS2 and NUTS3 according to the EU territorial division, respectively). On the basis of the information provided by BBV, a number of studies involving the analysis of regional growth in Spain have been produced. During the 1990s, and following the pioneering work of Barro (1991), and Barro and Sala-i-Martin (1991, 1992, 1995), there was a reawakening of interest in the study of the convergence between

[1] If the comparison is made with the present 27 member states, the Spanish economy rose from 96% in 1999 to 104% in 2009.

[2] In absolute terms, in 2007 Madrid's GDP per capita (PPS) reached 34,100 euros compared to 18,000 euros in Extremadura.

economic areas. In these years many empirical studies were published within the framework of the economic growth literature in which the existence of economic convergence between countries and regions was tested in terms of two basic notions: on the one hand, σ-convergence, which implies the reduction in dispersion of a variable (typically GDP per capita) over time; and, on the other, β-convergence, which implies an inverse relationship between the growth rate and the initial level of GDP per capita of countries and regions. Here, the empirical strategy normally adopted involves the estimation of growth regressions. These convergence equations are derived directly from the economic growth models in order to test whether the economies that initially present the lowest GDP per capita grow at a relatively more rapid rate than the richer ones in absolute terms (unconditional convergence) or once distinct variables have been controlled for (conditional convergence). In the Spanish case, a good number of studies have examined the existence of regional convergence during the period for which the BBV data series offers information. The results point to the existence of convergence in both of its two forms (β and σ) from 1955 until the end of 1970s. After this date, the convergence process came to a halt and in recent decades the process has shown signs of exhaustion[3].

The convergence patterns that can be inferred from these studies of Spain during much of the second half of the 20th century are not exclusive to the Spanish economy. Since its foundation, a similar process has been documented in the European Union too. Various studies at the European level show a similar evolution to that recorded in Spain, i.e., the existence of regional convergence up to the end of the 1970s[4]. Yet, since the 1980s, the trend towards greater regional convergence was interrupted. As Puga (2002) showed, since the mid-1980s, inequality between countries within the European Union has fallen in a process of marked convergence at the same time as regional inequality within the countries has increased. As a result, the absence of regional convergence has converted the uneven geographical distribution of per capita income at the regional level

[3] At the regional level, de la Fuente (2002), Mas, Maudos, Pérez and Uriel (1994) and Raymond and García-Greciano (1994). At the provincial level, see Dolado, González-Paramo and Roldán (1994) and García-Greciano, Raymond and Villaverde (1995).

[4] Armstrong (1995), Barro and Sala-i-Martin (1991), Button and Pentecost (1995), Neven and Gouyette (1995), Fagerberg and Verspagen (1996), and Sala-i-Martin (1996). A review of the empirical analyses undertaken within the growth literature can be found in Magrini (2004). An alternative approach to the study of regional inequality can be found in Quah (1996), Caselli, Esquivel and Lefort (1996), Cánova and Marcet (1995), Rodríguez-Pose (1999) and Barrios and Strobl (2009).

within countries into a persistent feature of the European economies, a situation that remains cause for concern for today's policymakers[5].

Returning to the Spanish case, the study of long-term regional inequality has proved particularly difficult due to the paucity of available statistical data. For the period prior to 1955, the year in which the BBV series were first published, the availability of regional GDP estimates falls notably. For the whole of the 19th century and the first third of the 20th century, the only available data are those provided by Álvarez Llano (1986) who reports the distribution of GDP by Autonomous Community (NUTS2) for six different points in time[6]. Yet, the fact that the exact method employed by this author in preparing these data is unknown undermines considerably the reliability of these series[7]. In addition, Alcaide (2003) reported data, in this instance for both regional (Autonomous Communities) and provincial levels, for much of the 20th century (1930-2000).

Drawing on the information provided by Álvarez Llano (1986), Albert Carreras (1990a) undertook an initial evaluation of long-term regional inequality in Spain. The results allowed him to define the regional patterns of economic development and to analyse the dynamics of regional inequality in Spain since the 19th century by means of the inequality index. Carreras found a constantly increasing trend in inequality from 1800 onwards reaching a peak around 1950 or 1960. From that moment onwards, GDP per capita across regions began to decline and by 1983, the inequality observed was less than that recorded at the start of the analysis almost two centuries earlier[8].

At an international level, limited availability of long term data describing the regional distribution of income per capita hinders (in common with the Spanish case) the task of undertaking a historical analysis of regional income inequality, although there are a number of notable exceptions. Among these, the study conducted by Kim and Margo (2004) for the United States is perhaps the best example. The evidence available shows

[5] In Spain, Madrid's GDP per capita in 2006 was 1.92 times that of Extremadura. A slightly bigger difference was to be found between Île-de-France and the Languedoc-Roussillon (1.98) in France and between Bolzano and Campania (2.05) in Italy. Even more marked were the differences in Germany between Hamburg and Brandenburg-Nordost (2.65), and particularly in Great Britain where the GDP per capita in Inner London was 4.34 times that of the region of West Wales and the Valleys. In terms of the coefficient of variation, with the exception of France (0.18), regional inequality (NUTS2) in Germany (0.23), Italy (0.24), the United Kingdom (0.37) and the European Union (EU-27) as a whole (0.39) was higher than that recorded in Spain (0.19).

[6] Álvarez Llano (1986) provides estimations for 1802, 1849, 1860, 1901, 1921 and 1930.

[7] However, while these figures have been questioned by a number of authors, it has also been argued that these results approximate to the most qualitative information we have regarding the evolution in regional economies. A critique of these data can be found in Carreras (1990a).

[8] This evolution can be completed with Martín (1992) and Domínguez (2002). As in Carreras (1990a), the analyses undertaken were based on the estimates of regional GDP provided by Álvarez Llano (1986).

that in the colonial era the differences in regional income per capita in the US economy were not marked. Yet, in the early decades of the 19th century these differences had begun to increase, intensifying remarkably during the second half of the century. However, throughout the 20th century, and especially after the end of World War II, a substantial reduction in regional inequality occurred in the United States[9].

The evolution in long-term regional inequality, as outlined above, is very much in line with the little empirical evidence available at the international level. Here Williamson's (1965) study stands out. Following the line of analysis suggested by Kuznets[10], he examined the relationship between national economic development and the evolution in regional inequalities. Based on a sample of ten countries throughout the 19th and 20th centuries[11], Williamson concluded that in the initial stages of development an increase in regional inequality is observed, while in later stages a trend towards convergence emerges that leads to a reduction in spatial disparities: *"...the early stages of national development generate increasingly large North-South income differentials. Somewhere during the course of the development, some or all of the disequilibrating tendencies diminish, causing a reversal in the pattern of interregional inequality. Instead of divergence in interregional levels of development, convergence becomes the rule, with the backward regions closing the development gap between themselves and the already industrialized areas. The expected result is that a statistic describing regional inequality will trace out an inverted "U" over the national growth path; the historical timing of the peak level of spatial income differentials is left somewhat vague and may vary with the resource endowment and institutional environment of each developing nation*"[12]. Williamson tied this evolution to the presence of four mechanisms that acted either in favour of divergence or convergence at different points in time: migration, capital markets, public policies adopted by the central government and interregional links.

The empirical studies discussed in the preceding pages show that in both the Spanish case and that of the United States, as well as more generally, there would appear to be an increase in regional differences in the early stages of economic development. With time, however, these differences tend to diminish so that regional inequality in the long run describes an inverted U-shaped curve. In the section that follows the different

[9] See also Kim (1998), Caselli and Coleman (2001) and Yamamoto (2008).
[10] Kuznets (1955). Also Hirschman (1958) and Myrdal (1957).
[11] The countries included in the sample are the United States, the United Kingdom, France, Canada, the Netherlands, Sweden, Norway, Italy, Brazil and Germany.
[12] Williamson (1965), p. 9 and 10.

theoretical approaches that have been adopted for the study of regional inequality are briefly presented so as to determine the extent to which the theoretical predictions are in line with the findings in the empirical studies described so far.

1.2. Economic theory and regional inequality. An initial examination

The theoretical study of regional inequality has traditionally been undertaken from within regional economics. In this field, neoclassical analysis has predominated with the use of models based on the existence of constant returns to scale and markets operating in perfect competition, although the theoretical developments made in recent decades have allowed for alternative approaches. The contributions to the study of regional inequality can be divided between those that originated from growth theory and those that originated from international trade theory.

In the case of growth theory, the neoclassical models (Solow, 1956) suggested that regional income could differ on the basis of differences in the capital-labour ratio between regions. Yet, economic integration can favour regional convergence. In short, under the aforementioned assumptions the factors of production behave as follows. Regional wage differentials generate forces of attraction that act on the workers, giving rise to migratory flows that with time end up equalising wages. Likewise, capital flows from those regions where capital is abundant in to those where it is scarce. Finally, the capital-labour ratio becomes equal, as do the returns on capital and labour. Consequently, the neoclassical literature holds that economic inequalities tend to eventually disappear and the theoretical models predict the existence of convergence in equilibrium.

In turn, the endogenous growth theory developed in the 1980s (Romer, 1986; Lucas, 1988), by incorporating increasing returns, suggests the possible existence of regional divergence. The reason lies in the uneven distribution over time and in space of technology, considered the engine of long-term growth. Therefore, economic growth can favour the increase in regional inequality, whereby the latter can present a non-linear evolution over time[13].

In the case of international trade theory, the Heckscher-Ohlin neoclassical model establishes that interregional differences in income are the result of differences in factor endowments and their prices. Here, economic integration leads to a convergence in

[13] Galor (1996), Pritchett (1997) and Lucas (2000) show the non-linear nature of the convergence processes.

regional incomes due to the factor price equalisation across the regions. However, as a consequence of factor endowment differences, the regions can specialise in different industries in accordance with their comparative advantage. Thus, if the factor endowment differences become more marked over time, the industrial structures of the regions will diverge, as will regional incomes.

The New Trade Theory models are based on alternative assumptions including imperfect competition and increasing returns to scale in the manufacturing sector. In this case, economic integration accounts for the appearance of differences in the regions' productive structures according to differences in market size. The greater factor endowment enjoyed by the larger market (determined exogenously) of the 'core' region compared to that of the 'peripheral' region would result in the relative specialisation of the 'core' in the production of manufactured goods while the 'periphery' specialises in agricultural goods. Thus, the respective productive structures of the regions would experience divergence giving rise to differences in their total incomes.

The more recently developed framework of New Economic Geography (hereinafter, NEG) focuses its analysis on the determinants of the spatial localisation of economic activity and their dynamics over time (Krugman, 1991). In broad terms, in the NEG models, transport costs and increasing returns interact in a framework of monopolistic competition favouring the spatial agglomeration of economic activities, which gains in strength once it is set in motion[14]. Here, the reduction in transport costs and the progressive integration of the markets of goods and factors play a key role, since they can favour the spatial concentration of economic activities and, as a result, increase regional inequalities. Yet, studies such as that undertaken by Puga (1999) show that the relationship between regional economic integration and the degree of spatial concentration of economic activity can describe a bell-shaped, non-monotonic evolution once the forces of dispersion are activated (for example, congestion costs, wage differentials and the fragmentation of firms). In this way, the progressive integration of the market can give rise after a certain point has been reached to regional convergence.

[14] Krugman (1991), Fujita, Krugman and Venables (1999). Unlike New Trade Theory, NEG can explain the mechanisms by which sizeable differences can be generated in the regions' productive structures and income levels, even when these regions present similar factor endowments. What makes the models of new economic geography attractive is the fact that the cost parameters and the level of demand are endogenous and differ between locations as they depend on the location decisions taken by all the agents. This distinguishes these models from those of international trade with imperfect competition in which the location of the factors of production is given and fixed (exogenous). Combes, Mayer and Thisse (2008), p. 47.

The forces stressed by Traditional Trade Theory and NEG are, nonetheless, not exclusive and they might be in operation simultaneously. In this regard, several empirical studies have attempted to disentangle the relative strength of comparative advantage in factor endowments and increasing returns as the driving forces behind the spatial distribution of manufacturing. Is it possible to explain the observed patterns in the location of industry on the basis of these two alternative theoretical approaches? Kim (1995, 1999) suggested that in the case of the US, where a remarkable concentration of manufacturing activities in the north east of the country took place, the role of increasing returns was limited and most of the explanation relied on differences in factor endowments across states. However, Klein and Crafts (2009) have recently questioned this explanation, and they argue that NEG forces were the main determinants in the emergence of the manufacturing belt in the US between 1880 and 1920. They used an alternative approach based on the proposal of Midelfart-Knarvik, Overman, Redding and Venables (2002). These latter authors examined the spatial evolution of manufacturing activities in the last decades in the European Union, where the process of economic integration has raised concerns about the potential impact of the process on the geographical redistribution of economic activities. They found that both Heckscher-Ohlin and NEG forces were in operation between the 1970s and the 1990s. An analogous result was obtained by Wolf (2007) for interwar Poland after World War I. The changes in the manufacturing sector after the reunification and the creation of the Polish national market were generated by the same forces that are shaping the distribution of manufacturing activities in the European Union in recent times. Finally, the case of Britain in the Victorian period has also been analysed (Crafts and Mulatu, 2005, 2006). The British manufacturing showed, one century after the Industrial Revolution had started, a relatively stable geographical pattern. From 1870 onwards, factor endowments were determining the location of industry, although a role for increasing returns was also found. Overall, the analysis of different national experiences and different periods has shown that the forces behind the spatial concentration of industry and regional specialisation may be diverse, including those suggested by both Traditional Trade Theory and the New Economic Geography.

The evolution of the industrial sector is however, only part of the story. For the study of regional inequality the whole economy needs to be considered. At an international level, in terms of income per capita, Redding and Venables (2004) have shown that market potential as suggested by NEG is a powerful variable to explain the differences recorded in GDP per capita across countries in current times. Mayer (2008) has recently extended this conclusion to the post-World War II era. Yet, the analysis of the role played by economic geography and market potential at a regional level within countries has not been so abundant. Most of the regional analyses conducted within NEG are based on the verification of Krugman's (1991) wage equation. However, the absence of long-term estimates of regional income for most of the countries hinders the study of the impact of market potential on income inequality.

This brief review allows extracting a number of conclusions with respect to the possible factors that might result in regional inequality, and its persistence over time. But when were these inequalities first generated? Based on the theoretical models, in particular the predictions derived from the NEG models, two elements, to a certain point parallel in time, emerge as possible explanations. On the one hand, the role played by economic integration, which in this case would take us back to the period in history in which the respective domestic markets were created. On the other hand, the processes of industrialisation, since it is in the industrial sector where increasing returns to scale tend to operate with greatest intensity.

1.3. Market integration and industrialisation in Spain

According to Combes, Mayer and Thisse (2008), the spatial inequalities in the distribution of economic activity and income observed today are the result of a long-term evolution that can be traced back to the Industrial Revolution. Furthermore, the processes of industrialisation that in many cases were set in motion in the 19th century coincided with domestic market integration. The reduction in trade costs between different areas of the same country was, on the one hand, linked to the elimination of institutional obstacles that hindered the free movement of goods and factors between regions and, on the other, to the fall in transport costs derived from the technological improvements made during the Industrial Revolution and their application to transport.

In this section, the major advances in these two fields are examined: firstly, the main characteristics in the integration of the Spanish market are described, and then the process of industrialisation in Spain during the second half of the 19th century from a regional perspective is presented.

1.3.1. The integration of the Spanish market

One of the features that economic theory has highlighted in the genesis of regional inequality is the emergence of national markets. In the case of Spain, the economic integration of the various regional economies was completed during the second half of the 19th century. Before that date, during the '*Antiguo Regimen*' (Ancient Regime), the Spanish market was fragmented into various local and regional markets that were largely unconnected. Historians have stressed two key elements to account for this situation: on the one hand, the persistence of institutional obstacles to interregional trade and, on the other, the relative backwardness and deficiencies suffered by Spain's transport system.

The persistence of barriers and limitations to internal trade led to the fragmentation of the Spanish market during the 18th century and the early decades of the 19th. This fragmentation was the result of extensive market regulations and excessive intervention that hindered the transport of goods. The existence of taxes, tolls and tariffs and even the survival of domestic customs borders to enter the Basque Country and Navarre (the so-called 'dry ports') were some of the obstacles obstructing the free movement of goods between different areas of Spain (Madrazo, 1984; Simpson, 1995a). The result of all these barriers was that during this period the regional markets were characterised by their low level of integration.

Yet, the second half of the 19th century was witness to a progressive integration of the domestic market thanks to the institutional reforms undertaken by the various liberal governments. These reforms were aimed at strengthening property rights and fostering a reduction in transaction costs that interfered in economic relations and impeded the free movement of goods within Spain's borders. Important here was the elimination of the main restrictions on interregional trade, including tariffs and the domestic customs, the suppression of the guilds and the "Mesta" (a medieval association of cattle farmers), the disentailment of real estate and land ('*desamortización*'), the abolition of entailed estates ('*mayorazgos*'), and the unification of the system of weights and measures that until that time had varied from region to region (Tedde, 1994; Carreras and Tafunell, 2004).

As for the transport system, Spain has traditionally had to confront geographical obstacles that have hindered its development. The difficulties faced by traditional land transport derive from the peninsula's mountainous relief. In addition, Spain is characterised by the absence of navigable inland waterways due to the climatic conditions that result in rivers of poor and irregular flow. Thus, the Spanish economy was deprived of an alternative mode of transport such as inland navigation, which in other countries played a fundamental role for the transport of goods and passengers. In turn, cabotage (coastal shipping), which could have compensated for this lack of navigable inland watercourses, did not show signs of life in the first half of the 19th century. As for overland transport, historians have stressed the inappropriate radial pattern of the road network that hampered the connection between the various regional markets. Not only the design of the network, but also the poor state of conservation of the roads would have delayed the establishment of a modern transport system. In this case, the reasons can be found in the difficulties that the Spanish public finances had to face during this period and which reduced the investment devoted to the maintenance and repair of the roads. Finally, the means of land transport were also out-dated and the supply of transport, provided by farmers who combined this activity with their farm work, was also insufficient. Together, these factors resulted in a system of land transport that was slow, expensive and, in general, inefficient (Ringrose, 1970; Gómez Mendoza, 1982)[15]. As a consequence, domestic market integration had not progressed very far before the arrival of the railway, although some authors claim that the process had already started in the Modern period.

Yet, the improvements made to the transport system, especially those completed in the second half of the 19th century, proved to be a determining factor for the integration of the Spanish market, thanks both to the introduction of the railway and the advances made in other modes of transport. First, the road network was improved substantially. The total length of paved roads, which at the end of the 18th century extended over little more than 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).

Coastal shipping also underwent far-reaching changes in the second half of the 19th century. Among the main technological innovations adopted for cabotage were the introduction of iron, which allowed the capacity of transport to be increased, and the substitution of sail with steam, which meant considerable timesavings. These innovations were accompanied by the improvements to the infrastructure of some of the peninsula's main ports, which allowed the docking of ships of much greater tonnage. As a result, the total volume of goods transported by cabotage rose from the meagre level of 690,000 tonnes in 1857 to 2.02 million tonnes in 1900. And this despite the competition it faced from the 1850s onwards from the railway, especially along the Mediterranean routes. This

[15] By contrast, Frax and Madrazo (2001) present a more optimistic perspective of road transport before 1850, claiming that the flow of transport on Spanish roads presented a more dynamic behaviour.

competition was also reflected in the transport prices. During this period the railway companies tried to attract a greater volume of trade away from the coastal shipping companies by lowering their rates, which resulted in a reduction in the transport prices in the second half of the 19th century (Frax, 1981; Gómez Mendoza, 1982; Pascual, 1990).

However, the integration of the Spanish market received its greatest impulse from the construction of the railway. The expansion of the rail network brought with it major changes that favoured the progressive development of the domestic market. The Railway Acts of 1844 and 1855 established the legal framework for the construction of the railways and resulted in their radial design centred on Madrid. The first line was finished in 1848, covering the 28 kilometres that separated Barcelona and Mataró. Over the following decades the basic network was completed, so that by 1901, with the completion of the stretch that linked Teruel up to the network, all the provincial capitals were connected to the railway (Wais, 1987). The Spanish historiography similarly stresses the great steps taken in the expansion of the railway during the period from 1855, the date of the introduction of the Railway Act, until 1866 when the railway linked up Spain's main economic centres. In barely ten years, the length of the broad gauge network grew from 440 to 5,076 kilometres. In a second stage, between 1873 and 1896, the railway arrived to the rest of the country and by the end of the 19th century the railway network covered a distance of 10,827 kilometres (Gómez Mendoza and San Román, 2005; Herranz, 2005) [16].

As a result, in the second half of the 19th century the basic rail network was completed and it became a key element in the integration of the Spanish market. One of the most notable effects of the construction and expansion of the new railway infrastructure was the fall in transport costs. According to the calculations made by Herranz (2005), in 1878 the ratio between the unit price of goods transported by rail and the price using an alternative mode of transport was 0.14, which represented a reduction of up to 86% in haulage costs thanks to the introduction of the railway. At the same time, the expansion of the telegraph also helped not only in the acceleration of the transmission of information but also in the reduction of the firms' transaction costs (Calvo, 2001). Between 1855 and 1900 the length of the telegraph lines in Spain expanded at a considerable rate, rising from 713 to 32,320 kilometres (Herranz, 2004).

The impact of all these improvements, including the abolition of institutional obstacles to interregional trade and the development of the transport system, resulted in

[16] Likewise, the country's infrastructure stock as a share of its GDP rose from 4.27% in 1850 to 27.21% in 1900. Herranz (2001).

the gradual integration of the goods market during this period for the main traded products. Indeed, the integration of the Spanish market was characterised by a convergence in regional prices. Various studies have analysed the evolution in Spain's grain markets from its early beginnings in the 18th century until its culmination in the second half of the 19th century (Sánchez-Albornoz, 1975; Peña and Sánchez-Albornoz, 1983, 1984; Barquín, 1997; Martínez Vara, 1999; Reher, 2001; Llopis and Sotoca, 2005; Matilla, Pérez and Sanz, 2008).

The integration of the factors markets, in common with the goods market, also underwent marked advances. In the case of the capital markets, the main events affected the monetary and banking systems. From the end of the 18th century there existed a great number of coins in circulation within the Spanish market originating from different regions and periods of history. This situation lasted throughout the first third of the 19th century (Sardà, 1948) and even by 1864 there were still 84 different coins in circulation (Martorell, 2001). However, the decree enacted in 1868 by the then treasury minister, Laureano Figuerola, unified Spain's monetary system which from that time on was founded on a single currency: the peseta. In this way, the peseta became the new official monetary unit within a bimetallic system linked to the Latin Monetary Union.

In addition to the monetary unification, various advances were made in the banking system. During the 18th century and much of the 19th, the Spanish banking system was characterised by its backwardness, which explains the survival of a system based on bills of exchange[17]. The evolution and possible convergence in interest rates in the different local markets have been analysed on the basis of the price of short-term bills of exchange. The results show that in the second half of the 19th century interregional variations fell, a sign of the gradual integration of the money markets (Castañeda and Tafunell, 1993; Maixé-Altés and Iglesias, 2009). This fall was accompanied by profound reforms in the banking system. The beginnings of this process of modernisation can be traced to the 1840s and 1850s and the introduction of new banking legislation. This new legal framework paved the way for the foundation of private banks as companies with

[17] "*The highly specific structure of the Spanish banking system was one of the reasons why a transfer system based on local markets dealing in bills of exchange survived*". Maixé-Altés and Iglesias (2009), p. 501.

limited liability (Tortella, 1973) and, after 1856, the provincial banks received the right to issue bank notes (Sudrià, 1994)[18].

With the 'Restoration' of the monarchy in the last quarter of the 19th century, new legislative measures were adopted that meant a major step forward in completing the integration of the capital markets. The Echegaray Decree of 1874 put an end to the plural system based on various banks of issue and granted the Bank of Spain the monopoly. At the same time, the Bank of Spain gradually opened branches in various provincial capitals[19]. Finally, in 1885 a transfer system was established between accounts held at different branches of the Bank of Spain (Tortella, 1970), with the effect that short-term bills of exchange could eventually be replaced at the same time as the integration of Spain's capital market was completed (Castañeda and Tafunell, 1993).

Spain's labour market integration has been analysed in a context characterised by the low intensity of its interregional migration during the 19th and the early decades of the 20th centuries. In the case of Spain, as in most pre-industrial societies, throughout the 18th century and much of the 19th, the few internal migrations recorded were primarily temporary in nature and occurred over short distances. The predominance of agrarian activities meant that the workers' mobility was restricted to rural areas and, in turn, depended on the seasonality of the harvests.

It was in the 1860s when the number of permanent internal migrations rose due to the early industrialisation of some areas, the effects of which were felt on agricultural employment (Erdozáin and Mikelarena, 1996). Yet, during the second half of the 19th century and up to the 1920s, the number of internal migrations remained small. In the decades between 1877 and 1920, the percentage of permanent migrations was remarkably stable with figures between 2 and 2.9% of the whole population (Silvestre, 2005). Urban growth also failed to show significant changes in this period (Luna, 1988; Reher, 1994). The late industrialisation and the slow structural change in the Spanish economy meant, as in other countries of southern Europe and the Mediterranean (Hatton and Williamson, 1998; O'Rourke and Williamson, 1999), that the level and variation of the internal mobility of the labour factor over time were only moderate[20].

[18] Before 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, twenty banks had issuing rights.

[19] Yet, Spain's commercial banks retained their regional character and did not develop a nationwide network of branches until well into the 20th century. Anes, Tortella and Schwartz (1974).

[20] Spanish historians offer two types of explanations for the low rate of internal migration. On the one hand, supply-based explanations support the idea that it was the lack of dynamism in the agricultural sector and

It was in the 1920s that a great increase was recorded in internal migratory flows, with rates that practically doubled those observed in the preceding decades (4.3% of the overall population)[21]. In this case, the increase was caused by the greater opportunities for work outside the agricultural sector resulting from Spain's industrial and economic growth as well as from the increase in regional wage differentials (Silvestre, 2005).

However, permanent internal migrations were accompanied by another type of migration that was also to play a key role in this period. International migrations increased in the 1880s, although this trend was interrupted in the 1890s. However, the greatest period of international migration coincided with the years between the first decade of the 20th century and the start of World War I, when these flows reached considerable volumes (Sánchez-Alonso, 2000). In the interwar years, emigration fell in a general context characterised by the distortions generated in the international markets (the so-called globalisation backlash), just when permanent internal migrations underwent an intense growth. The number of international migrations was virtually twice that of internal migrations at the end of the 19th century. In the interwar years, this situation changed substantially:

Table 1.1. Migratory flows in Spain.

	1888-1900	1901-1910	1911-1920	1921-1930
(1) Permanent internal migrations	428,253	565,830	583,123	968,581
(2) External migrations	903,023	1,349,037	1,813,317	1,128,312
(3) = (1) / (2)	47.42%	41.94%	32.16%	85.84%

Source: Silvestre (2005) and Sánchez-Alonso (1995).

However, Silvestre (2007) highlighted the importance and persistence of temporary internal migrations. In a primarily agrarian economy, and in which therefore, seasonal activities linked to farming were predominant, temporary migration continued to have a considerable weight. Even in the 1920s, when the number of permanent internal

its incapacity to free up the workforce that impeded migration from rural to urban zones, frustrating any structural changes. By contrast, other authors forward demand-based explanations arguing that the lack of mobility in the workforce within Spain was the result of the weak process of industrialisation and the low level of attraction generated by the cities. This debate and the main contributions to it can be followed in Silvestre (2005).

[21] In absolute terms, internal migrations increased gradually from 369,424 people in the intercensal period between 1877 and 1887 to 583,123 between 1911 and 1921. In the 1920s, the numbers rose to almost a million (968,851). Silvestre (2005).

migrations increased considerably, the flows of a temporary nature also continued to grow[22].

As regards Spain's internal migratory flows, a further aspect that stands out alongside its intensity is its geographical distribution. Madrid and Barcelona consolidated their position during this period as the main areas of attraction, becoming the principal destinations for migrants. By 1930, they accounted for 45.97% of all those '*born in another province*' as registered in the Population Census, followed at a considerable distance by Sevilla (4.35%), Vizcaya (4.29%) and Valencia (3.07%)[23]. Furthermore, studies of the geographical patterns of internal migration in Spain highlight the fact that the poorest regions, including Andalusia, Extremadura and even Castilla-La-Mancha, contrary to expectations, recorded a small number of emigrants[24].

In this sense, and having evaluated the intensity of migratory flows within Spain, Rosés and Sánchez-Alonso (2004) claim, in line with Boyer and Hatton (1994), that an increase in labour market integration can occur even in a context of limited mobility of the labour factor, as in the Spanish case. These authors studied labour market integration in terms of the evolution shown by interregional wage differentials and not on the basis of worker mobility. In so doing, Rosés and Sánchez-Alonso (2004) focused on real wages in Spain for a sample of regions between 1850 and 1930 that included agricultural workers, unskilled urban workers and urban industrial workers. Based on the growth literature published in the 1990s, their results pointed towards the existence of β-convergence in regional real wages throughout the period of study with the sole exception of the 'shock' caused by World War I. Only in the years following World War I did wage differentials increase albeit that in the 1920s the convergence in regional real wages reappeared with considerable intensity. This evolution led, in turn, to a reduction in the dispersion of wage differentials measured in this case by the coefficient of variation, i.e., σ-convergence, and it was considered evidence in favour of the integration of the labour markets.

[22] Temporary migrations were made between agrarian areas on the basis of regional agricultural practices and different farming calendars, but some farm workers migrated to non-agricultural zones and sectors to undertake industrial activities, mining, construction work, transportation and services. Overall, between 1877 and 1930 temporary internal migration, registered as '*transeúntes*' (transients) in the Population Census, accounted for around 3% of the total population. Silvestre (2005).

[23] By this date, almost half the residents of Madrid claimed to have been born in another province (46.84%), in common with more than a third of the population of Barcelona (35.98%). The share was not as great in the provinces of Vizcaya (24.91%), Sevilla (15.24%) and Valencia (8.31%). Silvestre (2003), p. 78.

[24] The migrant workers in Barcelona came primarily from the Mediterranean provinces and the Ebro Valley (and Almería). By contrast, those in Madrid came from the two Castillas and some of the provinces in the north of the Peninsula, while the capacity of attraction of Sevilla, Bilbao and Valencia was limited to their adjoining provinces (Silvestre, 2005)

1.3.2. The first stages in the process of industrialisation in Spain

Together with the integration of the Spanish market, the second aspect to consider in relation to the predictions that can be derived from the theoretical models is the process of industrialisation, which in the case of Spain occurred in parallel with the creation of the domestic market. The onset of the Industrial Revolution in Great Britain at the end of the 18th century, and its gradual diffusion to an increasing number of countries, allowed the various economies that joined this process to enter onto the path of what Simon Kuznets defined as 'modern economic growth' (Kuznets, 1966, 1971). This process is characterised by high and self-sustained growth rates of per capita income, which are often accompanied by an increase in population and almost always by structural change. Thus, in short, the transfer of resources from low-productivity agrarian activities to high productivity industrial activities, a sector that was gradually adopting technological change, created the conditions for economic growth. However, the differences in the growth rates also generated an increase in income inequality across countries, and given the self-sustained nature of 'modern economic growth', the differences in income levels became more accentuated over time (Acemoglu, 2009).

Sidney Pollard (1981) has suggested that the processes of industrialisation were unique and non-repetitive, and stood out, in turn, for their marked regional character[25]. A good number of examples illustrate the regional nature of the industrialisation processes: the Lancashire in Britain, the Sambre and Meuse Valley and the Scheldt Valley in Belgium, the Ruhr region in Germany, the triangle Genoa-Milano-Torino in north Italy or New England and the rest of the north east area of the US, to mention just some. The Spanish case provides further evidence of this. The Spanish economy, lying in the geographical periphery of Europe, sought from the early decades of the 19th century to join this race to industrialisation in which most of the countries of continental Europe took part. The outcome of these attempts in the 19th century (normally extended up to the outbreak of World War I) saw Spain trailing behind the leading European countries; in other words, the failure of the first Industrial Revolution in Spain, as stated by Nadal (1975). On balance, the Spanish economy had not witnessed the profound transformations that industrialisation implies.

[25] Unlike the assumption made by "*Gerschenkron, Kuznets, and others, that countries within their political boundaries are the only units within which it is worthwhile to consider the process of industrialisation*". Pollard (1981), p. vii.

Yet, and linked to the regional nature of the processes of industrialisation stressed by Pollard (1981), two exceptions emerged within this general view of economic backwardness: Catalonia and the Basque Country[26]. Both regions achieved a considerable degree of industrial development, even in comparison with the rest of the main industrial regions in Europe, with a high specialisation in two of the sectors that had leaded the Industrial Revolution in Great Britain, namely cotton and iron. In Catalonia, the cotton industry, with a tradition that stretched back to the 18th century, gradually became mechanised in the 19th century, so that by the end of the century the cotton industry, and by extension that of textiles, was concentrated almost exclusively in Catalonia. It was during those years that Catalonia became 'Spain's factory'. In the Basque Country, the iron and steel industry underwent rapid growth in the last quarter of the century, exploiting its proximity to the sources of iron ore minerals that supplied the factories in Vizcaya and the advantages of the non-phosphoric nature of these ores.

The share of the active industrial population offers an initial picture of the spatial pattern of Spain's process of industrialisation. According to the statistics reported in the 1860 Population Census, the industrial population in the province of Barcelona represented 31.5% of its total active population, a figure that was much greater than the Spanish mean, which stood at 12.5%. At this time, Barcelona was followed by Madrid with an industrial population that reached 21.1%, while the Basque Country as a whole presented more modest figures (14.3%). Forty years later, in 1900, the proportion of the population dedicated to industrial activities in Barcelona had increased to 35.5%, a figure that more than doubled the percentage of industrial population in the whole of Spain, which stood at around 16%[27]. In the Basque Country, the two provinces that presented the greatest industrial development, Guipúzcoa and Vizcaya, had industrial populations of 31.0% and 27.04% at this date, respectively. Meanwhile, Madrid had suffered something of a stagnation in the percentage of its population dedicated to industry in the second half of the 19th century (20.7%). Finally, in 1930 the industrial population of Barcelona once again recorded marked growth with respect to figures at the beginning of the 20th century, reaching a percentage of 58.6%. The Basque provinces of Guipúzcoa and Vizcaya had figures around 38%, while Madrid had also grown to 31%. All these provinces exceeded the Spanish mean, which despite the changes made in the distribution

[26] The evolution in the regional economies has been described in Nadal and Carreras (1990) and Germán, Llopis, Maluquer and Zapata (2001).
[27] The total figures for Spain in 1900 and 1930 come from Nicolau (2005).

by sector of its active population after 1910, recorded a figure of around 26.5% of its active population dedicated to industrial activities in 1930.

The overall result of this evolution was the gradual concentration of industrial activity in Spain from the second half of the 19th century onwards. This development is well documented by Spanish historians. Nadal (1987) studied the regional distribution of industry and its evolution between 1856 and 1900 using a fiscal source: the *Estadísticas Administrativas de la Contribución Industrial y de Comercio*[28]. The picture that emerges from the figures reported by Nadal highlights the relative weight of Catalonia and Andalusia in the middle of the 19th century, two regions that stand out in terms of the intensity of their industrial activity. In 1900, the predominance of Catalonia (and the Basque Country) was overwhelming, and among the other regions, only Valencia recorded figures greater than one in the industrial intensity index[29].

Tirado, Pons and Paluzie (2006) have conducted an in-depth analysis of regional specialisation and the patterns of localisation of Spain's industry in the second half of the 19th century and early decades of the 20th. Using the same sources as those employed by Nadal (1987) as well as the tax data provided by Betrán (1999) for the interwar years, they showed that in parallel to the integration of the Spanish market during the second half of the 1800s there occurred an increase in regional specialisation. Yet, in the interwar years this process came to a halt and the differences between the industrial productive structures of Spain's provinces did not register a further increase.

As for the geographical concentration of manufacturing activities, from the middle of the 19th century to the outbreak of the Civil War, industrial production gradually concentrated in a small number of provinces. In an initial period between 1856 and 1893 the inland provinces and those of Andalusia lost strength while an industrial axis emerged in the Mediterranean area. This process highlights the gradual shift of industrial activity to the coastal provinces of the geographical periphery of Spain. By contrast, in the interior of the peninsula, with the exception of Madrid, industrial development was scarce[30]. This localisation pattern, however, varied in the interwar years when certain inland provinces

[28] See also Nadal and Carreras (1990), Parejo (2001). An alternative approach to the study of regional patterns of industrialisation in historical perspective is based on the construction of industrial production indices, which are available for Catalonia (Carreras, 1990b; Maluquer, 1994), the Basque Country (Carreras, 1990b), Andalusia (Parejo, 1997) and Valencia (Martínez-Galarraga, 2009). For a comparative view of these regional indices see Parejo (2004).

[29] The industrial intensity index is defined as the ratio between each territorial unit's share of the country's total industrial output and each territorial unit's share of the country's total population. The Basque Country and Navarre were not included in this fiscal source since they had their own tax system.

[30] Sánchez-Albornoz (1987).

(Madrid and Zaragoza) and provinces in the north strengthened their industrial activity in a period in which the Mediterranean axis appears to have weakened.

The gradual concentration of the country's industry in a small number of areas, which is a well-documented fact between 1856 and 1929, did not stop at the end of that period. Paluzie, Pons and Tirado (2004) have completed the long-term overview of the geographical distribution of Spanish industry drawing on various sources and indicators. Their results show that between 1955 and 1975 the spatial concentration continued to increase slightly, albeit that since then there has been a clear tendency towards the geographical dispersion of industrial activity. Therefore, in the long term the spatial distribution of industry in Spain presented a bell-shaped evolution, with an initial phase characterised by an increase in industrial concentration and a shift in tendency since the 1970s in which a broad spatial dispersion of industry is observed[31].

On the one hand, this evolution coincides with the dynamics that can be derived from the models of New Economic Geography, in which a non-monotonic relationship exists between the fall in transport costs and the spatial concentration of activity. Further, it is similar to that which emerges in the empirical studies, including Williamson (1965), although this study is centred on regional inequality measured in terms of per capita income. This result reinforces the idea of the relevance of the processes of industrialisation as regional differences are defined. On the other hand, the Spanish case shows a dynamic that is similar to that recorded in other countries for which studies are available of the long-term evolution in the spatial distribution of manufacturing, such as the United States (Kim, 1995) and France (Combes, Lafourcade, Thisse and Toutain, 2008). Finally, it is worth stressing that in the Spanish case, entry into the downward part of the inverted 'U-shaped' curve in the 1970s occurred later than in other countries, such as the United States, where it commenced in the 1930s and 1940s.

1.4. Objectives and the development of the thesis

This thesis is an attempt to achieve a deeper understanding of the long-term evolution of regional inequality in Spain and its determinants during the first stages of economic development. This period is characterised by the difficulties experienced by the Spanish economy to converge with the core European countries. According to Maddison's database, in 1850 the Spanish GDP per capita was 60% of the average of

[31] In chapter 4, section 4.2, a more detailed description of the regional patterns of industrialisation in Spain can be found.

some of the main economic powers in Europe: France, Germany and the UK. By the eve of World War I, this percentage had fallen to 51%, and, in 1935, prior to the break out of the Spanish Civil War (1936-1939) the Spanish relative GDP per capita had increased to 55%, still below the levels recorded in the mid-19th century.

In this period, the process of industrialisation gradually spread across Europe, and structural change, i.e., the reallocation of resources from low productivity activities in the agrarian sector towards a modern industrial sector, was crucial for entering onto the path of 'modern economic growth'. However, Spain failed to fully develop the Industrial Revolution (Nadal, 1975) as the '*first comers*' in Europe did, and industrialisation only arrived in a small number of regions. At the same time, it was during these years that the Spanish market became finally integrated. The application of the industrial advances to the transport system resulted in a fall in transport costs and together with the institutional reforms implemented by the liberal governments in the 19th century, these events favoured the integration of the domestic market. In this context, it has been argued that the increasing concentration of manufacturing activities in a few territories generated an increase in regional inequality within Spain from the mid-19th century onwards, although some disparities already existed at the end of the Modern period[32].

The first shortcoming in the study of the long term evolution of regional inequality in Spain is the lack of reliable information before 1930. The only available estimates of regional GDP per capita before that date are those of Álvarez Llano (1986) for the Spanish Autonomous Communities (NUTS2 according to the Eurostat territorial division), starting in 1800. However, the methodology followed in the construction of these estimates is not made explicit by the author and therefore researchers have been reluctant to fully exploit this database. A first contribution of this thesis is the construction of a new GDP database for Spanish provinces (NUTS3) in different benchmark years between 1860 and 1930 following the standard methodology developed by Geary and Stark (2002). The first results confirm the increase of regional inequality from 1860 onwards, although this tendency was reversed in the first decades of the 20th century, and a period of convergence started until the arrival of the Spanish Civil War. This U-inverted shape in the relationship between economic development and regional disparities in Spain fits quite well with the evolution described by Williamson (1965) for

[32] Llopis (2001).

an international sample of countries at different stages of development in the 19th and 20th century.

The emergence of the New Economic Geography (NEG) in the last decades (Krugman, 1991) provides a valuable analytical framework for the study of the localisation of economic activity in space and therefore for the study of regional inequality and its evolution over time. NEG models are based on a number of alternative assumptions to those used in the literature that adopts a more neoclassical approach. Taking into account the presence of increasing returns and transport costs, NEG models stress the existence of a circular, cumulative process in which initial advantages of a location can be strengthened over time. Hence, NEG highlights the relevance of understanding the historical processes that have shaped the spatial distribution of economic activities and it appears to be a particularly appropriate theoretical framework for undertaking historical studies.

In short, NEG models suggest that the relationship between market integration and regional inequality follows a bell-shaped curve, like in Williamson (1965). In the first stages of the process, agglomeration forces make economic activity concentrate in a limited number of locations but as integration proceeds, economic activity becomes more disperse across space and a pattern of convergence is thus expected. But where will production take place? Krugman states that "*...in a world characterized both by increasing returns and by transportation costs, there will obviously be an incentive to concentrate production of a good near its largest market, even if there is some demand for the good elsewhere. The reason is simply that by concentrating production in one place, one can realize the scale economies, while by locating near the larger market, one minimizes transportation costs*"[33]. Therefore, large markets will be more attractive for firms and workers, and accessibility or market potential becomes an essential variable within NEG analyses.

In this framework, most of the empirical research has focused on the industrial sector, where economies of scale, a key feature of NEG models, tend to be more present. The Spanish historical experience is a good example. The evolution of the spatial distribution of industry in Spain and its remarkable concentration from the beginning of the process of industrialisation has been analysed from a NEG perspective. Rosés (2003), Tirado, Pons and Paluzie (2002) and Betrán (1999) have offered evidence that the

[33] Krugman (1980), p. 955.

economic forces stressed in the NEG models were in operation in the industrial sector in Spain in the mid-19th century, increasingly during the second half of the 19th century, and in the interwar years, respectively.

As regards the location of industry across space at an international level, a significant contribution in this field was made by Midelfart-Knarvik, Overman, Redding and Venables (2002). These authors carried out an empirical exercise to test whether NEG forces played a role in the distribution of industrial activities within the European Union in recent decades. The results confirmed that the interactions between market potential and economies of scale and between market potential and intermediate goods were driving the location of industry during the European process of integration. This approach has been applied to different historical cases like interwar Poland (Wolf, 2007), Victorian Britain (Crafts and Mulatu, 2005, 2006) and the US at the time that the manufacturing belt emerged in the north east of the country (Klein and Crafts, 2009). The results showed that NEG-type mechanisms (as well as comparative advantage) were at work during the process of industrialisation in these countries, although with different intensity and a changing relative strength over time.

This type of exercise can therefore be applied to the Spanish industrial sector to gain a deeper and sounder knowledge of the forces shaping the industrial map of Spain. Furthermore, the evidence available for other countries allows comparing the different experiences in order to obtain a broader picture of the determinants of industrial location at an international level. When compared to the previous studies conducted for the Spanish industry in the period under study, the Midelfart-Knarvik, Overman, Redding and Venables (2002) strategy has some noteworthy advantages. First, it is based on a theoretical NEG model. Second, it allows discriminating the relative strength of Heckscher-Ohlin forces and NEG-type mechanisms. Third, a good number of variables considering region and industry characteristics are included in the equations. And fourth, the relevance of market potential, as stated by NEG models, can be tested directly.

In this context, the availability of an indicator of market accessibility is essential for the empirical research. As mentioned above, when economies of scale, monopolistic competition and transport costs are considered, NEG models conclude that production will tend to take place and concentrate in high market potential locations. Hence, another contribution of this thesis is the construction of market potential estimates for the Spanish provinces between the 1860s and 1930. The production of these estimates is based on Harris' (1954) market potential equation, where the potential of a province is measured as a weighted sum of the economic size of other provinces and overseas markets, and the weights are a decreasing function of distance or transport costs. This methodology has recently been used in a good number of NEG empirical papers and also in historical perspective for Britain (Crafts, 2005b) and the Austro-Hungarian Empire (Schulze, 2007). In this regard, the work by Crafts has become the reference in economic history for the construction of market potential estimates. In addition, albeit the Harris equation is an *ad hoc* measure developed by geographers, it is possible to establish a close relationship between this indicator and the measures of market accessibility derived from the NEG models.

The availability of an indicator of market accessibility at a provincial level in Spain becomes therefore a key tool to improve our knowledge of the determinants of the spatial location of economic activity and regional inequality. First, it allows implementing the Midelfart-Knarvik, Overman, Redding and Venables (2002) empirical strategy in order to confirm or reject the relevance of the NEG forces in shaping the location of industry in Spain in the period under study, bearing in mind the results obtained in previous exercises for the Spanish economy. Once this aim had been fulfilled, the next step will be to analyse whether the mechanisms stressed in NEG models had an impact not only on the industrial sector, but when aggregate income per capita is considered. Can geography explain the different growth rates in provincial GDP per capita in the first stages of economic development in Spain?

At an international level, Redding and Venables (2004) concluded in a seminal work that market potential explains a good share of the present cross-country inequality in terms of GDP per capita. In other words, the big differences observed in income per capita across the world can be partially explained by the difference in the market accessibility of the countries. Being close to rich countries makes it very unlikely that a country will be poor, while being far away from large markets and prosperous countries is a considerable disadvantage for economic development. The results obtained by these authors were subsequently confirmed by Mayer (2008) for a sample of countries that covered the second half of the 20th century. Yet, the study of the role of economic geography forces as drivers of income inequality within countries has received less attention. Most of the research has focused on verifying the Krugman (1991) wage equation. Here, the existence of higher wages in high market potential regions and the decreasing wages with distance has been confirmed in a large number of countries. As for the Spanish case, the existence of a wage gradient has been demonstrated for present times but also for the 1920s (Tirado, Pons and Paluzie, 2003, 2006, 2009)[34].

However, a recent paper by Ottaviano and Pinelli (2006) constitutes a first attempt to test whether geography and market potential played a role in explaining the differential growth rates in income per capita at a regional level. Their exercise examined the experience of Finland between the end of the 1970s and the 1990s. The collapse of the neighbouring Soviet Union brought major changes to the Finnish pattern of trade, and had an asymmetric impact on the different regions of Finland. Therefore, these authors explored the significance of market potential before and after the shock that the fall of the Soviet block represented for the Finnish economy. This empirical strategy opens up new possibilities to examine the impact of economic geography on regional income per capita growth rates. When applied to the Spanish case, the new database constructed for provincial GDP and market potential allows analysing the causes behind economic growth at a regional level and, for the first time, investigating the sources of regional inequality in Spain during the second half of the 19th century and the first decades of the 20th century.

This brief introduction contains the general objectives of this thesis. The process of industrialisation and the integration of the Spanish domestic market between the second half of the 19th century and the first decades of the 20th century generated a notable and well documented increase in the spatial concentration of industry in Spain, and the emergence of the persistent regional disparities that still today are characteristic of the Spanish economy. The New Economic Geography theoretical models can shed light on the forces behind the spatial concentration of economic activity in a context characterised by falling transport costs and the increasing presence of economies of scale. Thus, within this analytical framework, locations with good access to markets attract both firms (capital) and workers (labour) leading to a cumulative process of agglomeration, and therefore, to the increase in regional inequality in the first stages of development.

In this general context, the thesis is organised as follows. In chapter 2, a survey of the New Economic Geography literature is conducted beginning with a review of the theoretical literature. First, some background is provided since the study of the role of geography in economy precedes the emergence of the New Economic Geography in the

[34] In addition, a positive relationship between market potential and migration decisions by workers in the 1920s was also found by Pons, Paluzie, Silvestre and Tirado (2007). In turn, Martínez-Galarraga, Paluzie, Pons and Tirado (2008) showed the existence of an 'agglomeration effect' in the industrial sector in line with Ciccone and Hall (1996) and Ciccone (2002).

1990s. The work undertaken by geographers is briefly introduced and then, the review focuses on the main features and the evolution of international trade theory until the departure of New Economic Geography from New Trade Theory to consolidate as an independent field. Then, the main theoretical contributions of the NEG models and the implications in terms of market accessibility and regional inequality are introduced. Next, the empirical exercises carried out within this analytical framework are presented, paying special attention to those conducted from an economic history perspective, where the Spanish case stands out for the abundant empirical research available.

This literature review highlights the key role played by market potential in the location decisions by agents. Thus, in chapter 3, a measure of market accessibility for the Spanish provinces is constructed based on the Harris' (1954) market potential equation, following the work by Crafts (2005b). In so doing, data on provincial GDP in the period under study is needed and thus, a new GDP dataset has been assembled following the proposal of Geary and Stark (2002). Then, the methodology applied in the calculation of market potential is fully detailed and the main results presented. Finally, market potential becomes a necessary empirical tool for the exercises proposed in the next chapters.

In chapter 4, the model developed by Midelfart-Knarvik, Overman, Redding and Venables (2002) is applied to the Spanish provinces between 1856 and 1929. This exercise combines both comparative advantage as stated in the Heckscher-Ohlin theorem and the mechanisms stressed by NEG models as determinants of industrial location. Did NEG forces play a role in the remarkable increase in the spatial concentration of industry in the first stages of the industrialisation process in Spain? The results may be useful to confirm or re-interpret the results reached in previous studies for the Spanish industry in this historical period and may also help to complete the picture available at an international level.

Next, in chapter 5, the focus moves from the industrial sector to overall regional inequality in terms of GDP per capita. Here, the empirical strategy suggested by Ottaviano and Pinelli (2006), where growth regressions are derived from a NEG model, is applied to the Spanish case in order to examine whether geography had an impact on provincial GDP per capita growth rates. Does geography matters in explaining the increase in regional inequality recorded in Spain during the period under study? Growth regressions allow exploring the proximate causes of growth as well as wider influences on economic growth considering a subset of explanatory variables where geography variables are included. Finally, the thesis closes with a final chapter where the main conclusions of the research are summarised.

Chapter 2. New Economic Geography (NEG): theory and empirics

2.1. Description of the theoretical framework: New Economic Geography

2.1.1. Background

The uneven spatial distribution of economic activities is one of the main characteristics of economic development, not only across countries but also within them, with such activities tending to concentrate in certain areas. Yet, for a long time the spatial component has been virtually absent from the main lines of economic research. Despite important contributions to this field[35], the infrequent inclusion of the spatial element has been the result of the difficulty in integrating space within economic theory, a fact that made economic geography virtually intractable. The main problem came from the need to incorporate increasing returns, considered key to explaining international trade, and regional and urban development, within theoretical models[36]. The underlying reason is that on assuming the existence of increasing returns at the firm level, the assumption of perfect competition becomes unsustainable. Therefore, the difficulties lay in the inappropriateness of increasing returns to the predominant paradigm of competition, in which the market structure was characterised by constant returns to scale.

The neoclassical theory of international trade, one of the areas in which some consideration has been given to the spatial component, has explained the existence of

[35] Some exceptions to this general trend include a number of classical studies from various fields such as urban economics (Von Thünen, 1826; Alonso, 1964), regional science (Christaller, 1933; Lösch, 1940), industrial location (Weber, 1909; Hotelling, 1929; Kaldor, 1935), and geography (Harris, 1954). These issues are dealt with in Fujita, Krugman and Venables (1999) and Combes, Mayer and Thisse (2008), whose work is followed in this section.

[36] "*…increasing returns as a cause of trade has received relatively little attention from formal trade theory. The main reason for this neglect seems to be that it has appeared difficult to deal with the implications of increasing returns for market structure*". Krugman (1979), p. 469.

flows of goods between countries and international specialisation based on the assumptions that markets operate in perfect competition, the existence of constant returns to scale and identical and homothetic preferences between countries. Broadly speaking, this theory suggests that countries will specialise in the production of a good according to their comparative advantage, a comparative advantage that is linked to the fact that the space is considered heterogeneous. In Ricardo's model, trade results from cross-country differences in technology and productivity: each country specialises in the production of the good for which its opportunity cost of production is lowest, and in which its production is, therefore, most efficient. The Heckscher-Ohlin theorem, moreover, assumes a model comprising two countries, two sectors and two goods, in which the countries enjoy access to the same technology, but differ in their production factor endowments. In this case, with labour and capital as the production factors, in a situation of free trade, the country in which the labour factor is relatively abundant (i.e., the country that enjoys a relative advantage in its endowment of this factor) will specialise in the production of the labour-intensive good, while the opposite will occur in the country in which the capital factor is relatively more abundant. The trading outcome will, therefore, be in accordance with the Factor Price Equalization theorem (FPE)[37].

The neoclassical theory of international trade is based on the assumption of factor immobility, an assumption that might fit well with reality when explaining cross-country trade, but which is more difficult to accept in the study of the trade patterns between the regions of the same country, which are characterised by a greater mobility of the labour factor[38]. Furthermore, the Heckscher-Ohlin theorem could explain the appearance of inter-industrial trade between countries that differ notably in their underlying characteristics, but not the new trade patterns that took shape after World War II, characterised by growing intra-industrial trade, i.e., involving relatively similar goods and countries (Grubel and Lloyd, 1975)[39]. In addition, although spatial heterogeneity might be an important explanatory element, it is difficult to imagine that these differences alone

[37] This result would occur in highly limited conditions. "*These include the requirement that countries have identical technologies, and that there are at least as many traded activities (goods or mobile factors) as there are immobile factors*". Venables (2005), p. 3.

[38] Here, we should bear in mind the spatial impossibility theorem (Starrett, 1978), which states that whenever agents are mobile, a competitive equilibrium in which the regions trade does not exist, because factor mobility and interregional trade are incompatible in a neoclassical world. "*If space is homogeneous, transport is costly, and preferences are locally nonsatiated, then there exists no competitive equilibrium involving the transport of goods between locations*". Combes, Mayer and Thisse (2008), p. 38.

[39] Nonetheless, some authors claim that intra-industry trade was already important before that date, for example in textiles trade prior to World War I, like in Brown (1995).

can explain the formation of large agglomerations and the origin of the substantial regional inequalities that are observed in reality.

These limitations in explaining the new patterns of international trade and the difficulties in modelling increasing returns began to be overcome with the emergence of new analytical tools in the 1970s. Here, the monopolistic competition model designed by Dixit and Stiglitz (1977), developed within industrial organisation, allowed increasing returns and imperfect competition to be integrated in the theoretical models. This general equilibrium model, which included Chamberlin's (1933) concept of monopolistic competition, and in which there exist a large number of firms usually represented as a *continuum*, allows two seemingly incompatible goals to be reconciled: a firm can be insignificant relative to the economy as a whole and in turn can enjoy a monopoly in its own market. Therefore, increasing returns and imperfect competition are combined.

The application of the Dixit-Stiglitz model (1977) to international trade allowed the difficulties involved in integrating increasing returns to scale to be overcome and, thus, from the end of the 1970s onwards it became the foundation on which New Trade Theory was subsequently developed. In the models developed within New Trade Theory (Krugman, 1980; Helpman and Krugman, 1985), the economy consists of two regions (A and B) and two sectors: agriculture and manufacturing. On the one hand, the agricultural sector produces a homogeneous good under constant returns to scale while the good is sold in a market that operates in perfect competition. Moreover, the trading costs of the agricultural good are considered to be zero, so that its price is the same in both regions, as are agricultural wages. On the other hand, the manufacturing sector produces a differentiated good, under increasing returns in a framework of imperfect competition[40]. The varieties produced by domestic and foreign firms differ, although they have the same elasticity of substitution. Likewise, the consumption of foreign varieties involves a trade cost (τ). This is an iceberg-type transport cost, such as that proposed by Samuelson (1954), where a fraction of the good transported is lost on route[41].

A comparison with the neoclassical theory of international trade shows that here the comparative advantage has been eliminated. The technology between regions is identical and by considering the space to be homogenous, the relative factor endowments

[40] Given that the varieties exchanged between the two regions belong to the same sector due to the product differentiation, New Trade Theory includes the possibility that intra-industrial trade exists, which is in fact one of its main goals.

[41] Therefore, for each manufactured unit transported from one region to another, only a fraction $\tau<1$ reaches its destination. This type of transport is often illustrated with the example of a horse that consumes part of the merchandise during transit.

are also the same in both regions[42]. Furthermore, New Trade Theory assumes not only goods to be mobile, but also the capital factor, so that firms' location decisions become endogenous.

The assumption of capital mobility allows studying the impact of the size of each region on the spatial distribution of firms, since this becomes one of the main determinants in each firm's location decision. These decisions are the result of a system of centripetal (proximity to markets) and centrifugal forces (inter-firm competition). In the case of the former, there is a consensus that a large market will tend to increase the profits of the firms established in it, so that a location with good access to demand is preferred (*market access effect*). In the case of the latter, the increase in the number of firms in the larger market will lead to an increase in competition so that the firms have an incentive to separate geographically from each other so as to relax this competition (*crowding-market effect*)[43]. The balance between these two forces determines the agglomeration in the manufacturing sector.

In this case, when the location of firms is endogenous (mobile capital) the liberalisation of trade leads to an increase in regional disparities. This new result is linked to the 'home market effect'. Given that locations with a larger market are the preferred choice of firms, as the market increases in size, the larger region will attract (in the sector with increasing returns) a number of firms that is more than proportional to their relative size differences. Here the comparative advantage lies in the relative size of a region, and the 'home market effect' magnifies this initial advantage. Thus, economic integration will strengthen the concentration of manufacturing in the larger region, in which firms will tend to locate in order to exploit more fully scale economies. By contrast, the smaller region will export capital to the larger region, thereby suffering a process of de-industrialisation. The outcome is an increase in regional inequality. The fall in transport costs therefore lies at the root of the growing divergence between regions.

New Trade Theory provided an explanation for the existence of intra-industrial trade between countries, as well as for the differences observed at an international level in the uneven spatial distribution of economic activities. Yet, certain limitations continued to exist. On the one hand, it is assumed an initial regional asymmetry as regards market size

[42] "*When two imperfectly competitive economies of this kind are allowed to trade, increasing returns produce trade and gains from trade even if the economies have identical tastes, technology, and factor endowments*". Krugman (1980), p. 950.

[43] The sale of differentiated goods leads to lower levels of competition between firms, but it continues to exist.

so that the 'home market effect' appears. Further, although firms' location decisions have become endogenous on assuming capital mobility, the labour factor continues to be immobile. And while this assumption may be acceptable internationally where obstacles to migration are great, it is more difficult to accept in the study of inequalities within the boundaries of a country. These limitations first began to be overcome in the work of Krugman (1991), in which, by assuming labour factor mobility, both the firms' decisions and those of the workers became endogenous, being the forces that lead to agglomeration. This seminal paper marked the emergence of the so-called 'New Economic Geography' (NEG), which is surveyed in the following section.

2.1.2. 'New Economic Geography': economic integration, market access and regional inequality

'New Economic Geography' is concerned with the study of the uneven spatial distribution of human activity and the existence of regional disparities in income levels. In the models of New Economic Geography, transport costs and increasing returns interact in a framework of monopolistic competition favouring the spatial agglomeration of economic activities, and strengthening them once they are in motion. In this context, transport costs and the gradual market integration of goods and factors play a key role, since a reduction in transport costs can favour the spatial concentration of economic activities and, as a result, increase regional inequalities. Thus, analysing the impact of economic integration on spatial inequality is one of the main goals of the New Economic Geography. In this case, transport costs serve as a proxy for economic integration, so that a reduction in these costs means a greater degree of integration.

In this theoretical framework, the spatial distribution of economic activity depends on the interaction of two types of forces, which operate in opposite directions: the *centripetal* or *agglomeration forces* and the *centrifugal* or *dispersion forces*. The model developed by Krugman (1991) illustrates a similar cumulative process to that envisaged by Hirschman (1958) and the cumulative causation described by Myrdal (1957), in which the concentration of economic activity occurs as a result of the interaction of two centripetal forces linked to market access. In turn, agglomeration is subject to a snowball effect that results in a continuous strengthening of this spatial concentration once it is set in motion.

In accounting for this process, Krugman (1991) extended the New Trade Theory models (in which it is assumed that labour is homogeneous and mobile between sectors but not between countries), by considering two regions in which the immobile factor (farmers) is used as an input in the agricultural sector; the second factor (workers) is mobile and is used as an input in the manufacturing sector. Therefore, the labour factor is divided between unskilled farm workers (immobile) and skilled manufacturing workers (mobile).

In the New Trade Theory models, the capital owners repatriated their income and consumed in their region of residence, but even so, the largest market attracted a more than proportional share of manufacturing. However, when skilled manufacturing workers are mobile, individuals live and work in the same region so that production and consumption also take place in the same region (that of destination). Therefore, migration modifies the relative size of the markets and the regional distribution of demand changes with the distribution of manufacturing skilled workers, which is now endogenous.

In Krugman's (1991) core-periphery model two main effects linked to the factors of production operate: one related to the firms and the other to the workers. In order to study the location decisions of firms and workers, it is assumed that one region becomes slightly larger than another, thereby increasing its number of consumers. First, this increase in one of the region's market size leads to an increase in its demand for manufactured goods, so that for firms it becomes more convenient to locate near the location with higher demand so as to save on transport costs. This means that activities with scale economies concentrate in locations that enjoy good market access (backward linkages). As a result, the 'home market effect' introduced above ensures that this increase in market size generates a more than proportional increase in the number of firms in that location, pushing up nominal wages. Second, the presence of more firms means a greater variety of locally produced goods, where consumption incurs lower transport costs. A lower local price index and the consequent increase in real wages in the region attract new flows of workers to the large urban and industrial centres (forward linkages).

These two centripetal forces are mutually strengthened favouring agglomeration, and the proximity to the large markets stands out here as one of the main mechanisms, since producers and workers, *ceteris paribus*, prefer locations that have good access to demand. Therefore, market access establishes itself as a key element in NEG analyses since it has a positive influence on the location decisions of firms and workers alike, and it induces factor mobility: that of capital in the case of backward linkages and that of labour in forward linkages.

In this case, the result of economic integration is the emergence of a core-periphery geographical pattern. When transport costs are high, trade is so expensive that firms sell their products in the local market. As a result, a symmetric pattern emerges in which the firms are spatially dispersed and the manufacturing sector is distributed evenly between the regions, which have the same nominal wages and price indices[44]. However, when transport costs are low enough, a shift is made to an asymmetrical equilibrium, characterised by agglomeration. Thus, economic integration gives rise to a geographical concentration of economic activities. This results from worker mobility, which allows for the appearance of a cumulative causation that strengthens the agglomeration by increasing the market size advantage. The greater demand generated in the core region means that all the firms in the manufacturing sector, in which the increasing returns operate, locate in the same region, leading at the same time to de-industrialisation in the periphery. In other words, the economic integration generates an abrupt transition from dispersion to agglomeration.

The shift to a core-periphery structure leads to an increase in regional inequalities. Hence, Krugman (1991) offers theoretical support for the substantial and persistent territorial inequalities that can be seen in the real world. Furthermore, in this case and unlike the international trade theories, regions that initially present similar characteristics end up diverging considerably, since even a small transitory *shock* can give rise to permanent regional imbalances[45].

Finally, Krugman (1991) emphasises the pecuniary externalities as opposed to the technological. When firms and workers move from one region to another, this unintentionally affects the welfare of all agents. The shift from a disperse structure to one of agglomeration is caused by microeconomic decisions, where agglomeration is the involuntary consequence of the aggregation of a large number of individual decisions. As a result, the agglomeration and the consequent increase in inequality have to be considered as an economic factor.

In this model agglomeration lies in the mobility of the labour factor. Yet, one limitation that has been identified in Krugman's (1991) model is that agglomeration is

[44] Together with the '*crowding-market effect*', a further force could lead to the dispersion of manufacturing activities: as unskilled farm workers are immobile, a proportion of them will be located in the periphery and the demand of these workers for manufactured goods workers has to be satisfied.

[45] By assuming that the regions are symmetrical, NEG does not take into consideration primary geographical elements, so that the theory does not establish which region will become the industrialised core and which the periphery.

also present in areas characterised by a low spatial mobility of labour, both internationally and within countries. Later developments within NEG have completed these results.

The work of Krugman and Venables (1995) and Venables (1996) allows understanding the emergence of large industrial regions in economies characterised by low labour mobility, assuming that the labour factor is immobile. Furthermore, these studies have the virtue of adding to the analysis a key element that was not included in Krugman's (1991) pioneering study: the existence of intermediate goods. In this case, firms produce differentiated varieties incorporating labour and intermediate goods produced by other firms. Finally, labour is now homogeneous (no distinction is made between skilled and unskilled workers), and as there are no intersectoral mobility costs, the workers can be contracted in either of the two sectors. The other assumptions are the same as those made by Krugman (1991).

The consideration of the existence of intermediate goods provides a better fit to real patterns and implies that when taking their decisions, producers of intermediate goods prefer to locate in those places where the final goods are produced. Likewise, producers of final goods show a tendency to locate where the suppliers of intermediate goods are placed. This reciprocal influence captures the Marshallian externality related to the availability of specialised intermediate inputs, which Marshall (1890) considered a fundamental element for the existence of industrial clusters[46].

Now, when firms concentrate in a region, the high demand for intermediate goods attracts the producers of these types of goods. In addition, the lower price index of the regions that produce a greater number of varieties leads to a fall in the production costs of firms in the manufacturing sector. As a result, intermediate goods are supplied at a lower price in the core region, which leads to more producers of final goods moving to the core. Thus, the producers have an incentive to locate in the region that contains the highest number of varieties, since they will benefit from lower production costs, resulting in agglomeration. Moreover, the higher nominal wage in the region in which manufacturing concentrate generates an increase in final demand, becoming once again an agglomeration force, albeit that in this case the increase in demand comes from the increase in the wages of the workers (which are immobile) without there being an increase in this case in population as in Krugman (1991).

[46] Along with this externality, Marshall (1890) noted a further two: informational spillovers and the formation of a skilled work market.

Therefore, Krugman and Venables (1995) and Venables (1996) show an alternative mechanism that can bolster agglomeration when there is no labour mobility: the presence of input-output linkages. If the production of intermediate goods represents an important proportion of industrial output, firms will have an incentive to locate near their suppliers and their consumers, which can favour agglomeration in a given region. If up to this point the agglomeration occurred endogenously because of the size of the local markets and was caused by the mobility of consumers/workers, the presence of input-output linkages in the industry leads to the emergence of new forces that play a relevant role when shaping the spatial pattern of economic activity.

Among these new forces we find not only those that tend to favour agglomeration, but also centrifugal or dispersion forces. On the one hand, there exists greater competition in the manufacturing sector of the core region derived from the greater number of firms resulting from the agglomeration (*crowding-market effect*). On the other hand, the dispersion force linked to the increase in the level of nominal wages in this region and the consequent increase in labour costs has to be added. Finally, given that the workforce is immobile, it is necessary to take into account that the demand for manufactured goods in the periphery remains substantial. Together, these factors can lead to the relocation of industry from the core to the periphery, where lower wage costs can offset the lower demand for the firm's goods. In this case, by choosing the periphery, a producer will face less competition, as there are fewer firms located in the region, and lower wage costs. By contrast, the firm will confront a weak final demand because of the workers' lower purchasing power, a low demand for intermediate goods and, as a result, higher costs in the acquisition of intermediate inputs, since transport costs affect a greater fraction of the varieties used as inputs.

By including these new forces in the analysis, the relation between economic integration and the spatial concentration of manufacturing is no longer monotonic and shows a bell-shaped evolution. If in Krugman's (1991) model the fall in transport costs led to the emergence of a core-periphery pattern that exacerbated regional inequality, in this case the pattern is different. When transport costs are high, a symmetric equilibrium is recorded in which manufacturing is distributed equally between the two regions, without there being any spatial inequality. When the transport costs fall the symmetric equilibrium is broken and a core-periphery structure like that described by Krugman (1991) appears. However, specialisation in industry in the core will only occur in the case in which the share of the manufactured good in the final consumption is high. As a result of the high demand for the manufactured good the agglomeration forces cause the regions to diverge. Yet, this asymmetric equilibrium is no longer stable when transport costs reach a sufficiently low value. In this case, the dispersion forces bring the agglomeration process to a halt, or even invert it, resulting in the reindustrialisation of the periphery and the simultaneous deindustrialisation of the centre. Consequently, the economic integration initially generates an increase in regional disparities but as the integration proceeds, this inequality tends to disappear introducing a period or regional convergence.

Puga's (1999) study confirms this result according to which the relationship between the regional integration process and the degree of concentration of activity in the territory can describe a non-monotonic bell-shaped evolution. The author combines the two earlier cases by assuming interregional labour mobility (Krugman, 1991) and input-output linkages (Krugman and Venables, 1995; Venables, 1996), considering also the presence of intersectoral mobility. It is, therefore, a more appropriate framework for studying regional inequality.

In the long-term equilibrium, when there is labour mobility, the results of Puga (1999), by incorporating input-output linkages and intersectoral migration (while enabling to understand the determinants of economic agglomeration) provide a similar pattern to the one described in Krugman (1991), i.e., a pattern characterised by initial dispersion and a subsequent concentration of economic activity. However, when labour mobility does not exist, the bell-shaped evolution between economic integration and regional inequality is confirmed. In this case, when transport costs are high, manufacturing is dispersed across the regions. Yet, with the fall in transport costs firms can decide to locate in those sites with a larger market, where they can also take advantage of the possibility of locating near other firms and purchasing cheaper intermediate goods since these do not incur any transport costs. However, the savings generated from buying intermediate goods fall as transports costs fall, whereas the interregional wage differentials persist. When transport costs reach a sufficiently low value, firms can obtain benefits from relocating in the deindustrialised region in the periphery where the immobile factors are cheaper, combining imported intermediate goods with cheaper local labour. In this case, the firms can choose to delocalise production in order to reduce their production costs, giving rise to a spatial fragmentation of firms: production activities are transferred to the regions with lower wages while some strategic functions will remain concentrated in a few urban regions.

Therefore, as transport costs change, the relative intensity of the agglomeration and dispersion forces vary, giving rise to different degrees of spatial inequality. In the early stages of integration centripetal forces predominate producing an increase in the spatial concentration of economic activity, and therefore, of inequality. Once a certain level of integration has been reached, this trend is reversed, giving rise to a dispersion of economic activity and a reduction in regional inequality. However, this result depends to a large measure on the assumptions that are made in relation to the existence of worker mobility at the regional level in response to wage income differentials. When the workers decide to migrate to places with a greater number of firms and higher real wages, agglomeration is intensified. By contrast, when the workers stay where they are, interregional wage differentials persist. Consequently, the relationship between integration and agglomeration is no longer monotonic, since reductions in transport costs make firms more sensitive to cost differences generated by the wage differential, leading to the spatial dispersion of industry.

All those factors that limit interregional labour mobility become *dispersion* or *centrifugal forces* that work against the concentration of economic activity. These forces can be diverse. First, the literature has pointed to the importance of the appearance of congestion costs derived from agglomeration, due to the fact that a large quantity of consumer goods and services are non-tradable. Land competition in large urban agglomerations can lead to an increase in housing prices. This generates not only an increase in the cost of living in the larger regions, but it also increases the number of workers that have to commute each day as the cities expand (*commuting costs*)[47]. Thus, the existence of growing urban costs generated by the agglomeration in the core becomes a centrifugal force causing firms and workers to disperse to avoid these costs.

Likewise, so far it has been assumed that workers are homogeneous and that they only move as a result of potential differences in income as a function of the wage levels and prices in each region. However, workers are heterogeneous, which means that each potentially mobile individual will not react in the same way to interregional economic differences. Furthermore, the individual's decision to migrate is based on a large number of considerations, many of which are non economic. The worker's reticence to move can be linked to a range of factors. To mention just a few, there might be reasons related to their personal life or to the attributes of the region of origin such as proximity to the

[47] Combes, Mayer and Thisse (2008) estimate that in the US, rent and transport costs together represent around 40% of the family budget. In the Paris region, these costs can reach 45%.

family, the climate or ties to the land. Furthermore, it would be reasonable to assume that as workers' income rises and their basic needs are satisfied, they will value more highly these non-economic factors linked to the quality of life.

It is also interesting to note that when considering the presence of urban costs (Ottaviano, Tabuchi and Thisse, 2002) and the heterogeneity of individual attitudes as regards migration (Tabuchi and Thisse, 2002), to which we can add the transport costs that are positive for agricultural goods (Picard and Zeng, 2005), not only are some of the more restrictive, or less realistic, assumptions from earlier NEG models relaxed, but also the existence of a bell-shaped evolution in the relationship between economic integration and inequality in different contexts is confirmed.

Therefore, the initial impact of greater economic integration can be the strengthening of regional disparities. Yet, in the absence of interregional labour mobility, if the process of economic integration progresses, once a certain level of integration has been reached an inverse process is set in motion whereby greater economic integration leads to a reduction in regional inequalities. The theoretical models suggest that a reindustrialisation of the periphery can occur due to the fact that the dispersion forces start to act once the transport cost have reached a low enough level. Thus, progressive market integration can lead to a regional convergence in terms of both real wages and the structures of production. However, for this convergence to exist, the integration must have progressed sufficiently.

The political implications of these outcomes are not so alarming as regards the consequences that the process of European integration, for example, might have. In turn, the theoretical predictions seem to match more closely the patterns observed in the real world: so, rather than a catastrophic shift from a regular spatial distribution of industry to its complete concentration in a single region, which was typical of previous models, here a gradual process of change is ushered in, in which the regions have industrial sectors of different sizes.

In fact, this theoretical prediction is in line with a number of empirical studies including that of Williamson (1965). As mentioned earlier in the introduction, this author described, in line with the work by Kuznets, the existence of an inverted U-shaped relationship between the process of national economic growth and development and regional inequality. His regional analysis of countries at different levels of development showed that during the early stages of the process an increase in regional disparities is recorded. However, as an economy continues to develop, this relation tends to vary and regional inequality falls[48]. Therefore, the theoretical predictions of NEG seem to fit this evidence better.

2.1.3. Multi-regional NEG models: cross-country economic integration and the internal geography of countries

So far the NEG models reviewed have focused on the application of an analytical framework comprising two regions, where the mobility or immobility of workers has different consequences for the spatial distribution of economic activity. However, when considering more than two regions, the accessibility to markets varies across these regions[49]. Here, each region's capacity to attract firms and workers depends on its position in relation to the markets, so that size and market access, as well as competition from other firms, will affect a firm's location. Yet, as well as the integration of the national economy, it is necessary to consider the integration of national economies within international trade, which also has a significant impact on the location of economic activity within each country. In this sense the question needs to be raised as to the impact of trade policy on patterns of regional development within countries. Theoretically, this aspect has been analysed in a number of studies.

One of the first theoretical contributions to the debate within NEG was the study by Krugman and Livas Elizondo (1996), who sought to explain the effect of trade policies on the formation of large metropolises in developing countries in recent decades. Before World War II, the largest cities were to be found in industrialised countries, but since then large urban centres have proliferated in developing countries. Drawing on the experience of Mexico and the studies of Hanson (1996, 1997)[50], Krugman and Livas Elizondo (1996) developed a theoretical model aimed at explaining the effect of trade policies on the internal economic geography of countries.

In the case of Mexico, import-substituting industrialisation (ISI) policies adopted since the 1940s led to the agglomeration of economic activities in the capital converting it into one of the world's most populous metropolises. This economic agglomeration was

[48] "*...rising regional income disparities and increasing North-South dualism is typical of early development stages, while regional convergence and a disappearance of severe North-South problems is typical of the more mature stages of national growth and development*". Williamson (1965), p. 44.

[49] "*...the new fundamental ingredient that a multi-regional setting brings about is that the accessibility to markets varies across regions. In other words, spatial frictions between any two regions are likely to be different, which means that the relative position of the region within the whole network of interaction matters*". Behrens and Thisse (2007), p. 462.

[50] Hanson's empirical studies of the Mexican case are reviewed in section (2.2.2.2) in this chapter.

linked to political decisions that aimed to protect the Mexican domestic market. This situation, however, began to change in the 1980s. The abandonment of ISI policies and the liberalisation of the Mexican economy led to an increase in the decentralisation of manufacturing primarily away from Mexico DF to the northern zones of the country near the border with the United States.

Based on a formal model of two countries and three regions, Krugman and Livas Elizondo (1996) showed that a low level of openness in an economy leads to the spatial concentration of manufacturing activities because of the strong backward and forward linkages that arise from selling in a small domestic market. By contrast, when an economy becomes more open, the effect of this liberalisation is a spatial dispersion of manufacturing activities where the dispersion forces considered are land rents. As the importance of domestic demand is reduced, firms will tend to have fewer incentives to locate near this demand.

However, a series of studies suggests the opposite to be the case. Following the line of study of the emergence of major metropolises in developing countries, Alonso-Villar (2001), using a model comprising three countries in which the dispersion force considered is once more the congestion costs, suggests that the appearance of major agglomerations is not exclusively the result of protectionist trade policies. Other factors have also to be considered, such as the relative position of each country, in terms of industrialisation, with respect to other countries, and the competition to which firms on the international market are subject. Since manufactured goods produced in developing countries have to compete with those produced in the rest of the world, firms do not seek proximity with foreign markets where they might have difficulties competing with foreign products and they locate at a site that best allows them to supply the domestic market.

Elsewhere, Monfort and Nicolini (2000) developed a model that includes two countries and four regions, and in which the dispersion force is the existence of immobile consumers. Their results showed that the fall in trade costs derived from trade liberalisation leads to an increase in the agglomeration forces within a country. When interregional transport costs are low, strong competition between local producers means that it is not beneficial to locate in the periphery to supply the domestic market. With trade liberalisation, if international trade costs are reduced, competition in the periphery stems from the foreign producers and, as above, this location becomes less attractive. Consequently, greater economic integration can foster the emergence of regional economic agglomerations within countries. In this way, trade liberalisation works against regional convergence by strengthening the agglomeration forces within countries.

Paluzie (2001) also analysed the effect of trade policies on regional inequality patterns. In a model with two countries and three regions, her study examines the emergence of regional inequality in processes of integration such as that of the European Union. Paluzie (2001) concluded that trade liberalisation can give rise to a polarisation in the distribution of economic activities, and consequently, to an increase in regional inequalities within a state. This result can explain the interruption in the process of regional convergence observed in the European Union since the 1980s, a fact that corresponds perfectly with the evolution observed in the case of Spain at the regional level since joining the common market in 1986. As in Monfort and Nicolini (2000), in these last two studies, the dispersion force is not land rents and urban commuting costs as in Krugman and Livas Elizondo (1996), but rather the pull of a dispersed rural market resulting from the existence of immobile workers as in Krugman (1991), an assumption that appears more appropriate when the object of study is the European context and not the metropolises in developing countries.

Adopting a similar line to these last two studies Crozet and Koenig (2004a) suggested that the impact of an increase in foreign trade on an economy's spatial distribution depends on the internal geography of the country. They adopted a model with two countries and three regions, two domestic and one foreign, considering two alternative scenarios. First, they examined the effect of a reduction in international trade costs on the spatial distribution of activity for a homogeneous country in which the two domestic regions are equidistant from the border and, therefore, they have the same access to foreign markets. The different simulations undertaken showed that international economic integration gives rise to a spatially concentrated domestic industrial sector.

Then, it is assumed that one of the two domestic regions has a better foreign market access, so that the existence of two heterogeneous regions modifies the forces affecting the domestic economy. On the one hand, access to a larger foreign market reduces the incentive of local firms to locate near the domestic consumers, since now they represent a smaller share of their sales. In this case, a potential effect of trade liberalisation is to push domestic firms towards the regions closest to the foreign markets so as to benefit from a better access to foreign demand. Furthermore, better foreign market access not only means better export opportunities but also the opportunity to import cheaper inputs. But on the other hand, trade liberalisation equally generates an increase in the competition exerted by foreign firms in the domestic market. In this situation, trade liberalisation can push domestic firms to locate in interior regions further from the foreign market so as to protect themselves from foreign competition.

Thus, a gradual liberalisation of trade can generate two effects: a *pull-effect* towards the regions of the geographical periphery near the foreign markets and a *push-effect* towards the regions of the interior that are better located for supplying the domestic market. The impact of these forces depends on several factors: if the foreign demand for domestic products is high, domestic firms will tend to locate in the region with better access to international markets; by contrast, a large number of foreign firms exporting to the domestic market can favour the development of interior regions that are more protected from international competition.

Therefore, trade liberalisation gives rise to the appearance of economic forces that can operate in different directions, although the results obtained by Crozet and Koenig (2004a) suggest, based on simulations conducted for different model parameters, that regions nearer to the foreign markets are more attractive for the location of firms. Thus, trade liberalisation would lead to an increase in the concentration of economic activity. And this concentration would show a greater propensity to locate in regions closest to foreign markets. The only situation in which agglomeration would occur in the interior region would be when the initial distribution of activity strongly favoured that region.

In conclusion, there would appear to be no consensus as regards the theoretical predictions regarding the effects of trade liberalisation on the distribution of economic activity within a country. As noted, on the one hand, it is claimed that the effect of trade liberalisation is the dispersion of economic activity within a country (Krugman and Livas Elizondo, 1996). On the other, the studies of Monfort and Nicolini (2000), Alonso-Villar (2001), Paluzie (2001) and Crozet and Koenig (2004a) conclude that the possible outcome of trade liberalisation is an increase in agglomeration within the country itself. In this way, the trade policy adopted by a country is linked to the location of economic activity within it, and therefore, to its spatial inequalities[51].

[51] Among studies that analyse this subject from a different perspective see Ades and Glaeser (1995), Behrens (2003), and Behrens, Gaigné, Ottaviano and Thisse (2006).

2.2. A survey of NEG empirics: market potential and agglomeration

2.2.1. Market access: relevance and origins

A quick glance at a map showing the regional distribution of GDP per capita within the European Union reveals the existence of a marked core-periphery pattern. Most of the regions with the highest per capita income in Europe are located in the area known as the European '*blue banana*' which extends from London, through the Benelux countries and western Germany, into northern Italy. Outside the economic and geographic core of Europe, a number of other zones of high income stand out such as the Scandinavian countries (Denmark, Norway, Sweden and the Helsinki area), and the Paris region. By contrast, the economic periphery coincides in many instances with the geographical periphery, including Portugal, Greece, the recently incorporated countries of the East and the southernmost parts of Italy and Spain.

When explaining this core-periphery pattern in terms of GDP per capita and its persistence over time, a number of studies have focused on the role played by market access. The reason for this is the strong correlation between the two variables as is emphasised by the regional structure of market potential in the European Union: the regions in the economic core of Europe are, in turn, those that enjoy the best market access. Thus, it is highly unlikely that the regions near these areas of greatest market potential will be poor, since their proximity to the main markets converts them into attractive zones for both firms and workers. Given that the theoretical models of NEG also identify market access as a determining factor in the genesis and evolution of regional inequalities, the review of empirical studies undertaken in this section focuses on those studies that seek to test the theoretical predictions of NEG most closely concerned with market access.

However, there is a long tradition of studies of market access that predates the development of the theoretical models of NEG. A good many of these studies have been undertaken in geography, since it was in this field where the question was first raised as to how to measure market access and the effects that this has on industrial location. Harris' (1954) pioneering study sought an explanation for the high concentration of US manufacturing in the country's northeast industrial belt, which at that time, with an area that was just a fourteenth of the country's total area, concentrated around 70% of industrial employment. As Harris suggested, manufacturing had primarily developed in areas and regions near the large markets and in turn, the size of these markets had increased due to the size of the industrial base, indicating a circular causation similar to that maintained by NEG. In order to examine this relationship, Harris (1954) proposed a measure of market access based on the following formula:

$$P = \sum\left(\frac{M}{d}\right)$$

where market potential *(P)* is defined as the summation of markets *(M)* accessible to a point divided by their distances from that point, where *'M'* is a measure of the economic activity in each area, and *'d'* the distance (or the transport costs) between areas and regions. The concept of market potential is, therefore, an index representative of accessibility to the markets based on the idea (taken from physics) that as the distance between two points becomes greater the attraction between them weakens, represented in this case by a smaller number of economic relations between the two points.

Harris' analysis was followed by a series of studies that shared his concern for the geographical distribution of economic activity and, more specifically, that of manufacturing. Clark, Wilson and Bradley (1969) analysed the evolution in market potential in Western Europe. Their aim was to investigate the effects generated by European integration, following the Treaty of Rome (1957), on the location of European industry, based on the analysis of market potential and the variation in this indicator over time. To do this, they examined the accessibility of the European regions prior to the Treaty, the changes in this indicator after its signing and the predicted situation of the future incorporation of new countries within the European Economic Community (Denmark, Great Britain, Ireland and Norway), in a context moreover, affected by developments in transport technology. Their results showed an increase in the concentration of industrial activity in the central regions of Europe that enjoyed greater market potential following European integration and indicated a possible strengthening of this tendency with the future enlargement of the Community. Furthermore, Clark, Wilson and Bradley (1969) warned of the possibility that should Britain not join the common market it might find itself distanced from the European economic centre.

Based on a specification similar to that proposed by Harris (1954), these authors calculated market potential in terms of regional income *(M)* and transport costs *(d)*. A significant advance in this study with respect to that of their predecessor, in which market access was measured solely in terms of the US economy, was the inclusion in the analysis of countries with which trade links existed. The consideration of these countries introduced the need to add tariffs to the calculation of market potential, understood as a barrier that increased transport costs and which, therefore, had to be added to the denominator in the equation.

A further seminal study in this field was that undertaken by Keeble, Owens and Thompson (1982)[52]. This study, also centred on the European Economic Community, measured market accessibility in the EC regions, assuming that greater market potential favoured investment decisions and, hence, the rate of regional growth. In their study, the most inaccessible regions, understood as those with the least market potential, found themselves in the geographical periphery of Europe, while the regions with the best market access, represented by a high market potential, lay in the European core in an area around the Netherlands, Belgium and West Germany. Their results pointed to major differences in the accessibility of the European regions and to a widening of these regional disparities in the second half of the 1960s and throughout the 1970s, even after the enlargement of the EC in 1973. Therefore, the authors concluded that the impact of European integration on market access had been only moderate but that, by contrast, the pattern of market potential was a closer reflection of historical processes of industrialisation and urbanisation.

A key aspect of the methodology adopted in this study is its consideration of the most appropriate measure for each region's 'self-potential'. In line with Harris' equation, the economic activity of region *r* has to be incorporated and a decision taken as to the most appropriate intraregional distance to be used. Here, as a measure of a region's internal distance, Keeble, Owens and Thompson (1982) adopted a formula similar to that previously used by Rich (1980). This expression provides an internal distance in each region that corresponds to a third of the radius of a circle that has a similar area to that of region *r*:

$$d_{rr} = \frac{1}{3}\sqrt{\frac{areaofregion}{\pi}}$$

[52] Later complemented by Keeble, Offord and Walker (1988).

In recent years, studies of market accessibility have proliferated in relation to the gradual expansion of the European Union. The relevance of the subject is marked by the fact that the integration of new members can generate changes in the patterns of market access causing a variation in the spatial structure of economic activity[53].

Other studies have recently been undertaken within transport economics (Gutiérrez and Urbano, 1996; Vickerman, Spiekermann and Wegener, 1999). Here, the interest is centred on the impact that trans-European road networks might have on changes in the accessibility of European regions. The development of new regional transport infrastructure has been a main goal of European policy aimed at cohesion, and a sizeable share of the structural funds received by Europe's less prosperous regions has been invested in the construction of transport infrastructure. These studies examine the evolution in transport networks as a means of reducing distances and transport costs between regions, and analyse the impact that they can have on market accessibility by bringing the peripheral regions closer to Europe's core regions. Here, it is hoped that the construction of supranational infrastructure can reduce the initial locational disadvantage in terms of the accessibility of the European periphery, enhancing the capacity of these zones to attract economic activity, thereby leading to a reduction in territorial inequality within the Union[54].

2.2.2. Market potential and the empirics of NEG

Having reviewed some of the main studies in the early analysis of market potential undertaken in the field of geography, in the sections that follow the focus shifts to examine a series of studies that have sought to test empirically some of the main theoretical predictions derived from the NEG models, linked in varying degrees to market access[55]. However, as is typically stressed, the empirical studies still lag behind the theoretical development of NEG which has been much more prolific in recent decades. This is due to the difficulties that are often found when studying the empirical relevance and predictive power of the NEG models.

First, NEG suggests that firms locate where they expect their profits to be greatest, i.e., sites that are characterised by better market access and, therefore, by greater demand

[53] This is the case, for example, of studies by Brülhart, Crozet and Koenig (2004), Crozet and Koenig (2004b), Behrens, Gaigné, Ottaviano and Thisse (2007) and Lafourcade and Thisse (2008).

[54] A more detailed review of the most recent studies conducted from a European perspective can be consulted in Combes and Overman (2004).

[55] A recent review of the empirical literature examined in the following section can be found in Head and Mayer (2004b), Combes, Mayer and Thisse (2008) and Redding (2010).

(section 2.2.2.1). In turn, the theory predicts that the increase in the number of firms located in such a region, as a result of its greater market potential, causes factor prices to increase. The relationship between nominal wages and market access has been empirically validated both internationally and within countries (section 2.2.2.2). Finally, and as a consequence of the increase in real wages, the theoretical models uphold that labour mobility is related to the market potential of the different locations (section 2.2.2.1).

2.2.2.1. Market potential attracts production factors

As discussed in the theoretical review section, the spatial agglomeration of economic activity in Krugman (1991) is the joint result of the interaction of two centripetal forces, related with the two production factors: capital (backward linkages) and labour (forward linkages).

In the case of the first of these two forces, interest has focused on studies that have examined the determinants of firms' location decisions by empirically testing the existence of backward linkages. This centripetal force, which has its origin in the presence of transport costs and scale economies, suggests that firms prefer to locate in a region that enjoys good market access. In this way, a firm's profitability is directly linked to market potential.

Of the various studies that have examined this relationship, the one that adopts a line closest to the theoretical models of NEG is Head and Mayer (2004a)[56]. These authors analysed the location choices of Japanese multinational firms in Europe for a total of 57 European regions in nine countries (Belgium, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain, and the United Kingdom) in the period between 1984 and 1995. With this goal in mind, Head and Mayer (2004a) estimated a profit equation as a function of the relative accessibility of the regions, controlling for differences in regional costs. The results showed that consumer demand does indeed matter for location choice: a 10% increase in market potential increases the chances of a region being chosen by 3 to 11%. Yet, the authors concluded that these backward linkages might not be the only, or even the most important, determinant of the firms' behaviour[57].

[56] Other key studies include Friedman, Gerlowski and Silberman (1992), Henderson and Kuncoro (1996), Devereux and Griffith (1998), Head, Ries and Swenson (1999), and Crozet, Mayer and Mucchielli (2004). A review of these studies can be found in Head and Mayer (2004b).

[57] Not only do land prices and the skill level of the work force have to be considered, but also other factors such as the regional differences in taxes and, therefore differences in the cost of capital, the subsidies

On the other hand, a larger number of firms in a region will push up nominal wages while the local price index will fall as the number of locally produced varieties increases[58]. Together these two factors cause real wages to rise. If the wage gap between regions is sufficiently wide, workers will be attracted by the regions with greater market potential where they can maximize their real wages after deducting their mobility costs. This attraction will lead to positive migratory balances in the regions of greatest market potential (forward linkages), thus strengthening the agglomeration tendency.

Among the empirical tests for the existence of forward linkages, the work by Crozet (2004) should be highlighted[59]. He examined whether market access and real wage differentials in Europe have a positive influence on the decisions of migrant workers. To do so, he assumed that workers choose locations on the basis of regional real wage differentials, considering workers to be heterogeneous and bearing in mind the effects that regional unemployment can have. In addition, the workers that decide to migrate have to deduct the costs of migrating, and these costs grow in proportion to the distance between the regions.

The estimation of an equation directly derived from a theoretical model is applied to the study of five European countries (Germany, Italy, the Netherlands, Spain, and the United Kingdom) during the 1980s and 1990s. The results obtained by Crozet (2004) offer solid evidence in favour of the existence of a forward linkage, i.e., that the regions with greatest market potential attract workers. Yet, the simulations based on the estimated parameters show that the agglomeration forces are limited geographically and the study predicts that the distance at which a region is likely to begin to attract workers from distant zones is small. These forces are, therefore, too weak to counter the high barriers to migration that affect the location decisions of individuals in Europe. Consequently, and in part due to the low propensity to migrate, the forward linkages appear unable to engender core-periphery-type structures at the large spatial scale within the European Union in the short term, at least while the workers remain so sensitive to mobility costs.

---

granted by European regional policy, and first nature geography elements might also affect a firm's location decision. Combes, Mayer and Thisse (2008).

58 The greater nominal remuneration of the factors of production in the regions that enjoy better market access is dealt with in the following section (2.2.2.2).

59 Other studies in Kancs (2005, 2010), Pons, Paluzie, Silvestre and Tirado (2007), Paluzie, Pons, Silvestre and Tirado (2009), and Hering and Paillacar (2008, 2009). For a recent survey of empirical studies in this field, see Clark, Herrin, Knapp and White (2003).

2.2.2.2. Market potential raises the price of the production factors

If the region that enjoys the best market access attracts capital and more firms locate in that region, the increase in demand will push nominal wage levels up, increasing, therefore, the retribution of the labour factor. In this field, the main empirical studies are based on the wage equation proposed by Krugman (1991), in which he established the relationship between factor prices and market potential. The wage equation determines the zero-profit condition for firms and implicitly it defines the maximum factor price level that a representative firm can pay in each region or country given its market access. In other words, the equation captures the idea that regions or countries with better market access can pay relatively higher wages.

Different versions of the wage equation have been applied in the empirical studies. They can be differentiated between: a) those that are concerned with inequality in per capita income across countries; b) those which seek to study wage differences across regions within a country, and; c) finally, other applications linked to the wage equation will be briefly commented.

a) Market potential and cross-country income inequality. Redding and Venables' (2004) study analysed the impact of access to markets on cross-country variation in per capita income. They adopted a version of the wage equation that examines which part of the cross-country variation in GDP per capita can be explained by market access (*MA*) and supplier access (*SA*). This equation reflects that both market potential and proximity to suppliers (input-output linkages) have a positive relationship with GDP per capita[60]:

$$\ln PIBpc_r = \frac{1}{\sigma}\ln \overline{MA}_r + \frac{\alpha}{\sigma - 1}\ln \overline{SA}_r + \zeta + \varepsilon_r$$

The estimation of this equation was undertaken for a cross-section of 101 countries in 1994, using a strategy that is applied in a number of consecutive steps. First, bilateral trade flows between countries are used to estimate a gravity equation, from which they obtained the coefficients that enable them to construct measures of market access (*MA*) and supplier access (*SA*). Thus, distance in addition to other trade barriers mean that the countries most distant from the economic core suffer a world market access

[60] In this case, GDP per capita is considered a *proxy* of wages (w_i), ζ is a constant and ε_t is the disturbance term.

penalty on the sales of their products. Furthermore, they also face additional costs for the importation of inputs and intermediate goods.

Second, the aforementioned wage equation was estimated. The results showed that approximately 73% of the cross-country variation in GDP per capita can be explained by market access differentials. Yet, a regression of this type can present problems of endogeneity especially when the domestic component ϕ_{rr} predominates within the measures of *MA* and *SA*. To overcome this problem of endogeneity, two additional estimations were conducted.

First, the domestic component was excluded from measures of *MA* and *SA*. Although this procedure allowed them to reduce the possible bias in the estimation linked to endogeneity, eliminating the domestic market makes less sense from an economic point of view. In this case, the explanatory power of the market access for cross-country per capita income falls to 35%. Although this result is lower than the earlier one, it allowed the authors to conclude that a country's development depends significantly on that of its neighbours.

Second, they sought to overcome the problem of endogeneity by including both the domestic and foreign components and by using instrumental variables (IV). In this case, *MA* and *SA* are instrumented with exogenous geographical variables: the distance to the world's three main markets represented by New York, Brussels and Tokyo. With this new estimation, the initial results were maintained[61].

Finally, Redding and Venables (2004) also analysed another of the mechanisms highlighted by NEG models and which is concerned with the effect of geographical location on per capita income through the manufacturing price index. Countries that are remote from suppliers of manufactured goods incur greater transport costs and have, as a result, a higher price index (G_i). The results obtained in this case coincide with the theoretical predictions: countries with greater supplier access (*SA*) are characterised by lower relative prices of equipment and machinery.

Recently, Mayer (2008) has analysed the impact of market potential on a country's economic development in greater detail following the proposals of Redding and Venables (2004). In so doing, he expanded the time period analysed, drawing on international trade data for the period 1960 to 2003, in order to evaluate the persistence of

[61] Faced with the possibility that the estimation might be affected by the problem of omitted variables, the authors controlled for this by including a number of typical variables from the growth literature, such as resource endowments, levels of education, physical geography and the quality of a country's institutions.

these patterns over time. His results showed that economic geography measured using a market potential index is a powerful driver of economic development, leading the author to conclude that the cross-section results of Redding and Venables (2004) are robust when a panel estimation is undertaken[62].

b) Market potential and regional wage inequality. While Redding and Venables (2004) sought to explain differences in cross-country wages in terms of GDP per capita within the NEG framework, Hanson (1998, 2005) centred his study on the impact of market access on the spatial distribution of regional wages within countries[63]. Redding and Venables (2004) worked from the assumption of labour immobility as in Krugman and Venables (1995), where the agglomeration force is linked to the presence of intermediate goods. This way of proceeding appears more adequate when the object of study is cross-country inequality. By contrast, Hanson (2005), by focusing on interregional inequality, assumed labour mobility as in Krugman (1991) and replaced the agricultural good in household consumption with housing costs in line with Helpman (1998).

As already seen in the description of the theoretical framework, the NEG models show the existence of a wage equation that links the wages of a particular location with a market potential function. Thus, the theoretical models predict a spatial structure for wages, in which the latter tend to be higher in regions near to large markets since they can serve a high demand by assuming lower transport costs.

Hanson (2005) centred his study on the 3,075 US counties in the 1970s and 1980s testing via a structural estimation the verification of the Krugman wage equation in two versions. First, he estimated the simple market-potential function, an equation that is very close to Harris' (1954) definition of market potential:

$$\log(w_{rt}) = \alpha_0 + \alpha_1 \log\left(\sum_s Y_{st} e^{\alpha_2 d_{rs}}\right) + \varepsilon_{rt}$$

where w_{rt} is the nominal wage in region r in time period t, Y_{st} is the GDP in region s in period t, d_{rs} is the distance between regions r and s, and α_0, α_1, and α_2 are parameters to be

[62] Mayer (2008) also seeks to solve the problems of endogeneity by only computing the foreign market potential and instrumenting geographical centrality and total transaction costs. He also controls for the education levels of the population as a factor that varies over time but which is omitted from the regression.
[63] Hanson (1998) corresponds to the earlier working paper version of Hanson (2005).

estimated. In turn, the augmented version of the market-potential function includes in its equation a version of the market potential structurally derived from a theoretical model, the housing stock *(H_{st})*, and the wage *(w_{st})* in the neighbouring regions[64].

Hanson's results proved the existence of a spatial wage structure in the US counties. There exists, therefore, a wage gradient where a county's wage positively correlates with that county's market potential. However, the parameters estimated show that the agglomeration forces are limited to the geographical scale. The economic influence of the wages in the neighbouring areas for any county falls rapidly with distance and is only effective in a radius of less than 1000 kilometres. The income of the zones lying beyond this distance does not exert a positive influence on the determination of local wages.

Following this line of research a large number of studies have replicated Hanson's pioneering analysis for other countries with the aim of testing the wage equation. This is the case of Roos (2001) and Brakman, Garretsen and Schramm (2004) for Germany; De Bruyne (2003) for Belgium; Fally, Paillacar and Terra (2010) for Brazil; Hering and Poncet (2006) for China; Paluzie, Pons and Tirado (2005, 2009) and Garcia Pires (2006) for Spain; Knaap (2006) for the US; Combes, Duranton and Gobillon (2008) for France; Amiti and Cameron (2007) for Indonesia; Mion (2004) for Italy; Kiso (2005) for Japan; Head and Mayer (2006) and Niebuhr (2006) for the European Union.

In general, these studies confirm the existence of a within-country spatial wage structure and show the success of the empirical testing of the wage equation, an important mechanism within NEG. Together, therefore, market potential seems to maintain a positive relationship with wages both internationally and regionally within countries.

At this juncture, it is important to bear in mind certain considerations regarding the measurement of market potential. In the empirical studies reviewed market access is basically measured in two alternative ways. In some cases, when the information so permits, market potential can be derived from a structural model linked to New Economic Geography. In other cases, market potential is measured using Harris' equation. Some studies have compared both approaches to discern which is most appropriate. The results, however, are inconclusive. In the case of the wage equation estimation undertaken by Hanson (2005), the comparison between the two approaches shows that the structural

[64] Similarly, the equation is estimated in time differences, considering the skills of the workers to be heterogeneous and using instrumental variables to solve the problem of endogeneity. However, as Hanson recognises, other factors might be affecting his results, such as technology spillovers. Hanson (2005), p. 21.

alternative closer to the theoretical model gives better results than Harris' market potential. By contrast, in Head and Mayer's (2004a) study that seeks to corroborate the existence of backward linkages in the investment of Japanese multinationals in Europe, Harris' measure, which is not derived from a theoretical model, provides a better fit.

c) Other applications of the wage equation and trade liberalisation. A series of studies have attempted to empirically test the effect of trade liberalisation on factor prices and the distribution of economic activity within countries that open up to foreign trade. First, Hanson (1996, 1997) focused on the effect of changes in trade policy on regional wages in Mexico, although on this occasion the study was not based on a structural estimation derived directly from NEG models.

As discussed above, in the 1940s Mexico began an import-substituting industrialisation (ISI) policy at a time when most of its manufacturing activity was concentrated in the capital. In the 1980s, the turnaround in Mexican trade policy and the gradual opening up to foreign trade with its admission to the GATT (1986) and later its integration in the NAFTA (1994) led to a change in the location of manufacturing that gradually moved from Mexico DF to the north of the country in areas close to the US border. In this way, trade liberalisation, by altering the spatial structure of market potential, contributed to dispersing economic activity, in the line suggested by Krugman and Livas Elizondo (1996).

In Hanson (1997), the dependent variable was the relative wage in each Mexican region with respect to that of Mexico DF in different industrial sectors. The explanatory variables included distances to the capital and to the border crossings with the US. The results showed that a spatial wage structure exists whereby relative regional nominal wages fall as we move away from these two industrial centres: a 10% increase in distance from the capital reduces wages by 1.92%, while a similar increase in the distance from the US border, reduces wages by 1.28%. Again, regional wages are related to market accessibility. On the other hand, trade liberalisation in the mid-1980s should have contributed equally to a weakening in the wage gradient around Mexico DF. In this case, however, the evidence of a variation in the gradient was not so strong.

A similar line of analysis, linked to the impact of trade liberalisation on the internal distribution of economic activity within a country, was undertaken by Brülhart, Crozet and Koenig (2004) and Crozet and Koenig (2004b). However, in this instance the strategy adopted is different and distances itself from the wage equation. Brülhart, Crozet and Koenig (2004) study examined the possible effects of the enlargement of the European Union on the income of the less developed regions (Objective 1 regions). To do so, they measured the changes that would occur in the market access of these regions as a result of European enlargement and its effects on regional per capita income considering three different scenarios: EU-15, EU-25 and a hypothetical EU-33[65].

The results of this study indicate that the effect on the per capita income of Objective 1 regions depends on their geographical location relative to that of the new member states, located principally in Eastern Europe and the Balkans. This effect on GDP per capita would be more marked in the Objective 1 regions located closest to the new members, such as Burgenland (Austria), where the impact would be six times greater than that recorded in Objective 1 regions lying further away from Eastern Europe and the Balkans, as would be the case of South Yorkshire in the United Kingdom. Although the effect in these regions is moderate, Brülhart, Crozet and Koenig (2004) suggested that the Objective 1 regions in Greece could benefit from the future enlargement of the European Union into the Balkans.

Finally, Crozet and Koenig (2004b) tested the theoretical predictions from their theoretical model, according to which, economic integration could foster an increase in the agglomeration of economic activity in the region that enjoys the best international market access, except when competition from foreign firms is too high[66]. They based their test on the example of Romania, in order to examine whether, as a result of the country's incorporation to the process of European integration, its activity is concentrating in the regions nearest to the border with the European Union in the west of the country, and whether the existing agglomeration in the interior region around Bucharest is weakening. Their study showed that urbanisation rates are greater in the regions closest to the EU and that they, as a result, enjoy greater market potential. These results strengthen the theoretical predictions of the model developed by these authors.

2.2.3. Market potential, NEG and economic history

Various studies have undertaken long-term analyses of the geographical distribution of industry from within the framework of New Economic Geography. These studies seek to examine either the whole process of industrialisation, or one or more of its stages, which occurred in parallel, from the middle of the 19th century, with the

[65] The study uses Harris' (1954) market potential equation.
[66] The theoretical model developed by Crozet and Koenig (2004b) is discussed in section 2.1.3.

integration of national economies and the rapid fall in transport costs. The initial aim has been, therefore, to analyse the evolution over time in the spatial distribution of industry and its determinants.

The key study in this field is Kim's (1995) pioneering work examining long-term trends in manufacturing location in the United States, in which he concluded that regional specialisation intensified in the second half of the 19th century, reaching a peak in the interwar years. However, there was a shift in this trend in the 1930s and since then regional specialisation has shown a substantial and constant decline. Likewise, the spatial concentration of the manufacturing sector followed a similar pattern, as the regions initially became more and more specialised. Despite the slight fall in the second half of the 19th century, the long-term dynamics shows the existence of a bell-shaped relationship in the spatial concentration of industry in the US during the long industrialisation process, which peaked in the 1920s. Having described this evolution, the next step was to identify its determinants. To do this, Kim (1995) undertook an exercise using panel data for five different years (1880, 1914, 1947, 1967, 1987) and 20 industries in which he estimated an equation where the endogenous variable, the spatial concentration index, was regressed against a measure of scale economies (plant size) and resource endowments (raw material intensity). Kim reported that scale economies had a significant impact on spatial concentration, which provided evidence in support of NEG. Yet, he pointed out that the relative interregional differences in resource endowments, in line with the Heckscher-Ohlin model, would have been the main element accounting for the long-term evolution of the US manufacturing sector, thereby limiting the role of increasing returns in shaping the spatial distribution of industry.

A similar study has been conducted for the Spanish economy by Tirado, Paluzie and Pons (2002) in which they showed that during the second half of the 19th century, coinciding with the progressive integration of the Spanish economy, there occurred a gradual concentration of industrial activity in a small number of territories. When broadening the timescale, in a study that extended over 150 years from the middle of the 19th century to the end of the 20th, Paluzie, Pons and Tirado (2004) confirmed the marked rise in the geographical concentration of Spanish industry between the second half of the 1800s up to the Civil War (1936-1939). This concentration continued to slightly increase until the 1970s when there was a shift in the trend as shown by the reduction in the coefficients of industrial location. Therefore, this evolution, in common with that recorded in the United States and in the theoretical models of NEG, also shows the existence of a non-monotonic relationship between market integration and industrial concentration over time in Spain.

Tirado, Pons and Paluzie's (2002) study, focusing on the second half of the 19th century, undertook an analysis of the explanatory factors of spatial concentration in Spain in line with Kim (1995). They identified scale economies and market size as the determinants of the industrial geography in 1856. At the end of the century, factor endowments (in this case the accumulation of human capital) were added to the explanation of industrial location, at the same time as the elements of New Economic Geography (scale economies and market access) increased their explanatory power with the parallel advance in the process of economic integration.

Rosés (2003) also identified the influence of NEG forces in the regional specialisation of Spanish production in the middle of the 19th century. Based on an exercise that combined the advantages afforded by factor endowments and NEG forces, the author tested, in accordance with the line of analysis proposed by Davis and Weinstein (1999, 2003), the existence of a 'home market effect' during the early stages of Spanish industrialisation[67]. Rosés (2003) concluded that during the rise of Catalonia as a centre of industrial production in the first phase of Spain's industrialisation two types of basic explanatory elements coincided: factor endowments, tied to the availability of human capital, and home market size, which resulted in advantages for the location of manufacturing around Barcelona.

Similarly, Betrán (1999) studied the interwar period suggesting that the relative increase in industrial activity in provinces such as Vizcaya, Guipúzcoa, Madrid and Zaragoza during this period could have been linked to the presence of agglomeration economies derived from market size. Taken together, these studies of the Spanish case, suggest that agglomeration forces were already present by the middle of the 19th century, and that they grew stronger in the second half of that century, with their impact being maintained into the interwar years.

[67] Davis and Weinstein (1999, 2003) analysed the existence of Krugman's (1980) 'home market effect' based on two variables (*share* and *idiodem*) in a specification that also included a set of variables that capture the resource endowments. The results are centred on the analysis of β_2 coefficient associated with the *idiodem* variable (defined as the deviation in a country's expenditure on a given good with respect to the expenditure of the rest of the world). If it can be shown that $\beta_2 > 1$, then this is considered evidence in support of the existence of the 'home market effect'. Davis and Weinstein's (1999, 2003) results provided empirical support for the presence of the 'home market effect' at the international level. A critique of these results is available in Combes, Mayer and Thisse (2008).

These studies have opened up the path for the analysis of regional specialisation and industrial concentration in Spain in the framework of New Economic Geography. However, as in the case of Kim (1995), the study of the determinants of industrial location has not been based on sufficiently solid theoretical foundations, since the econometric specifications remain distant from the theoretical models[68]. A number of studies have tried to overcome this weakness with the empirical testing of the theoretical predictions of NEG.

Recently, Martínez-Galarraga, Paluzie, Pons and Tirado (2008) have provided evidence in support of the existence of an 'agglomeration effect' linking the spatial density of economic activity and interregional differences in the productivity of industrial labour in Spain for the period 1860 to 1999. In line with Ciccone and Hall (1996) and Ciccone (2002), the study showed that the estimated elasticity of employment density with respect to labour productivity, as the agglomeration effect has been defined, played a key role from the middle of the 19th century, i.e., during the early stages of industrialisation. However, its evolution presents a progressive decline over time and in the final period considered (1985-1999) the agglomeration effect is no longer significant.

Similarly, Combes, Lafourcade, Thisse and Toutain (2008) have offered a long-run perspective of the location of industrial activity in France at the territorial level of the *départements*. First, they showed that the fall in transport costs since the mid-19th century led to a bell-shaped evolution in the spatial distribution of activity in the manufacturing and services sectors, which underwent an increase in concentration between 1860 and 1930, before dispersing between 1930 and the year 2000. On the other hand, they also found evidence of an agglomeration effect in the French economy between 1860 and 2000. The intensification of economic density led in turn to an increase in labour productivity in both manufacturing and services. In a first phase between 1860 and 1930, this agglomeration effect was linked to market potential, while between 1930 and 2000, it could be explained by the difference in educational attainment recorded in the *départements*. The parameters estimated in this study suggested that doubling the employment density in a French *département* would result in labour productivity gains of around 5%. This result is in line with those obtained for industry in the Spanish provinces, which fell from 5 to 3% in the period between 1860 and 1985 (Martínez-Galarraga, Paluzie, Pons and Tirado, 2008). The results of these long-run analyses

[68] An exception is Rosés' (2003) study, based, as we have seen, on Davis and Weinstein (1999, 2003).

coincide with the pioneering studies of Ciccone and Hall (1996) and Ciccone (2002). In the former, the effect of doubling the employment density in a US county at the end of the 1980s was a 6% increase in labour productivity. In the latter, Ciccone (2002) applied the same empirical strategy with a sample of five European countries (France, Germany, Italy, Spain and the United Kingdom) at the NUTS3 regional level at the beginning of the 1990s. The analysis provided slightly lower values than those for the US fluctuating between 4.5 and 5%.

Another notable area for the empirical testing of the theoretical predictions of NEG has been the verification of Krugman's (1991) wage equation. For the Spanish case, following Hanson (2005), studies have analysed whether there is a relationship between provincial market potential and their nominal wages, as can be derived from the NEG models. The existence of this type of relationship in which wages are higher in regions with greater market potential constitutes an unequivocal sign of the presence of an effect associated with domestic market size. Paluzie, Pons and Tirado (2009) have shown that industrial nominal wages for the Spanish provinces depended positively on their proximity to large markets in the period 1955-1995.

In turn, they showed, by estimating a reduced form of the equilibrium-wage equation, evidence in support of the presence of a spatial wage structure in the interwar years (Tirado, Pons and Paluzie, 2009)[69].This last study not only verified the existence of a wage gradient centred on Barcelona (the peninsula's leading industrial centre in the interwar years), but it also examined whether this gradient varied at a time when protectionist policies were intensified following the introduction of the Cambó tariff in 1922. It is, therefore, an example of an opposite case to that studied by Hanson (1997) in relation to the Mexican economy, which was characterised by economic liberalisation from the mid-1980s. Here, during the 1920s, evidence is found for a weakening of the wage gradient centred on Barcelona, a province lying near to the French border, and therefore, close to foreign markets. Furthermore, the authors suggest that this shift towards protectionist trade policies might explain the relative rise in the early decades of the 20th century of inland zones such as Madrid, that were better placed, thanks to their location in the geographical centre of the peninsula, to supply the protected domestic market[70].

[69] This hypothesis had been previously tested in Tirado, Pons and Paluzie (2003, 2006).

[70] In the theoretical debate, these results lie close to the theoretical predictions derived from the model proposed by Crozet and Koenig (2004a). See section 2.2.2.2 in this chapter.

Elsewhere, Pons, Paluzie, Silvestre and Tirado (2007), following Crozet (2004), have verified the presence of forward linkages in the internal migrations between Spain's provinces in the interwar years. The study established a direct relationship between workers' location decisions and the host regions' market potential. Yet, although the Spanish workers were attracted by industrial agglomerations, this attraction was limited to relatively close-lying zones. The high costs of migration would seem to have reduced the intensity of migratory flows and would have been a key factor in the workers' location decision. This would explain the apparently low intensity of internal migrations in Spain until the 1920s and the geography of these migrations in the interwar years. The migratory flows to the main industrial centres did not originate from the poorest regions in the south of the peninsula which lay furthest from these industrial centres and this was due to the migration costs that grew in relation to the distance that the workers had to travel.

Paluzie, Pons, Silvestre and Tirado (2009) conducted the same type of analysis for three different periods: the 1920s, the 1960s and the beginning of the 21st century. Their results showed that a forward linkage was present both in the periods of concentration and in the phases of spatial dispersion of economic activity after the 1970s. Spain's internal migrations increased in the 1950s and again, more markedly, in the 1960s and the beginning of the 1970s. Furthermore, during this period the migration did originate from the most economically backward regions (Andalusia, Extremadura and Castilla-La-Mancha). After this date, however, the intensity of migrations fell and the spatial pattern changed due to a weakening in the attraction of those regions that had traditionally been the recipients of migrant workers. In this case, the loss in weight suffered by the industrial sector at the expense of the services sector as the sector with the capacity to generate migratory flows, the extension of the territory that defines a region's market potential, and the reduction in the explanatory power of migration costs explain the changes in the migratory model during the 20th century.

These last two sets of studies have sought to test empirically, in the context of the Spanish economy, first, the wage equation, i.e., the existence of higher wages in regions that have greater market potential resulting from the agglomeration of manufacturers in core regions (backward linkages), and second, the attraction of these wages for generating migratory flows of workers (forward linkages). These are some of the centripetal forces stressed by New Economic Geography (Krugman, 1991) and which are responsible for agglomeration in the early stages of economic development. However, these might not be the only forces operating. Differences in the comparative advantages between regions in terms of their natural resources and factor endowments, as highlighted in the neoclassical theory of international trade (Heckscher-Ohlin model), might simultaneously shape the spatial distribution of manufacturing. This is a question that was first raised in the pioneering study for the US undertaken by Kim (1995), who included variables in the regression associated with factor endowments as well as with NEG, concluding, as we have seen earlier, that, despite the significance of scale economies, the weight of the explanation lay in the endowment elements stressed by Traditional Trade Theory.

However, the approach adopted by Kim (1995) in his analysis of long-term industrial location in the US is not without its problems[71]. First, there is the recurrent problem of endogeneity attributable to the circular causation typical of the processes of agglomeration that NEG seeks to describe. A further weakness lies in the possibility that additional variables to those considered in this type of analysis, and therefore omitted from the regressions, can affect the spatial distribution of manufacturing. Among these are variables related to factor endowments and the intensity of their use and the presence of intermediate goods. Kim (1995) sought to solve this problem by applying industry and time fixed effects to his panel data[72]. In addition, Kim did not test directly the role of market access in industrial location decisions. Finally, the regression estimated by Kim is a non-structural or reduced specification, which is not derived directly from a theoretical model.

The approach adopted by Midelfart-Knarvik, Overman and Venables (2000) and by Midelfart-Knarvik, Overman, Redding and Venables (2002) in their analyses of the European Union allowed tackling some of the problems that emerge from earlier studies such as Kim's (1995). On the one hand, they use a structural specification, i.e., one derived directly from a theoretical model. On the other, the number of variables considered is wider. It is not a regression that includes plant size or raw material intensity in industries on an index of production location as explanatory variables, but rather it takes into consideration region and industry characteristics.

[71] See Combes, Mayer and Thisse (2008), chapter 11.

[72] In order to deal with the potential problem of omitted variables, Kim incorporated to the equation *industry* and *time fixed effects*. *Industry fixed effects* allow controlling for the variables that are constant over time but specific of each industry. However, this assumption becomes very restrictive when a long time period, as in Kim (1995), is considered. In the second case, the effects would be constant across industries but specific of each year considered assuming therefore that the omitted variables are constant over time. Thus, *time fixed effects* control for macroeconomic shocks assuming that they would affect equally to all the industries.

These studies are based on the idea that the effects of factor endowments and of economic geography on industrial location can be combined and that they are not excluding[73]. Thus, regions are heterogeneous in various characteristics such as their natural resources endowment or their proximity to the markets. And in the same way, industries differ in their attributes, for example, in the use of production factors, natural resources, or skilled labour, the size of the establishments and in their dependence on intermediate inputs, to mention just some. The most interesting aspect of this model is that it considers both factor endowments and NEG in terms of a series of interactions between region and industry characteristics, which together capture the role of both explanations in the determination of regional specialisation and industrial location allowing the relative importance of the Heckscher-Ohlin-type arguments against the NEG forces to be quantified.

Adopting this new approach, Klein and Crafts (2009) have questioned some of the conclusions previously drawn by Kim for the US. In this case, their study focused on the years between 1880 and 1920. For these authors, the emergence of the manufacturing belt in the US during this period was linked above all to the interaction of the forces stressed by NEG. Specifically, of the six interactions considered, only one of the three referring to Heckscher-Ohlin (HO) factor endowments was significant: agriculture endowment, and even then, only prior to 1900. By contrast, the skill level of the workforce and coal abundance were not statistically significant. Yet, the three NEG interactions related to market potential were significant determinants of industrial location. The interaction with scale economies appeared as a decisive factor throughout the period, to which are gradually added the intensity of sales to industry and the intensity of intermediate input use, whose combined effect in 1920 was greater than that of all the other variables. Thus, using a more appropriate methodology for the analysis of industrial location patterns, Klein and Crafts (2009) offered an alternative interpretation to that provided by Kim for explaining the development of the industrial belt in the northeast of the US, in which NEG factors are fundamental. In addition, they remarkably improved the empirical methodology developed in the original paper by Midelfart-Knarvik, Overman, Redding and Venables (2002) therefore increasing the robustness of the econometric results.

However, Klein and Crafts' (2009) study is the most recent in a series of studies that have applied the approach set out by Midelfart-Knarvik, Overman, Redding and

[73] Something that Davis and Weinstein (1999, 2003) also sought to capture.

Venables (2002) with a historical perspective. The first historical exercise was conducted by Wolf (2004, 2007). The reunification of Poland following World War I, and the subsequent integration of the Polish domestic market, meant that Poland in the interwar years provided a particularly appropriate case for testing the theoretical predictions of NEG. Thus, Wolf studied the determinants that lay behind the changes in the location of industry after the 1918 reunification based on the proposals contained in Midelfart-Knarvik, Overman and Venables (2000). The estimation of the model provided values for the interaction variables that showed that both types of mechanism, HO and NEG, acted simultaneously in the period of study. On the one hand, the resource endowments variables played a key role. Specifically, the availability of skilled labour was the mechanism that dominated industrial location, with an intensity that moreover increased over the period (1926 to 1934). Yet, the results showed that the interaction between market potential and demand for intermediate inputs corrected for plant size was also significant. This points to the existence of NEG-type forces, albeit that on this occasion, the impact of these forces was stable over time[74].

This study was followed by those conducted by Crafts and Mulatu (2005, 2006), who focused on the case of Britain between the end of the 19th century and the first decades of the 20th. The study of industrial location in Great Britain, where the Industrial Revolution had its roots, in a period marked by a sharp fall in transport costs, was undertaken, again, by estimating an equation based on Midelfart-Knarvik, Overman and Venables (2000). The coefficients of the interactions linked to factor endowments corroborated the importance of Heckscher-Ohlin elements when explaining the location of British industry in this period. Thus, the location pattern of British industry responded to the traditional variables of factor endowments (agriculture, human capital and coal abundance). However, these forces were accentuated by the NEG forces, since the interaction between market potential and scale economies also appeared as a significant variable. However, the scale effects weakened over time to the extent that they ceased to be significant in the observation corresponding to 1931. Finally, Crafts and Mulatu (2005, 2006) stressed the strong impact obtained for human capital endowment.

As seen above, Kim (1995) explained the bell-shaped evolution observed in the geographical concentration of manufacturing in the US in terms of factor endowments, limiting, therefore, the role of increasing returns in the analysis of the economic

[74] These results in Wolf (2004) were updated in Wolf (2007). For a more detailed description, see section 4.1 in chapter 4.

geography of the United States. However, Klein and Crafts (2009), following Midelfart-Knarvik, Overman, Redding and Venables (2002) and, therefore, adopting a more suitable approach for the analysis of the relative importance of Heckscher-Ohlin-type arguments and NEG forces, obtained different results. In their opinion, the forces highlighted by NEG explain the intensification of the concentration of manufacturing in the US centred on the industrial belt in the northeast of the country.

In the case of Spain, the marked geographical concentration of manufacturing in a small number of regions during the second half of the 19th century has been explained by Tirado, Paluzie and Pons (2002), following the proposal outlined by Kim (1995). The possibility of studying this phenomenon with an alternative empirical strategy, as in the case of the US, might be useful to validate the results obtained by these authors in order to examine in greater detail the determinants of industrial location in the early stages of Spanish industrialisation and so as to broaden the time period considered. This means that a similar exercise to that undertaken by Midelfart-Knarvik, Overman, Redding and Venables (2002) can be carried out to examine the Spanish case, an exercise that has already been applied to different historical periods such as interwar Poland, Victorian Britain and the US (1880-1920).

The literature review undertaken above highlights the key role played by market potential in the location decisions of both firms and workers, and as such, in the spatial agglomeration of production stressed in the NEG models. In turn, the uneven geographical distribution of economic activity has direct consequences on regional disparities. Therefore in chapter 4, the determinants of industrial location in Spain between 1860 and 1930 are studied following Midelfart-Knarvik, Overman, Redding and Venables (2002), in a period during which the process of industrialisation advanced considerably at the same time as the domestic market achieved greater integration. Next, in chapter 5, the focus moves from industry to GDP per capita and the relationship between market potential and regional inequality in the early stages of economic development in Spain is explored following Ottaviano and Pinelli (2006). However, before undertaking these analyses in chapters 4 and 5, it is first needed to produce estimates of the regional market potential for the Spanish regions, which in turn makes it necessary to obtain an estimation of regional GDP at a sufficiently disaggregated geographical scale. This is the goal set for the following chapter.

Chapter 3. Market potential in the Spanish provinces, 1867-1930

3.1. Introduction

In chapter 2 the significance of market potential within the New Economic Geography approach has been highlighted. In its theoretical models, the size of a particular location as well as its proximity to large markets (characterised by good access to demand) are key to the localisation decisions taken by firms and workers alike, since they favour the emergence of the agglomeration forces described by Krugman (1991). Similarly, when considering multi-regional models, the capacity of different locations to attract firms and workers varies according to their relative position in space. Moreover, the evolution in market access over time is a further factor to take into consideration since, as stressed, such access can vary for a number of reasons. The public policies adopted by governments can have an impact on trade costs[75]. On the one hand, investment in infrastructure will bring about changes - often presenting a regionally asymmetric pattern - in the quantity and quality of the lines of communication, thereby affecting transport costs. On the other hand, the market potential can also vary as a consequence of changes in trade policy or, for example, as a result of the decision taken by a group of countries to initiate a process of economic integration, such as that which led to the creation of the European Union.

Likewise, the previous chapter has served to recognise the importance of having recourse to an accurate indicator of market access in order to carry out empirical studies within a NEG framework. Market potential has been an essential variable in those studies that have sought to test for the existence of backward linkages (Head and Mayer, 2004a)

---

[75] They might also be affected by factors not directly linked to public policies. For example, taking a historical perspective, the emergence of new modes of transport and improvements resulting from the application of new technologies both affected transport costs. This was the case throughout the 19th and the early 20th centuries with the development of the steamship and the railway and, later, with the invention of the combustion engine that promoted the use of road transport.

and forward linkages (Crozet, 2004), as well as in those that have focused on providing an analysis of per capita income inequality between countries (Redding and Venables, 2004), regional wages (Hanson, 1998, 2005) or the effects of the enlargement of Europe to include the countries from the east of the continent (Brülhart, Crozet and Koenig, 2004), to mention just a few of the most relevant studies.

A further aspect that has been the focus of study is the question as to the most appropriate way of measuring market potential. Studies range from Harris' (1954) early classic to recent proposals for structural estimates that are derived directly from NEG models[76]. However, as we shall see below, the two measures are directly linked and are not, therefore, so different from each other.

On the one hand, a first alternative involves calculating the market potential following Redding and Venables (2004), i.e., based on the coefficients estimated in a gravity trade equation to construct the market access (*MA*) and supplier access (*SA*) variables. In this line, among the studies of economic history reviewed in which the market potential is obtained following this approach, Wolf's (2007) work stands out. This author constructed the measure of market potential using a gravity equation based on interregional trade in Poland between the wars.

Yet, the alternative to obtaining an structural estimate of market potential based on New Economic Geography models requires a volume of data regarding bilateral trade flows that all too frequently is not available. This difficulty becomes even more apparent when undertaking regional studies from a historical perspective, such as the one proposed here. For example, in the calculation of market potential estimates in Great Britain, Crafts (2005b) had to face the absence of data for interregional trade. In this instance, the author, in common with a great number of empirical studies conducted within NEG, turned to the market potential equation defined by Harris (1954). This equation has been adopted in a large number of studies published before the emergence of NEG (Clark, Wilson and Bradley, 1969; Keeble, Owens and Thompson, 1982), as well as in those conducted within this analytical framework.

Similarly, in the case of Spain, it is not possible to know the trade flows between provinces and regions in the second half of the 19th century and the first decades of the

[76] On this very question, as discussed in chapter 2, section 2.2.2.2, the evidence as to which indicator provides the best results is inconclusive as it varies according to the specific analyses. In this sense, Klein and Crafts (2009) in their study of the determinants of industrial location in the US obtain similar results when the market potential is calculated using Harris' equation and an alternative definition based on the distance coefficients from a gravity equation estimated for international US trade.

20th due to the lack of such information. Having ruled out, therefore, the possibility of constructing a measure of market access using a gravity equation as suggested by Redding and Venables (2004), market potential will be calculated following Crafts (2005b) on the basis of Harris' equation.

The market potential estimation undertaken by Crafts (2005b) for the British regions between 1871 and 1931 has paved the way for the study of this variable in different periods of history. Based on Harris' proposal, and bearing in mind the changes introduced by Keeble, Owens and Thompson (1982) to complete the original indicator, Crafts (2005b) considered in his calculations both the internal market potential in the British regions and their foreign market potential, in which are included their main trading partners. In this sense, the path opened up by Crafts in the historical application of the market potential *à la Harris* has recently been followed by Schulze (2007) in calculating this indicator for the regions that made up the Austro-Hungarian Empire between 1870 and 1910.

Thus, the aim of this chapter is to construct an indicator of market access based on the market potential equation defined by Harris (1954). The spatial unit chosen is the Spanish provinces, that is, a NUTS3 level of disaggregation according to the EU nomenclature for statistical territorial units. As for the period of the study, the analysis focuses on the years between the second half of the 19th and the first decades of the 20th centuries, a period in which the Spanish economy was in its early stages of economic development. Moreover, the integration of the domestic market was completed in these years, while at the same time the Peninsula began to witness a marked increase in the spatial concentration in its manufacturing sector (Paluzie, Pons and Tirado, 2004).

The availability of market potential estimates, as already stressed, is essential for the exercises that will be carried out in the following chapters: first, when studying, in Chapter 4, the determinants of industrial location in Spain prior to the Civil War following Midelfart-Knarvik, Overman, Redding and Venables (2002), and, then, in Chapter 5, when analysing the causes of regional inequality in terms of the differences in the per capita provincial income in Spain in the same period based on Ottaviano and Pinelli (2006).

The rest of this chapter is organised as follows. First, the relationship between the measure of market access derived from the NEG models and the market potential equation defined by Harris (1954) is clarified. Then, the construction of the market potential in the Spanish provinces for the years 1867, 1900, 1914 and 1930 is detailed.

This is the major contribution of this chapter. Finally, the results are presented and briefly described, and the main conclusions summarised.

3.2. Theoretical foundations underpinning market potential

Interest in the relationship between market access and industrial location is longstanding among geographers and economists. Harris' (1954) pioneering work sought to explain the creation of the industrial belt in the northeast of the United States and its persistence over time. Harris held that the area had experienced a process of industrial concentration characterised by a circular causation that was similar to that subsequently proposed by New Economic Geography. According to Harris (1954), the north-eastern areas of the country enjoyed an advantage in terms of better market access, which would have attracted manufacturers to these locations placed near the largest markets. In turn, the size of these markets would have augmented due to the concentration of manufacturers[77]. In order to analyse the importance of markets as an industrial location factor in the United States, Harris proposed an index for measuring market accessibility based on the following formula:

$$P = \sum \left(\frac{M}{d} \right) \qquad (3.1)$$

where market potential *(P)* is defined as the summation of markets accessible from a point divided by the distance to that point, where *'M'* is a measure of the economic activity in each area, and *'d'* the distance or the transport costs between areas and regions[78].

Yet, this measure of market access suggested by geographers and widely adopted by economists is an *ad hoc* indicator and is not built upon a solid theoretical foundation. As Krugman pointed out: "*Market potential analyses have been a staple of geographical discussion, especially in Europe (see, for example, Keeble, Owens and Thompson, 1982). The main theoretical weakness of the approach is a lack of microeconomic foundations:*

[77] "*...manufacturing has developed partly in areas or regions of largest markets and in turn the size of these markets has been augmented and other favorable conditions have been developed by the very growth of this industry*". Harris (1954), p. 315.

[78] "*The term market potential, suggested by Colin Clark, is analogous to that of population potential as proposed and mapped by John Q. Stewart. It is an abstract index of the intensity of possible contact with markets. The concept is derived ultimately from physics, in which similar formulas are used in calculating the strength of a field, whether electrical, magnetic, or gravitational*". Harris (1954), p. 321.

while it is plausible that some index of market potential should help determine production location, there is no explicit representation of how the market actually works"[79]. However, the advances made by New Economic Geography models can help overcome this lack of theoretical link between market access, the spatial localisation of economic activity and regional development. Thus, departing from a NEG model a mathematical expression can be derived. By adopting a series of assumptions, it can be proved that this expression is comparable to the original market potential equation proposed by Harris (1954). Therefore, it allows us to provide a theoretical foundation to the Harris' equation for market potential[80].

Combes, Mayer and Thisse (2008) focus their analysis on the determinants of industrial location in a context in which, as noted above, activities with scale economies tend to establish themselves in regions that enjoy good market access, since it is these locations that offer the greatest potential benefits. Thus, the study of the benefits accruing to a firm in a New Economic Geography theoretical framework allows deriving an expression for the Real Market Potential (*RMP*) from which it is possible to establish a relationship with Harris' (1954) equation. The NEG models show that, in equilibrium, the gross benefits of exploiting a firm (π^*_{rs}) are expressed as follows:

$$\pi^*_{rs} = (p^*_r - m_r)\tau_{rs} q^*_{rs} = m_r \frac{\tau_{rs} q^*_{rs}}{\sigma - 1}$$

where *r* and *s* represent the regions or countries, p_r refers to the price of a variety sold by a firm located in *r*, m_r denotes the marginal cost of production, q_{rs} the quantity that a firm sells in market *s*, τ_{rs} are the iceberg-type transport costs payable on a good on the route from *r* to *s*, and σ is the elasticity of substitution between any two varieties, an inverse index of product differentiation.

On the one hand, the equilibrium price is expressed as $p^*_{rs} = \tau_{rs} p^*_r = \tau_{rs} m_r \sigma / \sigma - 1$, while in the short-run, when the number of firms is exogenous and the benefits are positive, the quantity q^*_{rs} is determined using a CES-type demand function that adopts the following form:

[79] Krugman (1992b), p. 7.

[80] Here, the explanation follows Combes, Mayer and Thisse (2008). An alternative approach for deriving the market potential function using the wage equation is found in Krugman (1992b).

$$q_{rs}^{*} = (p_r^{*}\tau_{rs})^{-\sigma}\mu_s Y_s P_s^{\sigma-1}$$

where μ_s is a parameter representing the share of the good considered in the consumption of region s, Y_s denotes income in region s, and P_s is the CES-type price index in s, according to the following expression:

$$P_s = \left[\sum_{r=1}^{R} n_r \left(p_r^{*}\tau_{rs}\right)^{-(\sigma-1)}\right]^{-1/(\sigma-1)}$$

Taking this into account, the total net profit of a firm located in region r can be obtained by subtracting the specific fixed costs of each plant *(F_r)* from the gross profit obtained previously (π_{rs}^{*}), so that:

$$\Pi_r^{*} = \sum_s \pi_{rs}^{*} - F_r = cm_r^{-(\sigma-1)} RMP_r - F_r \qquad (3.2)$$

In this instance, $c = \sigma^{-\sigma}/(\sigma-1)^{-(\sigma-1)}$ and the abbreviation RMP_r corresponds to the real market potential of region r, which would be given by the expression:

$$RMP_r \equiv \sum \phi_{rs}\mu_s Y_s P_s^{\sigma-1} \qquad (3.3)$$

where the term ϕ_{rs} measures the accessibility of the goods from r into market s as a function of transport costs, which are represented by $\phi_{rs} \equiv \tau_{rs}^{-(\sigma-1)}$.

Once this expression of Real Market Potential has been derived from a NEG model, it is possible to establish the relationship between the latter and the market potential equation defined by Harris. To do this, three assumptions have to be made. First, it has to be accepted that $\phi_{rs} = d_{rs}^{-\delta}$, where d_{rs} is the distance between locations r and s, and the exponent δ corresponds to the estimated parameter for distance in the gravity equations that analyse the determinants of the volume of bilateral trade. Thus, it should be stressed that the estimation of gravity equations usually generates values that are close to one for parameter δ. Recently, Disdier and Head (2008) have tried to quantify the magnitude of the effect of distance on international trade by compiling a total of 1,467 coefficients obtained for distance derived from the estimation of gravity equations in different studies. The authors obtain a measure for the coefficient for distance of 0.9 for the post-World War II period, so that a 10% increase in distance between two countries would reduce trade between them by 9%. This is a very similar finding to those reported in other studies, for example in Mayer (2008) for an international sample of countries between 1860 and 2003[81]. Other studies, including Hummels (1999), Anderson and Van Wincoop (2003), and Redding and Venables (2004), corroborate the proximity of the distance coefficient to one in recent times. Therefore, the available empirical evidence supports the assumption regarding a coefficient for distance close to one, so that, as Head and Mayer (2004a) claim, Harris' assumption of an inverse distance relation, where $\phi_{rs} = 1/d_{rs}$, appears to be a reasonable approximation to reality.

However, the impact of distance on trade has increased over time[82]. In the case of Spain, the focus of this thesis, the period considered includes the second half of the 19th and the first few decades of the 20th centuries. In a similar vein, Estevadeordal, Frantz and Taylor (2003) have analysed the period between 1870 and 1939, reporting distance coefficients that are slightly lower, oscillating between -0.64 and -0.79 depending on the specifications employed[83].

Second, it is assumed that the share of each good within the total consumption does not vary between regions, so that $\mu=1$[84]. Finally, an important aspect is the inclusion

---

[81] "*The average coefficients on trade costs are very much in line with existing findings. The coefficient for distance is very close to -1*". Mayer (2008), p. 6.

[82] In line with Combes, Mayer and Thisse (2008), p. 111, the impact of distance on trade has increased from a value of approximately 0.5 in 1870 to 1.5 in the year 2000. This rise, which would seem counterintuitive in a context marked by increasing globalisation, is a recurring result in the empirical literature, which also shows a considerable increase in the impact of distance on trade in the post-World War II era. The reason typically given to explain this evolution is that distance is taken as a proxy of trade costs, while the latter can be affected by other factors. On the one hand, transport costs are linked to elements in the physical geography (access to the sea, the relief, as well as border effects). On the other hand, it is necessary also to consider the trade policy (tariffs and non-tariff barriers), information costs (existence of business and social networks) and cultural differences (sharing a common language or otherwise), since all these factors can affect trade costs.

[83] Flandreau and Maurel (2005) estimate a value for distance between 0.79 and 0.99 in Europe at the end of the 19th century. López-Córdova and Meissner (2003), by contrast, report a value of 0.661 for the period 1870 to 1910 in an international sample that includes between 14 and 28 countries depending on the year. For this same period, Mitchener and Weidenmier (2008) offer an estimation for the distance coefficient around 0.56. In the interwar years, Eichengreen and Irwin (1995) report a decreasing impact with values that vary between 0.51 and 0.78 in 1928 to 0.33 and 0.57 in 1938. However, Jacks, Meissner and Novy (2009) obtain lower values, between 0.31 and 0.38 for the period 1870-1913, and even lower ones for interwar years (1921-1939), ranging from 0.15 to 0.20.

[84] "*This simplifying assumption may be deemed acceptable when working with the consumption of final goods. However, regarding the consumption of intermediate goods, this assumption becomes more problematic, as it implies that either all sectors consume the same amount of each factor, or regional sectoral compositions are the same*". Combes, Mayer and Thisse (2008), p. 305.

in the Real Market Potential (*RMP*) of the price index $P_s^{\sigma-1}$, which is missing in Harris' equation, assuming, therefore, that there is no variation in the price indices from one region to another. Bearing these three assumptions in mind, it is possible to obtain, using the expression of Real Market Potential, Harris' (1954) equation.

3.3. The construction of provincial market potential in Spain, 1867-1930

3.3.1. Definition and selection of provincial and foreign 'nodes'

Market accessibility or market potential is measured using Harris' (1954) equation, in accordance with the following expression:

$$MP_r = \sum_s \frac{M_s}{d_{rs}} \qquad (3.4)$$

Based on this equation, the market potential of a province *r* can be expressed as the ratio between M_s, a measure of economic activity in province *s* (typically GDP), and d_{rs}, the distance or bilateral transport costs between *r* and *s*[85]. This indicator can be interpreted as the volume of economic activity to which a region has access after having subtracted the necessary transport costs to cover the distance to reach all the other regions. The total market potential, in turn, is divided between the internal market potential and the foreign market potential. In the case of the former, the economic potential of any Spanish province depends on the GDP of each of the other provinces adjusted by its proximity to these provinces, measured in terms of distance or, as in this case, in terms of transport costs. Likewise, it is also necessary to consider the market potential of each province, i.e. its self-potential. Besides, a province's foreign market potential must be added to its internal market potential. In this case, as outlined in detail below, the size of the external markets is considered in terms of GDP, distances and additional tariff costs.

[85] The measurement of transport costs has been and remains the subject of much debate. The geodesic, straight-line distance, real distance as a function of the available infrastructure, distance measured in time (Hummels, 2001), or the transport costs that include the distances and the freight rates, are the various alternatives used in empirical studies. A review of the literature from an NEG perspective can be found in Combes and Lafourcade (2005) and Lafourcade and Thisse (2008).

The period analysed extends from the second half of the 19th century to the first third of the 20th century, and includes specifically the years 1867, 1900, 1914 and 1930. The study is conducted for the Spanish provinces, that is, for a level of disaggregation that corresponds to the NUTS3 statistical territorial units of the European Union. However, for strictly geographical reasons, the territories lying outside the peninsula have not been included, so that the Balearic Islands, the Canary Islands and the autonomous cities of Ceuta and Melilla are not considered in the analysis. Consequently, the study includes a total of 47 provinces.

Calculating the provincial market potential requires the adoption of a city or a 'node' within each province to serve as its unit of reference. In this way this node is assigned the total volume of economic activity generated within that province. Therefore, the smaller the selected territory is, the less restrictive this assumption will be. The node assigned to each province is, in most instances, its administrative capital. However, there are some exceptions. In the case of the coastal provinces of Murcia, Oviedo and Pontevedra, the provincial capitals do not lie on the coast. The geography of the Iberian Peninsula is such that a good number of its provinces enjoy direct access to the sea, a characteristic that necessarily influences its transport costs, and not just those with the other Spanish provinces but also those with foreign ports. In this sense, direct sea access is a highly relevant geographical factor since it translates as a locational advantage resulting in lower transport costs and enhanced market access (Rappaport and Sachs, 2003). In order, therefore, to capture the coastal location of these provinces, alternative provincial nodes are chosen. These nodes are in all cases major centres of population and economic activity within the province and, furthermore, they possess a commercial port: Cartagena, Gijón and Vigo, respectively. By contrast, in the case of the provinces of Girona, Granada, and Lugo, three provinces with a coast but whose capital, once more, is not beside the sea, there do not exist other centres of important activity or any large ports in the coast. Thus, these three provinces are considered as inland provinces[86].

As for the exterior nodes, the first step is to select the countries that played an important role as trading partners for the Spanish economy. This selection is based on information regarding the geographical distribution of Spanish exports between the mid-19th century and the 1930s, which reveals a high concentration in the country's export

[86] The city of Girona is 35 km by road from the port of Sant Feliu de Guíxols, and approximately 50 km from the ports of Palamòs and Blanes. In this case, the rapid rail connection with Barcelona, which is just 100 km away, does not penalise Girona significantly for being an inland province. In the case of Granada, the distance to its closest port, Motril, is 68.5 km by road. Similarly, Lugo is 107 km from Ribadeo.

markets[87]. France and Great Britain together constituted the market for more than 40% of Spanish exports in the four years selected, reaching a maximum at the beginning of the period studied here with 57.1% of all exports. On the basis of this information, the decision was taken to include as foreign markets those countries that accounted for at least 5% of Spain's exports[88]. Thus, four countries are included in the calculation of the foreign market potential: Great Britain, France, Germany and the United States[89].

Having decided on which countries to include in the sample, the next step involves selecting a node to represent each of the four markets. In the case of Great Britain, London, the capital and economic centre of the country, is used[90]. For the USA, the node selected is New York, while in the case of Germany, for questions of geographical access and the size of its port, the city of Hamburg is taken as the node. However, in the case of France the way of proceeding must differ. As a consequence of its geographical location in relation to that of the Iberian Peninsula, the French market can be accessed from Spain both via the Atlantic and the Mediterranean seaboards. Therefore, localizing the French market in a single node would mean penalising the regions on one or other of these two seaboards. For this reason, the French market is divided so as to capture the various routes along which the Spanish provinces can access it. Thus, three regional nodes are considered: Le Havre and Nantes on the Atlantic seaboard and Marseille on the Mediterranean[91].

The calculation of the market potential of the Spanish provinces can be disaggregated into two components: the internal market potential to which each

[87] Prados de la Escosura (1982) and Tena (2005). See table A1 in the Appendix.
[88] Two exceptions include Cuba, a market that received a high percentage of Spanish exports, above all in the mid-19th century (18.5% of the total), and Argentina, whose market exceeded the 5% threshold in the years immediately before World War I. They are excluded here because both countries do not figure in the sample of countries for which Prados de la Escosura (2000) offers GDP estimations at current prices. However, it ought to be the case that the limited size of their markets and, especially, the great distance separating them from the Peninsula would minimize the cost of their exclusion.
[89] These countries account for 62.4% of Spanish exports in 1865/69, 57.8% in 1895/99, 58% in 1910/13 and 58.9% in 1931/35. The sample is smaller than the 14 countries considered by Crafts (2005b) for the British case (Austria-Hungary, Belgium, Denmark, France, Germany, India, Ireland, Italy, The Netherlands, Norway, Portugal, Spain, Sweden and the USA), and the 15 countries that Schulze (2007) included in his study of Austria-Hungary – from the countries above, he excluded Norway and included Russia, Switzerland and Turkey.
[90] Crafts' (2005a) study gives disaggregated information for regional GDP in Great Britain. Hence, it is possible to calculate the market potential, not by assigning all the economic activity in Britain to London but rather by distributing it between the nodes selected for each of the 12 regions. However, this approach gives similar results to those obtained when just considering London as the node for the whole of the UK.
[91] Below the figures corresponding to the foreign GDP are given in detail and this division of the French market is further explained.

province's self-potential needs to be incorporated, and the foreign market potential. Next, the detailed information needed for obtaining each of these two components is provided.

3.3.2. Internal market potential

Determining the internal market potential requires access to the following information: a) the GDP of the Spanish provinces, b) the transport costs between the provinces (inter-provincial internal market potential) and c) the self-potential of each province.

3.3.2.1. The GDP of the Spanish provinces. The provincial GDP figures are obtained adopting the methodology outlined in Geary and Stark (2002). They suggested an estimation of the GDP of the various countries making up the United Kingdom before World War I that consists in distributing the British GDP on the basis of the wage income. This methodology, applied to the Spanish case, is used to obtain provincial GDP estimates for the years 1860, 1900, 1914 and 1930.

The methodology proposed by Geary and Stark (2002) for the estimation of output is based on two variables: labour force and productivity, grouped by sector (agriculture, industry and services), and by country. When applied to Spain, the total GDP of the Spanish economy is the sum of provincial GDPs

$$Y_{ESP} = \sum^{i} Y_i \tag{3.5}$$

where Y_i is the provincial GDP defined as

$$Y_i = \sum^{j} y_{ij} L_{ij} \tag{3.6}$$

y_{ij} being the output or the average added value per worker in each province i, in sector j, and L_{ij} the number of workers in each province and sector. As no data are available for y_{ij}, this value is proxied by taking the Spanish sectoral output per worker (y_j), assuming that provincial labour productivity in each sector is reflected by its wage relative to the

Spanish average (w_{ij}/w_j). Therefore, we can assume that the provincial GDP will be given by

$$Y_i = \sum^{j} \left[y_j \beta_j \left(\frac{w_{ij}}{w_j} \right) \right] L_{ij} \tag{3.7}$$

where, as suggested by these authors, w_{ij} is the wage paid in the province i in sector j, w_j is the Spanish wage in each sector j, and β_j is a scalar which preserves the relative province differences but scales the absolute values so that the provincial total for each sector adds up to the known total for Spain. This model of indirect estimation, based on wage income, allows an estimation of GDP by province at factor cost, in current pesetas. This is of great interest for this study. The calculation of market potential at a specific point in time requires data expressed at current price levels, as it is these prices that guide the agents' decision making.

The equation described above requires the following data in order to estimate the provincial GDP: a series of the sectoral structure of employment at the provincial level, an estimate of output per worker in each sector for Spain, and finally a series of provincial wages per sector.

First, the series concerning employment by sector in each province are compiled from the information provided by the 1860, 1900, 1910 and 1930 Population Censuses[92]. Second, the output per worker in Spain requires the data for: a) the sectoral output at factor cost, which is obtained from Prados de la Escosura (2003); and b) the total amount of workers per sector in the Spanish economy – once more, this information is found in the respective Population Censuses. The third data set, nominal wages by province, presents greater difficulties due to the relative shortage of data regarding wages in the period studied. In fact, the restricted availability of nominal wages by province and economic sector explains why these four specific years have been chosen for consideration in this study, and as such, they limit the study of market potential to these four cut off points. The sources used and the steps taken to estimate the wages are described below, taking agricultural and industrial wages separately.

Agricultural nominal wages are drawn from two sources: Sánchez-Alonso (1995) and Bringas (2000). To begin with, agricultural wages for the year 1860 are given by

[92] The calculations follow Nicolau (2005) and Foro Hispánico de Cultura (1957).

Sánchez-Alonso[93]. Agricultural wages for the remaining years selected were obtained from Bringas (2000). For 1900, the original source is the emigration statistics published by the Instituto Geográfico y Estadístico: IGE (1903)[94]. For 1914 and 1930, the figures offered by Bringas come from official sources and are published in the Spanish Statistical Yearbooks (*Anuarios Estadísticos de España*) from 1914 to 1931[95]. The missing wage data for certain provinces in 1930 is rectified by using the interpolations reported in Silvestre (2003)[96].

Agricultural wage series occasionally show exceptionally high figures. This is the case in the provinces of Logroño and Pontevedra in 1897, and in those of Castellón and Lleida in 1930. In all cases, their wages exceed the standard deviation of the sample average and present higher values than those recorded in the nearest provinces. These wages are corrected by a simple average of the wages in the neighbouring provinces.

For nominal wages in industry, three sources have been consulted: Madrazo (1984), Sánchez-Alonso (1995) and Silvestre (2003). Industrial wages for 1860 are given by Madrazo[97]. Figures for ten professional categories involved in the building of roads are offered. Six of them are quite well represented across provinces (apprentice, unskilled labourer, mason, bricklayer, carpenter and miner). The industrial wage for 1860 comes from a simple average of three categories established according to the level of skill[98]: bricklayers (skilled workers), unskilled workers, and apprentices.

However, a number of problems arise. The geographical coverage of bricklayers is high but no information is available for their wages in six of the provinces[99]. In order to fill this gap, data for the most similar professional category for which Madrazo offers information, that is, masons, is used. The wages of bricklayers in these six provinces are calculated from the wages of masons, and their deviation from the average wages for masons in Spain, weighted by the industrial population of each province according to the Population Census of 1860.

93 Sánchez-Alonso (1995), p. 302-303. '*Salarios agrícolas, años 1849-1856*', irrespective of sex. The primary sources are Moral Ruiz (1979) and García Sanz (1980).

94 '*Jornal medio de los obreros agrícolas en las poblaciones de menos de 6.000 habitantes en el año de 1897'*. IGE (1903), p. xlvii-xlix.

95 '*Jornales medios diarios masculinos'*. Bringas (2000), p. 180-183.

96 Silvestre (2003), p. 338. This author offers information for Ávila, Badajoz, Madrid, Santander, Segovia and Valencia in 1930.

97 '*Jornales de los obreros de la construcción de carreteras durante el año 1860 en reales de vellón'*. Madrazo (1984), p. 208.

98 Here the aim is to obtain the highest degree of homogeneity with the wages in Silvestre (2003). A simple average is used, given that no data on active population in each occupation are available, and so the average cannot be weighted.

99 Guipúzcoa, Lugo, Ourense, Oviedo, Vizcaya and Zamora.

On the other hand, as no wages are provided for any professional category in Navarre, they must be estimated. It would be reasonable to think that there might be a wage gradient depending on geographical proximity. Indeed, for the rest of the years available, it is confirmed that the industrial wage in Navarre is close to the average wage of the neighbouring provinces. Therefore, the industrial wage in Navarre in 1860 is calculated as being the average wage in the neighbouring provinces[100].

Industrial wages in 1900 come from Sánchez-Alonso (1995)[101]. Regarding the primary data of IGE (1903), Simpson (1995b) defines them as semi-skilled workers and he points out two provinces with excessively high wages: Pontevedra and Toledo[102]. The values have therefore been corrected by re-calculating in both cases their wages as the average of the industrial wage in the neighbouring provinces.

Finally, industrial wages in 1914 and 1925 are given by Silvestre (2003)[103]. This author provides data for nominal wages per hour weighted by the active population in each occupation according to different categories: skilled male workers, skilled female workers, unskilled labourers and apprentices (male and female). For 1930, the hourly wages in 1925 are used, because for subsequent years the average cannot be weighted among occupations since no data are available on the active population in each one. Another obstacle now has to be overcome: the absence of industrial wage data for the Canary Islands. For 1914, it is assumed that the industrial wage in the Canary Islands is similar to that of the lowest one among the Spanish provinces (0.28 pesetas per hour). For 1925, the increase in the industrial wage is also assumed to be similar to that of the Spanish economy as a whole.

Finally, there is no information on wages in the service sector in Spanish provinces. Following Geary and Stark, who faced the same problem in their study of the British economy, the service sector wages are calculated as a weighted average of agriculture and industry series in each province, where the weights are each sector's share of the labour force[104]. Wages in the Geary and Stark equation are relative wages with respect to the Spanish total, defined as the ratio between the nominal wage by sector in a

[100] As a test for this result, wages in 1897, the closest available date, are used. In that year, the industrial wage in Navarre was 3% higher than the Spanish average weighted by the industrial population. If this percentage is applied to the Spanish value for 1860, the figure obtained almost coincides with the one calculated previously.
[101] '*Jornales fabriles en las capitales de provincia (pesetas) en 1896-1897*'. Sánchez-Alonso (1995), p. 294-295. The original source is, again, IGE (1903), p. xlvii-xlix.
[102] Simpson (1995b), p. 190 and 199, respectively.
[103] Silvestre (2003), p. 341-342. Data come from Ministerio de Trabajo, Comercio e Industria (1927).
[104] Geary and Stark (2002), p. 923.

province and the average nominal wage by sector in Spain. The Spanish wage is obtained as an average of the wages in the 49 provinces[105].

To conclude this section, a final point needs to be stressed. The methodology used and the lack of data, particularly for wages, force us to impose three main assumptions: first, relative wages accurately reflect the relative average productivity across sectors and provinces for all employees; second, the series of wages, not homogeneous throughout time, are representative of agriculture and industry; third, service sector wages can be represented by a weighted average of agriculture and industry wages[106]. Having collected these data and applying the equation proposed by Geary and Stark (2002), an initial estimation of the GDP of Spain's provinces at factor cost and at current prices for the years chosen is obtained[107].

3.3.2.2. Inter-provincial transport costs. In the mid-19th century, the Spanish railway was still very much in its infancy. However, the first basic phase in the construction of the country's rail network had been completed by 1866, with all the main population nuclei and centres of economic activity joined up[108]. Thus, by this date, 32 provincial capitals formed part of the network[109], which is the reason why 1867 has been chosen as the first year in this study[110]. Yet, given that a significant number of provinces remained unconnected to the rail network, it is necessary to consider an alternative means of land transport to that of rail for this particular year: namely, road transport. In addition, the geography of the peninsula implies that cabotage (coastal shipping) between the Spanish ports chosen as connecting nodes for the coastal provinces must also be included.

---

105 In this case, a total of 49 provinces are included, as the two provinces making up the Canary Islands are counted as one. Ceuta and Melilla are excluded from the analysis. The wages used can be consulted in table A3 in the Appendix.

106 "*As regards the use of nominal wages, to the extent that there are regional variations in price levels then there will also be bias. A priori, it is not possible to assess the net effect of these biases, tests confirm that the method produces acceptable results*". Geary and Stark (2002), p. 933-934.

107 See table A4 in the Appendix.

108 "*The volume of potential traffic that was to benefit from the substitution of the road by the railway was much greater in the case of the lines built before 1866 than in the case of those built later, which shows the crucial importance of this first period in the railway age in the reduction of Spanish transport costs*" [Own translation]. Herranz (2005), p. 188-189.

109 Cordero and Menéndez (1978), p. 245-256; Wais (1987), p. 255-262. The provincial capitals connected by rail can be consulted in table A2 in the Appendix.

110 Given that the GDP estimation corresponds to 1860, it is assumed that the structure in the territorial distribution of GDP in 1860 would have been maintained seven years later - the date chosen for computing transport costs and market potential. This assumption carries with it a high degree of uncertainty, since the early years of the 1860s were a period of growth in the Spanish economy. According to the data in Prados de la Escosura (2003), between 1860 and 1867, Spain's GDP grew in real terms (pesetas of 1995) by 6.35%.

As for the country's inland waterways, their role within the transport system was only minor, and so this means of transport is not included in the calculation of market potential[111].

Therefore, the estimation of transport costs in 1867 requires data for inter-provincial distances as well as for the rates applied to the transport of goods by rail, road and coastal shipping. By contrast, by 1900 all the provincial nodes had been connected up to the rail network[112]. Hence, beyond this date it is assumed that all inland commodity transport used either this mode of transport[113] or both, rail and cabotage between the coastal provinces. Although in the interwar years the motorisation of road transport was initiated, no great advances were made in Spain until the thirties, and for this reason it too is excluded from the analysis[114]. Consequently, in 1914 and 1930 it is accepted that rail and coastal shipping, as in 1900, were the means used for transporting goods between Spain's provinces[115].

Given that the coastal provinces could use both rail and cabotage to transport goods, it becomes necessary to know the respective volumes transported by each of these modes. This information, which is presented in table 3.1, is taken from Frax (1981), who compared the volume of traded goods transported by cabotage and rail between 1867 and 1920.

Having analysed the modes of transport considered in each year of the study, the transport costs between Spain's provinces can be calculated using the distances between the provincial nodes, to which can then be applied the mean estimated rates for the transport of goods, differentiating in both cases between the respective modes of transport.

---

[111] "*Navigable waterways always played a very small role in Spain's transport system in comparison with those of other countries [...] Moreover, when speaking about the use of canals for inland navigation, the historical literature has concluded that its impact was much greater in areas such as the generation of energy and the irrigation of farm land than in transport*" [Own translation]. Herranz (2005), p. 186.

[112] In fact, Teruel was not connected to the network until the 28th of June 1901, with the conclusion of the stretch that linked Puerto Escandón and Calatayud (Wais, 1987), but given the relative proximity of the date, and in order to simplify the calculations, it is considered as having been connected in 1900.

[113] The roads continued to be used for the transport of commodities between neighbouring provinces in this year. However, road haulage rates were higher than those charged on the railway and thus road transport only became an advantage over short distances thanks to the possibility of door-to-door deliveries.

[114] "*The 1930s would witness the beginning of the substitution of the train by the lorry, a process that would be temporarily interrupted during the post-war years but which was renewed with greater force in the 1950s*" [Own translation]. Herranz (2005), p. 198-99.

[115] The structure of the railway line was such that in the Atlantic seaboard rail and cabotage complemented each other, with the ships making good the lack of trains. However, in the Mediterranean, the trains ran right along the coast so that the competition between the ferry and railway companies must have been greater. Gómez Mendoza (1982), p. 82-83.

Table 3.1. Distribution of the volume of traded goods by coastal shipping and rail (%)

		Coastal shipping	Rail
1867	1867-1870	20.73	79.27
1900	1896-1900	12.38	87.62
1914	1911-1916	14.64	85.36
1930	1916-1920[116]	15.69	84.31

Source: Frax (1981), p. 40.

Distances. In 1867 Spain's rail network, as already mentioned, connected only a limited number of provinces. This means that all distances need to be ascertained according to the mode of transport under consideration: first, the distances by rail between the 32 provinces connected to the network; second, the distances by road for the 15 provinces that had no rail link; and finally, the distances by sea between the 17 ports chosen as nodes for the coastal provinces from where coastal trade was plied.

The distances by rail in 1867 can be obtained from information gathered by Wais (1987)[117], who reports the distances and the dates when the wide gauge stretches of the track were laid. By aggregating the various stretches of track it is possible to reconstruct the total distances between the 32 provincial capitals with a connection to the rail network. To calculate the distances by rail in 1900 – the year by which all the provincial nodes had been connected to the network –, an alternative source is used: the statistics published by the Ministry of Public Works: *Estadísticas de Obras Públicas*[118]. The inclusion of Teruel and the linking of Murcia, Oviedo and Pontevedra with Cartagena,

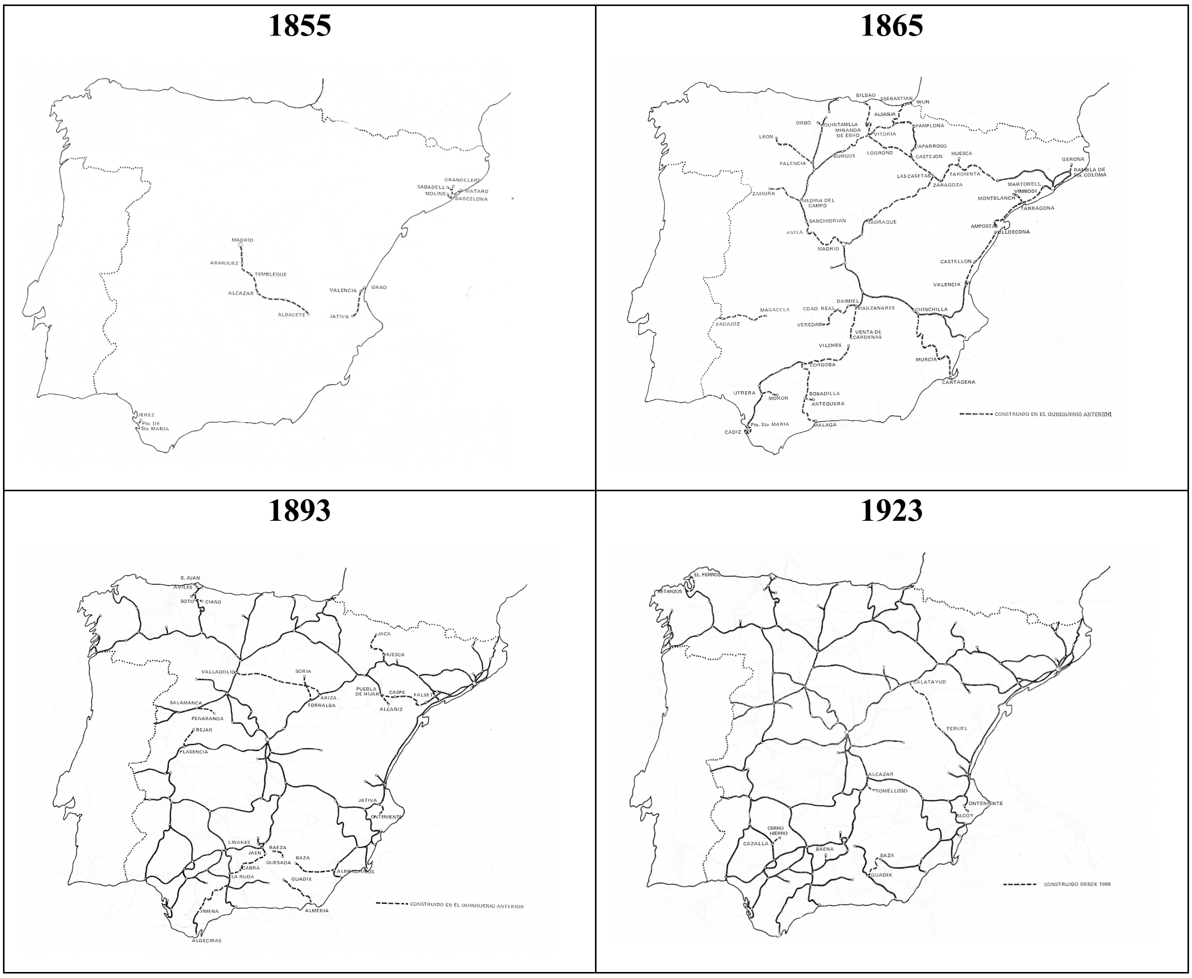
[116] Applying the percentage for 1916-1920 to the observation corresponding to 1930 does not give rise to any major distortions. The share attributable to railway and coastal shipping in this five-year period has been calculated using the cabotage data in Frax (1981), p. 70. Although the *Estadísticas de Cabotaje* stopped being published in 1920, the author provides figures for the total amount of merchandise transported by cabotage since that date drawing on data published in the *Estadística del Impuesto de Transportes por Mar y a la entrada y salida de las fronteras*. If to this information, the volume of merchandise transported by rail, which is taken from Anes Álvarez (1978), p. 492, is added, it is possible to calculate the mean percentage that the volume traded by cabotage represented with respect to rail, which in 1926-30 stood at 15.62%, very similar to the figure of 15.69% in 1916-1920.

[117] Wais (1987), p. 255-262.

[118] The statistics include '*Distances between the provincial capitals, following the shortest route, on the Spanish lines in use on the 31 December 1900*'. Ministerio de Obras Públicas (1902).

Gijón and Vigo, their respective coastal nodes, is undertaken, once more, using the data supplied by Wais (1987)[119].

Maps 3.1. Expansion of the railway network in Spain (1855-1923)



Source: Cordero and Menéndez (1978).

The expansion of the rail network between 1900 and 1914 meant that by this final date a number of changes had been made to some routes of the network[120]. The newly opened up stretches affected primarily the connection between Murcia and Granada via Guadix. The completion of this branch had a direct impact on three provincial nodes: Almería, Granada and Murcia (and indirectly on Jaén). Therefore, as the reduction in the

[119] In the case of the connections between Oviedo and Gijón and between Pontevedra and Vigo, the information corresponds, respectively, to the stretches that link Pola de Lena with Gijón, and Pontevedra with Redondela. Wais (1987), p. 255-262. The distances between Pola de Lena and Oviedo and between Redondela and Vigo are obtained using an electronic atlas and checked to the maps in the *Memorias de Obras Públicas*. Advances in the development of Geographic Information Systems (GIS) should enable these distances to be calculated more easily in the future.

[120] Cordero and Menéndez (1978) and Wais (1987).

distances between these three provinces and the other provincial nodes was quite considerable, the distances for 1900 have been corrected in accordance with the data supplied by Wais (1987)[121]. Finally, given that between 1914 and 1930 the rate of expansion of the rail network fell markedly, the same rail distances as those used for 1914 are employed here[122]. Based on these assumptions, a railway distance matrix can be built for each of the four years studied.

The distances by road in the 1860s - the only year for which they are necessary in this study - were obtained primarily from the Dirección General de Obras Públicas (1861) by comparing the routes and the distances with the road network map included in the *Memorias de Obras Públicas* published by the Ministerio de Fomento (1856). Where the distance by road between two provincial nodes was unavailable, an electronic atlas was consulted to verify that the route followed by the present-day road coincided with that in the *Memorias de Obras Públicas* for the mid-19th century[123].

In the case of cabotage, the distances by sea between the ports of the peninsula (corresponding to the nodes of the provincial capitals) were obtained from various web pages[124]. These distances, which obviously remained unchanged over time, are used for each of the years considered in this study.

The **transport freight rates** are calculated using the mean rates applied to the transport of commodities by each of the modes of transport: rail, road and cabotage, expressed in pesetas/tonne-kilometre (pts/tkm). The mean rates charged by the railway companies for goods transport have been calculated by Alfonso Herranz[125]. According to his information (table 3.2), the mean rate weighted by the traffic of goods by rail at current prices in 1867 was 0.111 pts/tkm. During the second half of the 19th century a

[121] In the case of Almería, the correction affects the distances with the rest of the capitals of Andalusia and those in the east of the peninsula. The direct connection between Granada and Jaén reduces the distances between the centre and the north of the peninsula. The connection with Murcia also reduces distances with the southeast and east of the peninsula. In the case of Murcia, the improvement comes from the connection with Andalusia.

[122] "*After 1914, [...] the Spanish railway could not expand anymore, due to the low traffic expectations on the routes that had yet to be linked to the network*" [Own translation]. Herranz (2005), p. 197. The only change that might have had some impact was the construction of the line between Burgos and Soria, the aim of which was to provide a faster link between Santander and the Mediterranean. This was completed in 1929.

[123] In this regard, "*it would seem that after 1868 the network of first and second order roads barely increased at all, in contrast with those of the third order, proof, if somewhat late, that the radial and tree-like pattern of the road network was being replaced by a reticular pattern*" [Own translation]. Frax and Madrazo (2001), p. 40.

[124] www.dataloy.com, www.distances.com.

[125] Unpublished data kindly provided by the author. This information, however, serves as the basis for the construction of Graph 3 in Herranz (2005), p. 192. In this study, the freight rates are expressed in constant prices of 1914. In this regard, see also Gómez Mendoza and San Román (2005), p. 543-544.

marked fall was recorded in the prices of rail transport, reaching a mean rate of 0.078 pts/tkm by 1900[126]. However, in the early decades of the 20th century, this fall was reversed[127]. Moreover, in the interwar years, the highly inflationary context in Spain's economy affected the prices of rail transport. Given that current prices are considered, this translates as an increase in the rates for the transport of goods by rail to 0.106 pts/tkm[128].

Table 3.2. Mean freight rates charged by the railway companies for the transport of commodities (current pesetas/tkm)

1867	0.111
1900	0.078
1914	0.079
1939	0.106

Source: Herranz (2005).

In the case of the mean freight rates charged by cabotage, the data come from Nadal (1975)[129]. This study reports prices in pts/t paid in the transport of Asturian coal from the port of Gijón to eleven other Spanish ports in 1865[130]. In order to obtain the price in pts/tkm, a potential fit was performed on the maritime distance data from Gijón to these ports[131]. This fit gives the following equation: $y=0.643x^{-0.5352}$. Substituting the

[126] This could be linked to two factors. Until the 1870s, the reduction in freight rates might have been due to the "*gradual development of the railway connections as the network was constructed [...] with the growing weight of long distance transport, for which cheaper rates were charged than those for short distances*" [Own translation]. And throughout the second half of the 19th century, it was the result of "*deliberate attempts on the part of concessionaries to reduce prices so as to capture a greater volume of traffic*" [Own translation] at the expense of other railway, road haulage and cabotage companies. Herranz (2005), p. 193-194. See also Pascual (1990).

[127] "*At the end of the 19th century, the companies were able to harness the autonomous dynamism of the demand for transport, related probably with economic growth, structural change and the relocation of activity within the Spanish economy, so as to maintain their freight rates stable and to gradually improve their financial position, until on the eve of World War I they were in a relatively healthy situation*" [Own translation]. Herranz (2005), p. 194. Even though the analysis was conducted with rates at constant prices of 1914, it is equally valid for the current prices considered here.

[128] When considering constant prices of 1914, Herranz concludes "*the prices of rail transport barely register any real fall in the interwar years, once the effects of World War I inflation had been overcome*" [Own translation]. Herranz (2005), p. 197.

[129] Nadal (1975), p. 137-138. The primary data are contained in the publication '*Información sobre el derecho diferencial de bandera y sobre los de aduanas exigibles á los hierros, el carbón de piedra y los algodones*'. Ministerio de Hacienda (1867), p. 23-27.

[130] San Sebastián, Bilbao, Santander, Coruña, Cádiz, Sevilla, Málaga, Adra, Cartagena, Valencia and Barcelona.

[131] The potential fit shows a higher R^2 than other options, including linear, logarithmic and exponential fits, which explains its adoption.

distance between each pair of ports (x) the transport costs for cabotage between the coastal nodes is obtained.

However, the literature provides alternatives for the estimation of cabotage rates. On the one hand, Gómez Mendoza (1982) bases his calculations on the information for Asturian coal, performing in this case an exponential fit[132]. On the other hand, Barquín (1999) adopts a different strategy, in which the estimated price for shipping is given by the expression: $y=11.26+0.008x$. Barquín (1999) derives the fixed component in this equation from various sources published in the *Boletín Oficial de Comercio de Santander*, for the years 1848-50, 1854, and 1866, while the cost per kilometre covered is derived from a linear fit of the data contained in Nadal (1975)[133].

The mean freight rates for coastal shipping calculated for 1865 are applied in this study to the year 1867. For the other years included, a number of modifications are needed. As Gómez Mendoza points out for the Spanish case, "*in 1867 there was clear predominance of sailing ships over steam ships. However, the use of iron hulls for shipbuilding and the replacement of sails by steam meant the freights could be reduced. In 1860, 96 per cent of the tonnage transported by the merchant navy was done so by sailing ships. A quarter of a century later, this percentage had fallen to 27 per cent*" [134]. Therefore, the advances made in maritime shipping have to be incorporated in the calculation of the cabotage prices, correcting the average for the 1867 freight rates for the years 1900, 1914 and 1930.

In recent decades there has been considerable international debate regarding the reduction in maritime transport costs in the years leading up to World War I, a period in which the world economy experienced a strong globalizing force (O'Rourke and Williamson, 1999). Among the various indices of ocean freight rates present in the literature, here the most recent is used: the index devised by Mohammed and Williamson (2004)[135]. This is a nominal index that includes information for a large number of routes

[132] Gómez Mendoza (1982), p. 86. The equation obtained is: $y=0.04(0.9993)^x$.
[133] Barquín (1999), p. 341.
[134] "*In addition, improvements to port facilities, including those made in the ria of Bilbao, Barcelona and Gijón, allowed boats of greater tonnage to dock in these ports without having to anchor outside. All these factors helped reduce mean fixed shipping costs and, as a result, freight rates*" [Own translation]. Gómez Mendoza (1982), p. 86.
[135] Isserlis (1938), North (1958), Harley (1988), and Mohammed and Williamson (2004). This index shows a fall in the size of the transoceanic freight rates of more than 50% between 1869 and 1900, which would reflect the increasing productivity of the sector. However, it should be taken into account that the various alternatives show different dynamics in the evolution of the maritime freight rates, which would inevitably affect our results. See table 3.3.

between Europe and the rest of the world based on information supplied by Angier (1920)[136].

Table 3.3. Maritime freight rates

	Isserlis		North		Mohammed-Williamson
1869	100	1865	100	1870-74	100
1900	76	1900	95	1895-99	48
1914	67	1910	48	1910-14	48
1930	93			1925-29	59

Source: Isserlis (1938), p. 122; North (1958), p. 549; and Mohammed and Williamson (2004), p. 188.

Finally, the lack of information on road freight rates in the middle of the 19th century is a feature that is frequently highlighted by Spanish economic historians[137]. In this instance, as discussed earlier, only information concerning the rates in force around 1867 is required[138]. First, Barquín (1999) has undertaken an estimation of road transport costs based on various sources for the period 1848-1884, differentiating between the prices paid for three different types of product: liquids 0.63 pts/tkm; coal 0.46 pts/tkm; and other products 0.30 pts/tkm[139]. These different prices have to be weighted to obtain a single mean price for 1867. Following Barquín, the weighting criterion is based on obtaining a mean freight rate as a function of the volume transported by railway for each of these three groups of products in 1869, the nearest year for which data are available[140]. Subsequently, this same distribution is applied to road transport. The resulting mean freight rate for transport by road in 1867 is 0.36 pts/tkm.

---

[136] Mohammed and Williamson (2004), p. 175-177 and 188. It should be borne in mind, therefore, that, an ocean transport cost index is used to approximate the fall in cabotage freight costs. In this regard, "*it is well-known that technological innovation in the maritime shipping industry reduced long-haul freight rates more than short-haul ones*". Jacks and Pendakur (2008), p. 4. These authors also question the impact of the fall in transport costs on trade prior to World War I: "*there is little room for maritime transport revolutions to be the primary drivers of the two global trade booms of the nineteenth and twentieth centuries*". Jacks and Pendakur (2008), p. 21.

[137] Frax and Madrazo (2001), Gómez Mendoza (1982), and Herranz (2005).

[138] "*As might be expected given the absence of significant technological changes in the sector [...] there does not appear to have been any great reduction in road haulage rates between the middle and end of the 19th century*" [Own translation]. Herranz (2005), p. 196.

[139] Barquín (1999), p. 339-341. Transport by cart, excluding pack animals.

[140] Anes Álvarez (1978), p. 496-501. Information for the companies MZA and Norte. Of the total volume transported by these companies in 1869, 10.94% corresponded to liquids, 15.51% to coal and 73.55% to other products.

However, it should be borne in mind that elsewhere other options have been adopted. Gómez Mendoza (1982), for example, bases his calculations on the social savings of Spain's railways by using an official survey conducted in the wine trade in 1884[141]. This information contains the transport prices in 27 provinces based on 178 entries for the transport of commodities by cart and 59 for pack animals[142]. Using these data, the author presents the mean transport price by road in pts/tkm, both for cart haulage and for pack animals for different distance intervals (every 15 kilometres, from 0 to 90 km)[143]. For distances over 60 km road transport has a constant mean cost around 0.64 pts/tkm, which is higher than that previously obtained.

The primary shortcoming in using these data is the consideration that wine prices might have been representative of road haulage costs in general. According to Barquín (1999), these prices were in fact higher than average transport costs. He reports that the main product being traded was grain and that the transport costs for this product were lower[144]. To this assessment, Herranz (2002) adds that the data record charges for transport by cart over short distances (below 89 km), and that the information in the questionnaire was based on the use of secondary roads that fed the railway stations, and which would have been of poorer quality than the main roads that ran parallel to the railway lines. He therefore supplements the data provided by Gómez Mendoza for the wine trade with the prices for the transport of grain provided by Madrazo (1984)[145]. The latter uses information for the transport of grain on routes that run from the coast to the interior of the peninsula estimating a mean price for the long-distance transport of grain by road in the middle of the 19th century of 0.41 pts/tkm[146]. Weighting these data, Herranz (2002) obtains prices for road transport of 0.54 pts/tkm, which he applies in his revision of the social savings of the Spanish railways.

From these three alternative estimations of the mean rate for road transport of commodities in the middle of the 19th century, the first one is selected. The selection is based on the greater amount of information contained in the Barquín's (1999) estimation and on the different purpose of the studies conducted by Gómez Mendoza (1982) and

141 The survey among wine producers, and which enquired about both road conditions and the transport costs for wine, is found in '*Información Vinícola'* (1886).
142 The raw data are published in Gómez Mendoza (1999), p. 733.
143 Gómez Mendoza (1981), p. 111.
144 Barquín (1999), p. 339.
145 Madrazo (1984), p. 749.
146 This figure is similar to estimates reported in Garrabou and Sanz (1985), p. 48 and Barquín (1997), p. 35.

Herranz (2002) whose primary interest lied in calculating the social savings generated from the construction of Spanish railways. Annex 1 in the Appendix shows details of the transport cost matrices constructed for each year for the different modes of transport.

3.3.2.3. The self-potential of each province. Having described the elements that enable the calculation of the inter-provincial market potential, now it is necessary to consider each province's self-potential in order to obtain its overall internal market potential. The computation of this figure is based on the ratio between the GDP of a province *r* and the estimated intra-provincial transport costs in that province. In this case, determining the internal distance d_{rr}, which is used to obtain the transport costs, is particularly important. This question is a highly controversial one among geographers and economists given that the final results of market potential are highly sensitive to the measure adopted[147]. Furthermore, the relative contribution between different provinces must inevitably be affected by the volume of economic activity in each zone, so that in the provinces with the highest GDP, the contribution to the total potential from the province's own market will be greater.

Most studies have adopted an expression in which the internal distance of each area under consideration takes the form of a circle in which all the economic activity is located at its centre. In line with Keeble, Owens and Thompson (1982)[148], the component d_{rr} is calculated using the expression:

$$d_{rr} = 0.333\sqrt{\frac{(areaoftheprovince_r)}{\pi}} \quad (3.8)$$

Thus, to obtain the self-potential of each province a value for the internal distance of each province that is equivalent to a third of the radius of a circle with an area equal to that of the province is taken. The choice of this measure is linked to the methodological similarity of this study to that of Keeble, Owens and Thompson (1982) and the more recent studies undertaken in economic history by Crafts (2005b) and Schulze (2007). However, the studies discussed in the review of empirical papers carried out within a

[147] Frost and Spence (1995).

[148] This study bases its proposal in turn on that made by Rich (1980), who, by contrast, proposed a constant of one half. However, their sensitivity analysis leads them to assign a constant of a third because if not they observe that smaller, highly urbanised regions suffer a penalisation. See Keeble, Owens and Thompson (1982), p. 425.

NEG framework, in some cases, adopted different strategies. In an international context, Redding and Venables (2004) used three alternative measures for calculating the internal distance, including a similar expression to that described before, albeit that in their particular case the constant used is equal to two thirds[149]. In other studies it is assumed that if the size of the regions considered is similar, then a fixed value for d_{rr} can be adopted[150].

To obtain the market self-potential of each province in the Spanish case requires the following information. On the one hand, to calculate d_{rr}, the area of each province is needed – data is obtained from the *Instituto Nacional de Estadística*[151]. On the other hand, given that transport costs are needed, the intra-provincial distance, d_{rr}, has to be multiplied by the various freight rates in force in each of the years considered. Thus, in 1867, given that not all the provinces had been connected to the rail network, the intra-provincial transport costs are obtained by multiplying the component d_{rr} by the rates applied in that year to the railway in the thirty-two provinces that were connected to the network. In the fifteen remaining provinces, the absence of a railway connection means that all transport within the province is assumed to be by road, and therefore the road haulage rates are applied. In the other years chosen, once the railway had linked up all the provincial nodes, the intra-provincial transport costs are obtained by applying the corresponding mean rates for rail transport. Finally, the market size of each province is measured using the GDP figures obtained following the Geary-Stark method (see section 3.3.2.1 in this chapter, and table A4 in the Appendix)

3.3.3. Foreign market potential

Here, an alternative strategy is used to that adopted above in calculating internal market potential. The strategy selected also differs in a number of ways from the method used for calculating foreign market potential in both Crafts (2005b) and Schulze (2007). In these two papers, external transport costs were obtained by using the ocean freight costs provided by Kaukiainen (2003) for grain and coal trade since 1870. These figures include a constant fixed cost for the work at the terminal, which includes the cost of loading and unloading the cargo as well as other port activities, and a variable cost linked

[149] Combes, Mayer and Thisse (2008), p. 313, propose a similar value. By contrast, Davis and Weinstein (2003) exclude this constant and simply take the square root of the area of a country divided by π.

[150] This is the case of Crozet (2004), who opts for an internal distance of 75 km. Kiso (2005), for example, takes a value across the Japanese prefectures of 15 and 25 km, respectively.

[151] www.ine.es.

to every 100 miles travelled. Tariffs are then added to these costs, since their existence represents an additional barrier to trade between countries and which, as such, must be considered an additional cost in the equation. Tariffs are included via the elasticities obtained for the distance and the tariffs in the gravity equations in Estevadeordal, Frantz and Taylor (2003). These elasticities allow them to convert the tariffs to equivalent transport costs that are then added to the fixed component of Kaukiainen's equation[152].

In this case, an alternative methodology is adopted, based primarily on the results of the gravity equation developed by Estevadeordal, Frantz and Taylor (2003). The option chosen by Crafts (2005b) and Schulze (2007), in which the GDP of foreign countries is divided by transport costs, implicitly assumes an elasticity of *-1* for the component d_{rs}. As discussed earlier, the gravity models are associated with distance coefficients that do not differ significantly from this value. In fact, in calculating the internal market potential of Spain's provinces, when dividing the provincial GDP by the transport costs an elasticity equal to *-1* is also assumed. However, the gravity equations for international trade allow us to estimate these elasticities more precisely. In order to exploit the quality of this information, here the elasticities obtained in Estevadeordal, Frantz and Taylor (2003) are used for calculating foreign market potential.

The gravity models seek to account for the volume of bilateral trade by using an equation that relates this variable with, among other factors, market size, distance and the tariff protection provided by the countries[153]. Thus, the intensity of trade flows between two countries is positively related with the respective size of the economies yet negatively with the distance and tariffs that separate them. Hence, shorter distances and lower tariffs will result in a greater attraction between two economies, thereby favouring their trade relations. The estimation of this equation in Estevadeordal, Frantz and Taylor (2003) generates coefficients for both variables, which, taken as an average for different specifications, show an estimated elasticity of *-0.8* for distance and *-1.0* for tariffs.

[152] See Crafts (2005b), p. 1162 and Schulze (2007), p. 10. Note that Estevadeordal, Frantz and Taylor (2003) equation makes use of panel data for the years 1913, 1928 and 1938, from a sample of 40 countries, which in some instances are reduced to 28 countries.

[153] In Estevadeordal, Frantz and Taylor (2003), a gravity equation is proposed to explain the volume of bilateral trade between two countries (*TRADE*) based on a series of variables: distance (*DIST*), GDP, GDP per capita, tariffs (*TARIFF*), the volatility in the nominal exchange rate between the two countries (*ERvol*), and a series of dummies that control whether the countries have access to the sea or not (*landlocked*), whether they share a border (*adjacent*), whether one or other of the countries is an island (*island*), whether the countries shared a colonial relationship during the period (*colonial*), whether both countries were on the gold standard (*gold*), and a final dummy (*diversion*) when one of the two countries but not both were on the gold standard.

On the basis of these results, the foreign market potential of Spain's provinces can be calculated, where foreign markets are considered potential destinations for Spanish exports. Therefore, taking a reduced version of the gravity equation, the volume of trade between a Spanish province and a foreign node would depend on the size of the foreign market (GDP_s), which is then modified according to the distance between both nodes (*distance*) and the mean tariffs operating at the foreign node (*tariffs*). Thus, the foreign market potential would depend on these variables in accordance with the following equation:

$$\varphi_{rs} = GDP_s \left(dista\eta ce\right)_{rs}^{\delta} (tariffs)_s^{\gamma} \tag{3.9}$$

or, expressed in logarithms, as in the original equation:

$$\varphi_{rs} = GDP_s + \delta \ln(dista\eta ce)_{rs} + \gamma(tariffs)_s \tag{3.10}$$

where, the distance and tariff coefficients would take the values δ=-0.8, and γ=-1.0, respectively. Thus, it can be supposed an extreme case in which if the distance to the foreign country were zero and there were no tariffs, the foreign market potential would be represented by the GDP in the foreign country (GDP_s). Therefore, any increase in distance and in the tariffs would bring about a reduction in the foreign GDP in line with the estimated elasticities. This way of proceeding allows considering the market size represented by each trading partner having subtracted, simultaneously, the effect of distance and tariffs on the volume of activity measured via the GDP.

3.3.3.1. Size of foreign markets. The size of the foreign markets is calculated on the basis of the respective GDPs of four countries (Great Britain, France, Germany and the United States) for 1871, 1901, 1911 and 1931. These figures, which are drawn from Crafts (2005b), are based, in turn, on the estimations made by Prados de la Escosura (2000). Given that the GDP figures reported by Crafts are expressed in millions of pounds sterling at current prices, the first step to be taken is the conversion of these figures to pesetas at current prices. To do this, the nominal exchange rate between the peseta and pound sterling in Martín Aceña and Pons (2005) is used[154]. The adoption of the nominal exchange rate to make this conversion is usual in studies of this kind, given that it is these rates that mattered to the agents at each of these points in time. However, the size of the foreign markets, in this case measured in terms of their GDP, is highly sensitive to the exchange rate chosen. As can be seen in table 3.4, the value of the peseta with respect to the pound suffered considerable fluctuations in the years selected for this study.

Table 3.4. Peseta/pound sterling exchange rate

1871	23.97
1901	34.78
1911	27.24
1931	47.64

Source: Martín Aceña and Pons (2005).

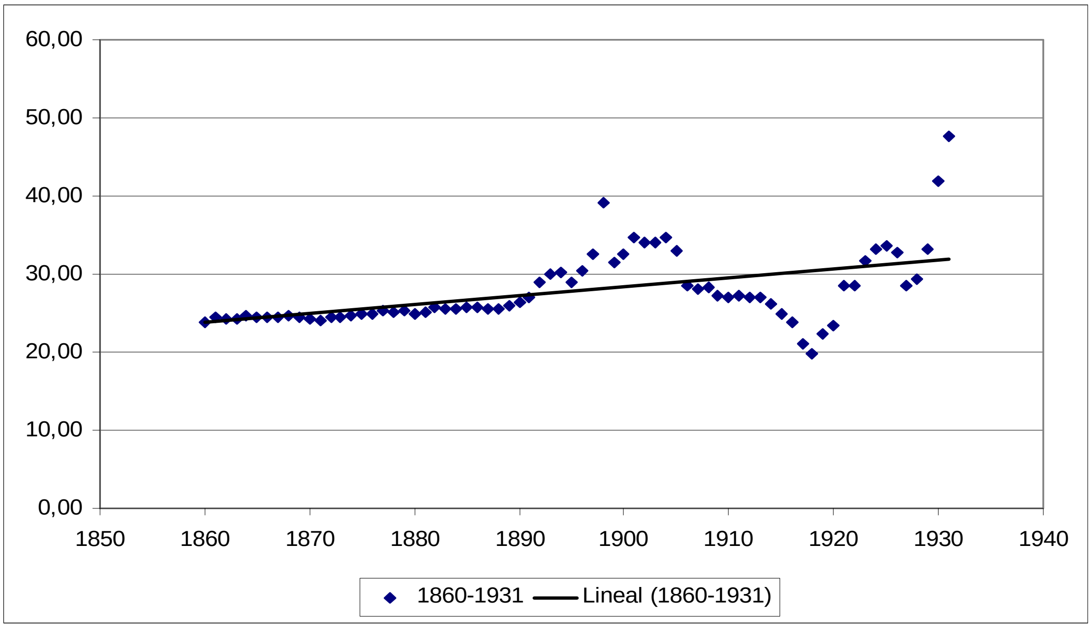
Table 3.4 shows that between 1871 and 1901 the value of the peseta depreciated against that of the pound. This fall in value actually began in 1892, with the peseta reaching its lowest point in 1898, reflecting the inflationary effects of financing the Cuban war. After this date, a period of gradual recovery in value was initiated as a result of the financial reform measures implemented by the respective governments. However, the 1920s ushered in a new period of depreciation of the Spanish currency. That situation was accentuated during the last few years of that decade, initially, by the deflation abroad as a result of the reintroduction of the gold standard and, later, in the early stages of the Great Depression, by the fact that the pound remained on the gold standard[155]. As this deflationary effect did not occur in the Spanish economy, the relative depreciation of the peseta was great, and in 1931 the value of the currency recorded an historical low.

Thus, these major variations in the exchange rate between the peseta and the pound sterling have marked effects on the calculation of the relative size of the foreign markets considered here. In order to ensure that these observations, which in some years might even be considered anomalous, do not have an extreme impact on the determination of foreign GDP expressed in pesetas at current prices, the study opts to capture the trajectory of the value of the currency by examining its trend throughout the period of study. The Purchasing Power Parity (PPP) theory holds that in the long term it

[154] Martin Aceña and Pons (2005), p. 703-706.
[155] The pound sterling abandoned the gold standard of the interwar years on 21 September 1931.

is the goods markets, through relative national and foreign prices, that determine currency values, recognizing, however, certain short-term deviations[156]. The linear estimate of nominal exchange rates between the peseta and pound sterling between 1860 and 1931 provides the coefficients to calculate the exchange rate value for the years being studied here, according to the overall trend in the series, as illustrated in figure 3.1 and table 3.5[157].

Figure 3.1. Peseta/pound sterling exchange rate, 1860-1931



Source: Author's own based on Martín Aceña and Pons (2005), p.703-706.

Table 3.5. Estimated peseta/pound sterling exchange rate

1871	25.00
1901	28.45
1911	29.60
1931	31.90

Source: see text.

Having obtained the GDP in pesetas at current value for the four trading partners considered in this study and for the chosen years, it is time to address a further issue –one

[156] The cointegration analyses show that the peseta fulfils the PPP theory in the long term. Serrano, Gadea and Sabaté (1998), Aixalá (1999), and Sabaté, Gadea and Serrano (2001).
[157] A linear fit is used in accordance with the following equation: *y=0.115x-190.17*.

already discussed earlier, and which affects the French market. The geography of the Iberian Peninsula means that the French market can be accessed both from the Atlantic and Mediterranean coasts, giving certain Spanish provinces a location advantage on the basis of their position with respect to this market. In order to capture the different possibilities of accessing the French market, here it is divided in three main markets, each of which is assigned a node of economic activity: Le Havre, Nantes and Marseille, respectively. The division of the French GDP is based on the regional population data contained in the Population Censuses for 1872, 1901, 1911 and 1931[158]. By proceeding in this way it is possible to obtain the foreign GDP figures used in the study:

Table 3.6. Foreign GDP (millions of pesetas at current prices)

	1871	1901	1911	1931
United Kingdom	30,194	58,284	68,956	139,030
France (N)	8,807	15,633	22,061	35,268
France (E)	7,137	11,874	16,465	25,370
France (W)	7,337	11,485	15,399	21,291
Germany	17,392	50,473	73,301	130,878
United States	36,775	115,694	196,250	623,975

Source: see text.

3.3.3.2. International distances. Most international trade studies that use gravity equations measure the distance variable between countries in terms of the geodesic distance, also known as the 'great circle distance'. This procedure involves calculating the distance in nautical miles as the crow flies while taking into consideration the curvature of the earth's sphere, which means including the longer distance that this curvature supposes. Yet, choosing this option here would give rise to a number of distortions in the results. For example, the geodesic distance between the ports of Bilbao and Le Havre would be similar to that between Bilbao and Marseille. However, to complete the latter route it is necessary to skirt around the peninsula on a much longer journey. In order to exploit our precise knowledge of the commercial routes, in this study the maritime

[158] www.insee.fr. Excluded from the total for France are Corsica and its overseas territories. The three large areas are built by aggregating the population in the NUTS1 regions as follows: North France (Île-de-France, Bassin Parisien, Nord-Pas-de-Calais); East France (Est, Centre Est, Méditerranée); and West France (Ouest, Sud Ouest).

distances between ports are considered[159]. In this case, unlike in studies of international trade, the mode of transport used to calculate the external market potential is exclusively shipping; hence, the external nodes are all located on the coast[160]. In addition, the sample of countries included is smaller than that considered by international trade studies, so that, with the exception of the United States, the distances to consider are not so great and, therefore, they are not so noticeably affected by the earth's curvature.

3.3.3.3. Tariffs. Finally, information about the tariffs operating between the trading partners considered in this study needs to be obtained. The mean tariffs are calculated as the percentage income from the tariffs with respect to total import volume. This indicator has been widely used in the international trade literature to measure a country's level of tariff protection[161]. The evolution in this measure over time and between countries, the problems it might give rise and alternative indicators have been studied by O'Rourke (2000)[162]. Moreover, the impact inflation can have on this measure and the determinants of global trade policy in the historical context of this study are analysed in Williamson (2006).

In this case, the calculation of mean tariff rates for the four countries included in the study sample is based on information drawn from O'Rourke (2000). The estimations for 1900 and 1914 use values relative to the five-year means between 1895-1899 and 1910-14, respectively. To obtain the mean tariffs for 1867 and 1930, the primary source used by O'Rourke (2000) is employed. The income data for tariffs and total volume of imports come from Mitchell (1998a, 1998b).

Given that for 1900 and 1914 the mean value for the preceding five-year period is taken, in this case, for 1867, the figures corresponding to the period 1860-64 are calculated. However, for these years no data for Germany is available[163]. Therefore, the German figures for this period are estimated on the basis of the changes in the mean tariff rates in another continental European economy: France. In fact, for the period from 1875

[159] www.dataloy.com, www.distances.com.

[160] Exports would also have arrived on the French market by rail, but their share as a percentage of all the trade between Spain and France seems to have been limited.

[161] "*Only one consistent measure of tariffs is available for the period from 1870 to 2000 in the form of the customs duties to declared imports ratio as in Clemens and Williamson (2001). This measure seems to be a reasonably good proxy for tariffs in the pre-World War I and interwar periods. However, after 1950 and the well-known rise of non-tariff barriers to trade, this measure becomes unreliable, sometimes registering unbelievably low levels of protection*". Jacks, Meissner and Novy (2009), p. 6.

[162] O'Rourke (2000), p. 461-464. Other leading studies include Bairoch (1989), League of Nations (1927), Liepmann (1938), and Estevadeordal (1997).

[163] The first data available for mean tariffs in Germany are for the period 1875-79.

to 1914, the German tariff is, on average, 90% that of the French figure. To calculate the mean German tariff in 1860-64, this percentage is applied to the French tariff in the same period. Finally, the mean tariffs used in calculating market potential are included in table 3.7[164].

Table 3.7. Mean tariffs (% respect to imports)

	1860-64	1895-99	1910-14	1926-30
United Kingdom	10.05	4.8	4.8	9.9
France	6.3	10.4	8.9	8.9
Germany	5.67	9.3	7.0	9.0
United States	22.78	22.7	18.3	14.6

Source: O'Rourke (2000) and Mitchell (1998a, 1998b)[165].

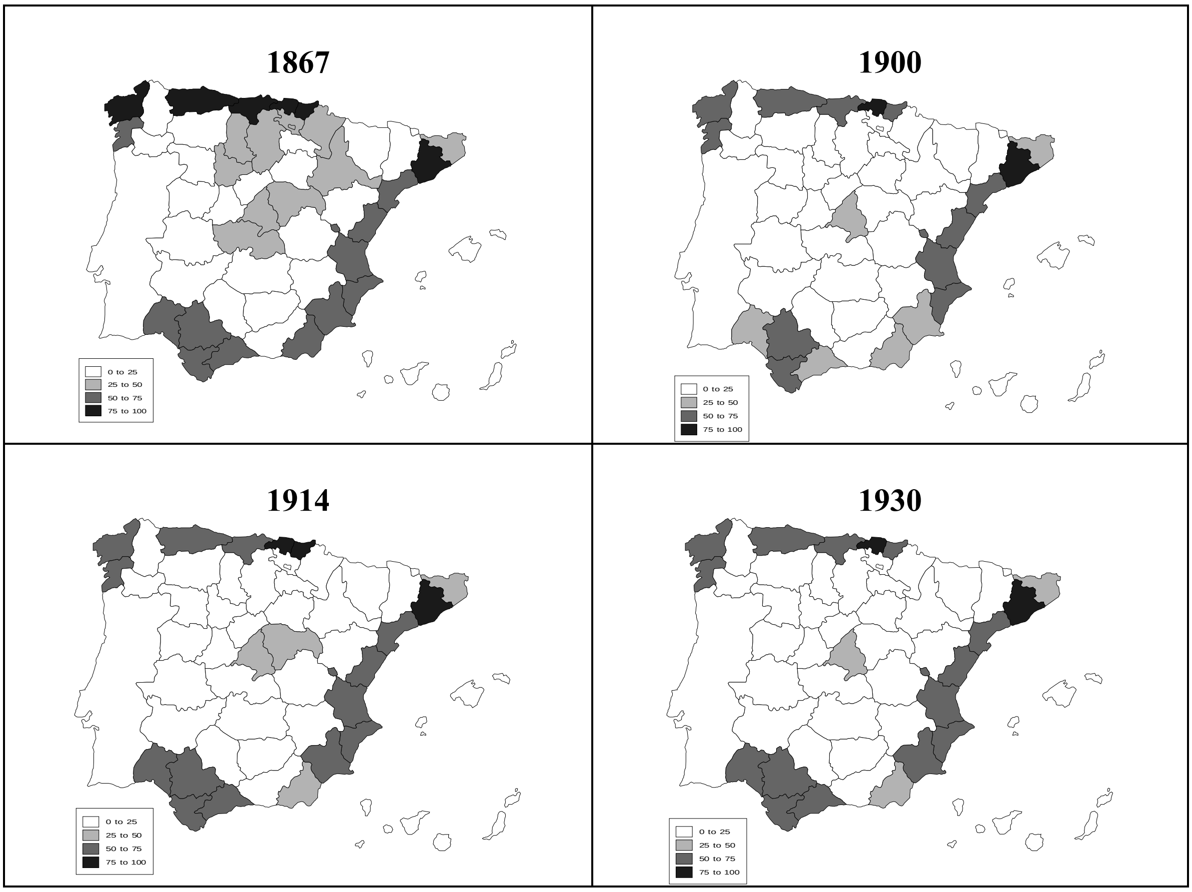
Using this information the foreign market potential of the Spanish provinces can be calculated. However, the procedure described has been used to obtain the foreign market potential in the seventeen coastal provinces, where the distances are calculated from the port of origin to the port of destination. For the remaining thirty inland provinces it is necessary to add the costs of transporting commodities from the provincial node to the nearest port. To do this, first, it is necessary to calculate the lowest transport costs from each inland provincial node to the nearest Spanish port. This is done by reducing the GDP of the country of destination by the transport costs, as was done previously for the internal market potential, and not via the elasticities that are applied to foreign trade. Second, the part of the foreign GDP that remains in the Spanish port of origin is deducted on the basis of the reduced expression derived from the gravity equation and the elasticities of Estevadeordal, Frantz and Taylor (2003), following the methodology used with the coastal provinces, but starting this time from a lower foreign GDP as the internal transport costs to the corresponding Spanish port have been deducted. The details of the calculation of the external market potential and the procedure used for the inland provinces can be consulted in Annex 2 in the Appendix.

[164] Note that in the interwar years, protection was centred on the proliferation of non-tariff barriers such as exchange controls, contingents and multilateral agreements.
[165] Tariffs are included in the gravity equations as *ln(1+t)*. Thus, in calculating the foreign market potential they have to be expressed on a per unit basis.

3.4. Results and initial hypotheses

Having gathered the information described in the section above, the market potential in Spain's provinces can be calculated for the years 1867, 1900, 1914 and 1930. The results obtained can be consulted in table A5 in the Appendix. An initial analysis of these results can be made by examining the geographical pattern presented by this variable. To facilitate this, maps of provincial market potential are drawn for the four years considered here. Throughout the period of study, Barcelona stands out as being the province with the greatest market potential, and therefore, maps 3.2 are expressed in relative terms with respect to this province. The maps show the evolution over time (1867-1930) in the market access of the Spanish provinces.

Maps 3.2. Market potential in Spain's provinces, 1867-1930 (Barcelona=100)



Source: see text.

An inspection of these maps allows drawing a number of conclusions. First, a notable change was experienced in the spatial distribution of market potential between

1867 and 1900 - a period characterised by a marked centrifugal tendency[166]. Thus, by 1900, the areas of greatest market potential were located in the coastal provinces of the geographical periphery of the peninsula, with the sole exception of Madrid. In that year, the main characteristic of the geographical pattern presented by market potential was the division between a group of coastal provinces, characterised by their better market access, and a second group of inland provinces of lower market potential. However, in the period 1900 to 1930 the spatial pattern of market potential showed hardly any variations and was characterised by a considerable degree of persistence. Thus, it can be concluded that, overall, by 1900 a clear polarisation was already apparent in the provincial distribution of market potential, with the most important changes in the evolution of this variable being concentrated in the first period (i.e., the second half of the 19th century).

In order to further examine this initial impression provided by maps 3.2, a test was performed to determine whether there had been convergence in the market potential in Spain's provinces for each of the two periods identified. This test was based on the convergence analysis proposed in the literature of economic growth by Barro and Sala-i-Martin (1991, 1992, 1995). In their studies, these authors tested for the existence of β-convergence, that is, an inverse relationship between the growth rate and the initial level of GDP per capita between countries and regions. An analysis of this kind, applied to market potential (unconditional β-convergence), allows evaluating not just the trends revealed in the maps but also to explore the provincial dynamics of this variable in greater detail, in particular for each of the two aforementioned periods.

Figures 3.2 and 3.3 show the results obtained from the convergence analysis. Figure 3.2 relates the market potential at the start of the study period with its growth rate in the period between 1867 and 1900. The relationship extracted from the two variables points to a clear situation of convergence[167]. In the graph it can be appreciated that, indeed, the provinces that began with the lowest market potential in 1867 were those that experienced the most substantial improvement in their market access between this date and 1900.

[166] Inland provinces such as Ávila, Burgos, Guadalajara, Navarre, Palencia, Toledo, Valladolid and Zaragoza that in 1867 presented a market potential that was between 25-50% that of Barcelona's saw their potential fall in 1900 into the lower interval (less than 25%).

[167] Together with the existence of β-convergence, the notion of σ-convergence implies a reduction in the dispersion over time of the observations that make up the sample. In the years studied the coefficient of variation in the market potential of the Spanish provinces, which in 1867 stood at 0.68, had fallen to 0.59 by 1900.

Yet, in terms of the market potential shown in 1867, three fairly distinct groups of provinces are seen to emerge. A first group of low market potential corresponds to those inland provinces that remained unconnected to the rail network in that year[168]. A second group, characterised by a medium level of market potential, is formed by the inland provinces connected to the rail network. The exception here is that of Madrid, which presented a market potential similar to that of some of the coastal provinces. This behaviour is maintained throughout the period indicating that, in the case of this province, the size of its own market and its good location within the peninsula compensated in part for the greater distance of Madrid from the international markets[169].

A third group, comprising the provinces with the greatest market potential includes all the coastal provinces which, as a whole, show a higher potential than that of the inland provinces. The greater proximity of these provinces to the foreign markets, as well as the fact that they had recourse to coastal shipping for their domestic transport, would appear to be the most reasonable explanations for this difference. Two further aspects should be stressed. On the one hand, as can be seen in figure 3.2, these provinces are more closely grouped together than their inland counterparts. On the other, within this group of coastal provinces those situated on the Atlantic seaboard have, in general, a greater market potential than their Mediterranean counterparts because of their closer proximity to the main foreign markets included in this study. In this case, a noticeable exception is Barcelona, which, despite its Mediterranean location, has a market potential similar to that of the provinces situated on the Atlantic seaboard. As with Madrid, this could be explained by the greater size of its own market.

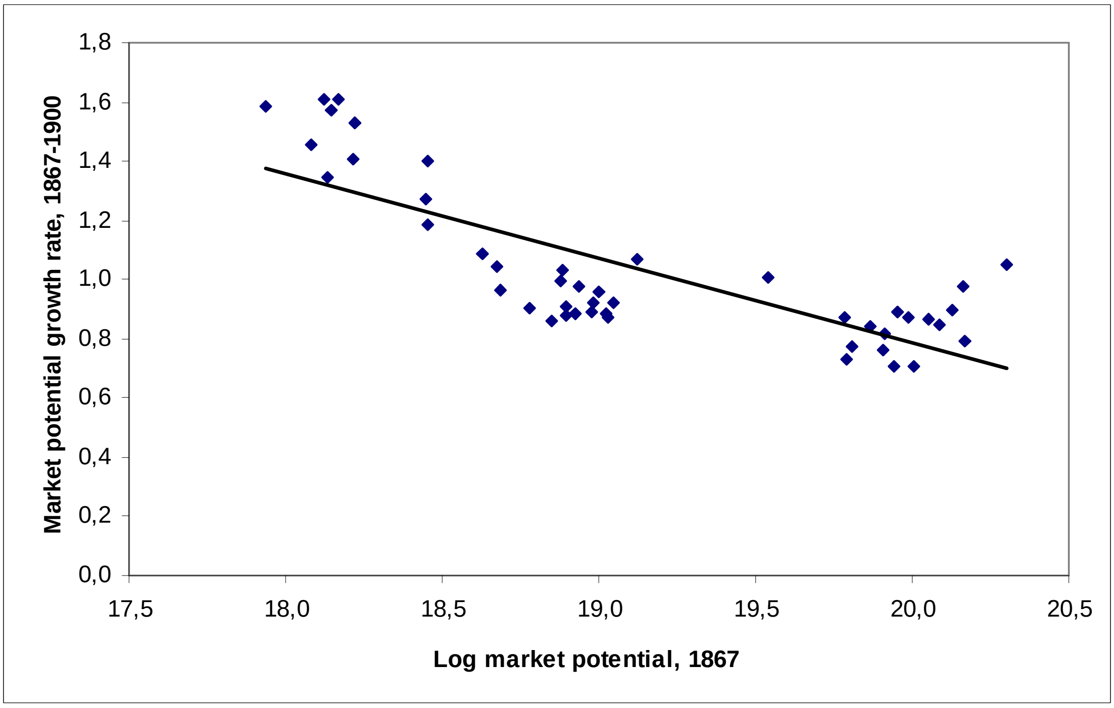
If the situation in 1867 is compared with the market potential of the Spanish provinces in 1900 (figure 3.3), the three groups of provinces identified in the first period of the study have become just two clearly differentiated groups. This coincides with a division between inland and coastal provinces, with the latter maintaining a higher market potential than their inland counterparts. Furthermore, in addition to the greater concentration observed in the inland provinces between 1867 and 1900, there is now a

[168] The exceptions were Segovia and Jaén, which although not on the railway line enjoyed a close road connection to the network that minimized the transport costs that these two provinces faced. Badajoz, by contrast, which despite being connected to the railway, presented a similar behaviour to these two provinces, reflecting its peripheral location in the peninsula and the structure of the rail network during this early stage of railway building.

[169] If the self-potential of the market in each province is excluded, Madrid would occupy an intermediate position among the group of inland provinces, at some distance from the greater market potential of the coastal provinces, strengthening, therefore, the perspective of the importance of the size of this province's own market. See table A7 in the Appendix for the market potential figures when self-potential is excluded.

higher degree of dispersion among the coastal provinces. Compared to 1867, the provinces of Barcelona, Vizcaya and Guipúzcoa stand out for their greater market potential, reflecting the industrial development experienced by these provinces in this period.

Figure 3.2. Convergence in market potential of the Spanish provinces, 1867-1900



Source: see text.

Overall, the greater stratification in the distribution of market potential in 1867 evolved towards a more polarised situation in 1900 (Quah, 1996, 1997). The growth shown by the group of provinces with the lowest market potential in 1867, corresponding broadly to those inland provinces without a connection to the rail network in that year, gave rise to a greater degree of concentration among the group of inland provinces. Thus, the convergence found in the earlier period was the result of the approximation among the inland provinces, due to the greater growth of the inland provinces of lesser market potential.

This evolution was caused by the advance in the construction of the rail network. The extension of the railway, which around 1900 linked up all the provincial capitals of Spain, eliminated the disadvantage that some of the provinces had suffered in 1867 by being excluded from the rail network. Their isolation had penalised them with higher transport costs, as they had to resort to road haulage, a more expensive transport mode.

Thus, the reduction in transport costs, linked to the extension of the railway, and, consequently, the progressive integration of the domestic market between 1867 and 1900, was the cause of the convergence that occurred in the levels of market potential of the two subgroups in which the inland provinces had been divided in 1867[170]. This gave rise, in turn, to a polarisation between the Spanish inland and coastal provinces, during this period.

Between 1900 and 1930, the negative relation between the initial market potential and the growth experienced by this variable in the second half of the 19th century had disappeared. In figure 3.3 a moderate tendency towards a divergence in the market potential of the Spanish provinces in the first third of the 20th century is observed[171]. The provinces that in 1900 had the greatest market potential (corresponding to the coastal provinces) saw their market potential increase at a rate of growth slightly higher than that of the inland provinces of lower market potential.

As described above, in terms of market potential at the beginning of the 20th century, the distribution of the provinces showed a clear polarisation with two groups of provinces - the inland and coastal provinces (with the notable exception of Madrid). This pattern was maintained in both 1914 and 1930[172], so that the differences between the market potential of the inland and coastal provinces was a trait that was consolidated over time.

The rapid expansion of the railway during the second half of the 1800s, when all the provincial capitals were finally connected to the network, was to slow down in the early decades of the 20th century. Between 1900 and 1930, the extension of the railway barely caused any changes in the structure of the transport network in Spain[173]. Therefore, the significant reduction in transport costs among the inland provinces between 1867 and 1900 did not continue into the next century. The basic rail network had by then been constructed. Hence, the inland provinces could not benefit from the expansion of the railway as they had in the previous period.

---

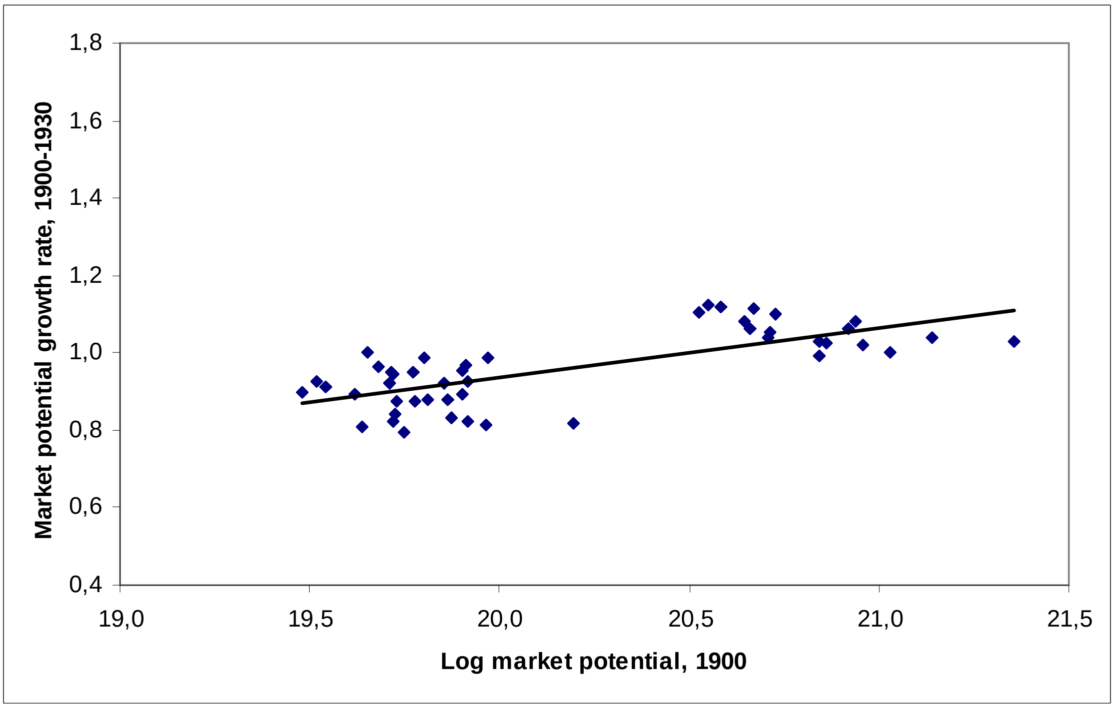
[170] And this despite the fact that the lines constructed before 1866 and which connected the country's main urban centres had a greater impact on market potential than those constructed after this date, according to Herranz (2007a), p. 465.

[171] The increase in the coefficient of variation from 0.59 in 1900 to 0.65 in 1930 reveals, similarly, a greater dispersion or σ-divergence.

[172] This is verified by figure 3.4 presented below.

[173] See section 3.2, in this chapter.

Figure 3.3. Convergence in market potential of the Spanish provinces, 1900-1930



Source: see text.

This description of the patterns of provincial market potential allows us to conclude that the reduction in transport costs was crucial to the dynamics of this variable in the period under study[174]. Thus, the geographical coverage of the railway would seem to have played a key role in the shift experienced by the patterns of market potential in Spain's provinces, especially in the second half of the 19th century. By 1900, the territorial structure of differential market access at the provincial level was well defined, with two groups of clearly differentiated provinces - the inland provinces and the coastal provinces with their greater market potential. In the early decades of the 20th century, not only was this structure maintained, as revealed by maps 3.2, but as the divergence shown in figure 3.3 highlights, it had been strengthened.

According to the theoretical predictions of NEG, better market access attracts the factors of production - both capital (backward linkages) and labour (forward linkages), giving rise to an agglomeration of economic activity in the locations with greatest market potential. As such, the NEG models link market potential directly with regional inequality. This relation between market potential and GDP per capita has been

[174] However, it should be borne in mind that the variation in market potential over time might also have been due to a change in the spatial distribution of GDP, a change in relative transport costs, or to both these factors.

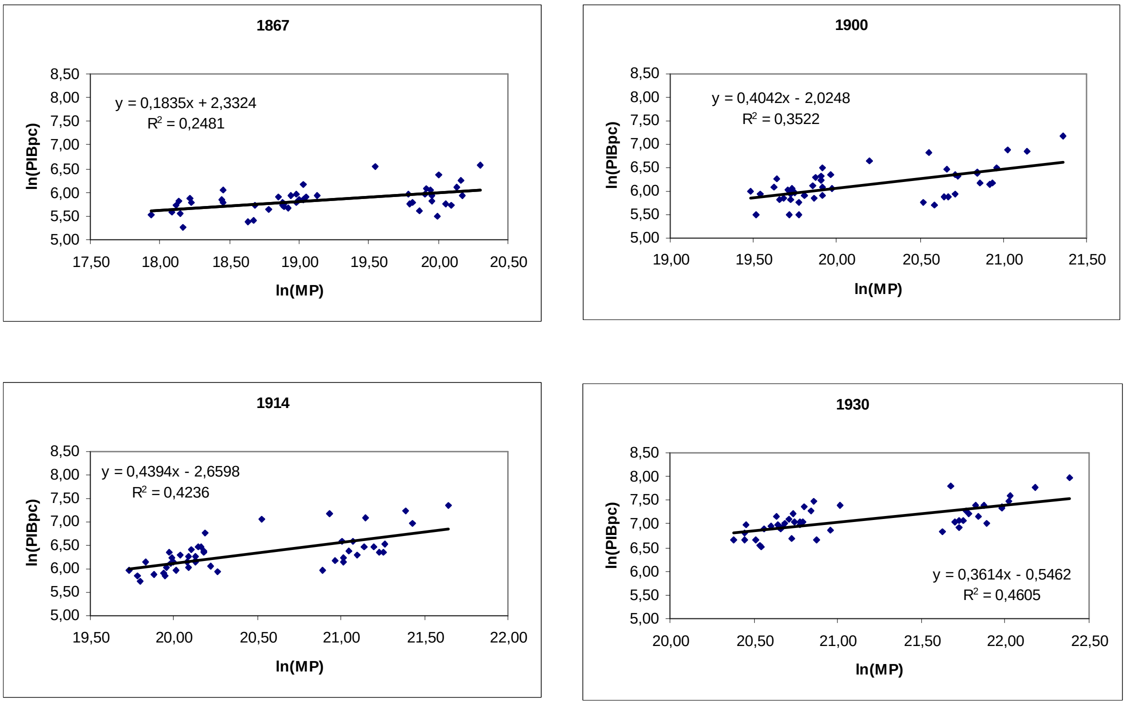
empirically tested in various studies within the NEG. International studies, such as Redding and Venables (2004) and Mayer (2008), establish a link between these variables, so that those countries located near the main markets, and therefore, enjoying better access to them, are the ones that possess a higher income per capita. This link has recently been studied for the Spanish economy by Faíña and López-Rodríguez (2006). They concluded that between 1991 and 2001, market potential and proximity to the centres of development would explain about 60% of the spatial distribution of provincial per capita incomes in Spain. In an initial exploratory attempt at determining whether this relationship existed in the second half of the 19th century and the first few decades of the 20th, figure 3.4 has been plotted.

Figure 3.4 shows that a positive relationship exists between the two variables, confirming that the provinces with the highest market potential are those that had, in turn, the highest GDP per capita throughout the period of study. Furthermore, in terms of the R^2, it can be observed that the explanatory power of market potential increases with time, thereby narrowing the link between the two variables[175].

In the graph corresponding to 1867 marked differences can be seen in the market potential in Spain's provinces. However, the degree of inequality in terms of GDP per capita is not particularly high. A first hypothesis might be the following. This weaker relation between both variables, when compared to the relation at the other dates included in this study, might indicate that in the middle of the 19th century the impact of market potential on provincial inequality was only moderate. This suggests that the NEG forces were not at that time the most important factor when explaining the spatial inequality in per capita income. However, since 1900, and in the years that followed, this relation gradually became more accentuated, increasing the impact of market potential on provincial per capita income.

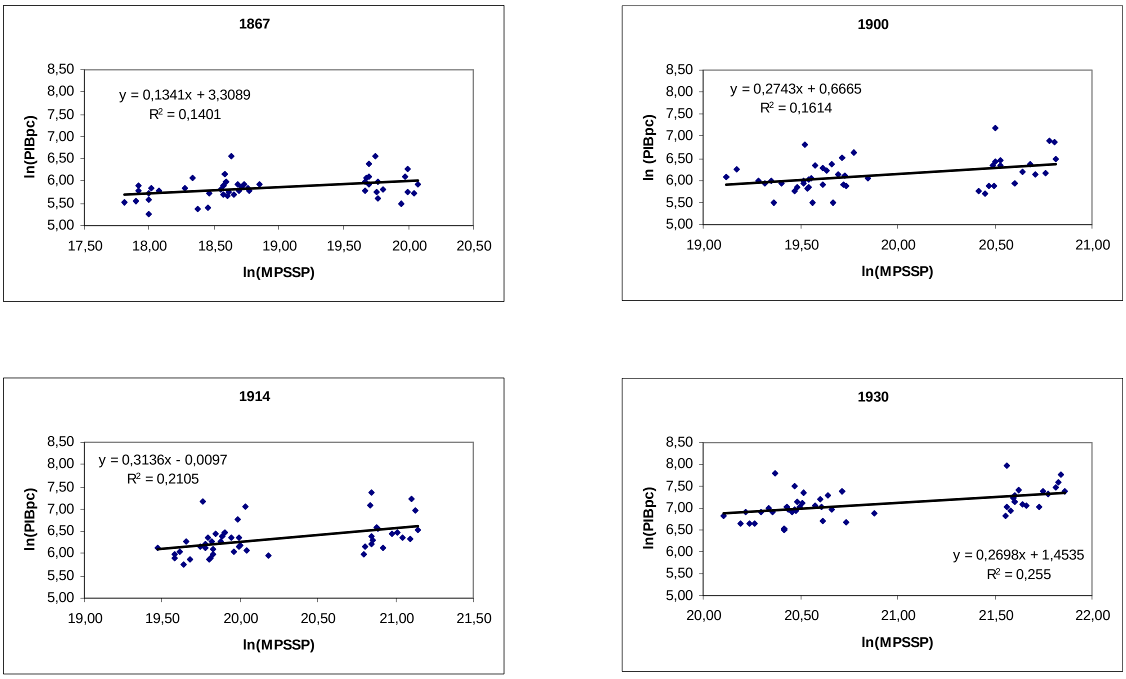
[175] The correlation between market potential and GDP per capita rose from 0.5 in 1867 to 0.59 in 1900, 0.65 in 1914 and reaching 0.68 by 1930.

Figure 3.4. GDP per capita and market potential (1867-1930)



Source: see text.

Figure 3.5. GDP per capita and market potential, excluding self-potential (1867-1930)



Source: see text.

Yet, these results must be treated with a certain amount of caution, because of the type of analysis undertaken[176], and because they relate the provincial GDP per capita with the market potential, and this latter variable includes the provincial GDP in its numerator. Therefore, it is unsurprising that both variables should show a high correlation. In order to avoid the potential problem of circularity in the results, the same graphs were drawn using an alternative measure of market potential, namely, one that excluded each province's self-potential. As can be seen in figure 3.5, the relationship between market potential and GDP per capita, and the patterns described earlier, are maintained, so that proximity to the large markets, even when each province's own market is ignored, appears to be a relevant factor when explaining provincial inequality in terms of income per capita. In this case, however, both the R^2 and the coefficients associated with the market potential experience a reduction with respect to the specification that considers the provinces' self-potential.

The analysis of the patterns of market potential in Spain's provinces has been based essentially on the variation in domestic transport costs. The main changes, arising in the second half of the 19th century, have been linked with the expansion of the rail network[177]. Thus, adopting a regional perspective, an indirect relationship can be established between the railway and economic growth. The existence of this relationship has given rise to a heated debate among Spanish historians. On the one hand, the pessimistic version of the impact of the railway on Spain's economic growth is based on the lack of profits recorded by the railway companies. The premature construction of the railway network which in fact anticipated demand, and the virtual absence of spillover effects for the industrial sector due to the concessions granted to the importation of inputs, minimized the positive impact of the railway on economic growth (Tortella, 1973, 1999; Comín, Martín Aceña, Muñoz and Vidal, 1998). By contrast, the social savings calculations made by Gómez Mendoza (1981, 1982) have led him to offer a more optimistic vision. The savings in transport costs attributable to the construction of the railway in Spain were considerable and even greater than those in other European countries. However, the social savings estimated by this author have been lowered in subsequent studies (Barquín, 1999; Herranz, 2002).

[176] The hypothesis that emerges from this first tentative exercise is tested with more robust instruments in Chapter 5, using conditional growth regressions derived directly from an NEG model.

[177] In addition, the benefits as a result of the possibility the coastal provinces had of exploiting coastal shipping, and their greater proximity in terms of transport costs to international markets, has to be taken into account.

In earlier studies of the Spanish economy carried out within a NEG framework, the integration of Spain's market during the second half of the 19th century has been linked with changes in regional specialisation and the increase in the concentration of economic activity (Tirado, Paluzie and Pons, 2002; Rosés, 2003). This process of market integration seems to have been driven by the reduction in transport costs thanks to the construction of the railways, to which should be added the integration of the labour and capital markets - the latter the result of the monetary and banking reforms implemented during this period, as discussed in detail in the first chapter.

In this study, following the line of research developed by these earlier papers, the elaboration of an indicator of provincial market potential allows studying the relationship between the railway, the reduction in transport costs and regional inequality. The construction and expansion of the railway in the second half of the 19th century would have brought about changes in the geographical structure of market access at the provincial level. This reduction in transport costs, in addition, would have facilitated the integration of the domestic market during this same period. The interaction between transport costs and market accessibility in a monopolistic competition framework might have generated, in accordance with NEG, the emergence of forces that favoured the agglomeration of economic activity in a limited number of territories, especially in the industrial sector, where scale economies gradually increased as industrialisation advanced. Finally, the greater industrial concentration would have led to a more uneven regional distribution of income.

In this sense, market potential has shown itself to be a useful tool for examining empirically some of the theoretical propositions that have arisen within the NEG. In this instance, being able to call on an indicator to measure the differential market access of the Spanish provinces and its evolution overtime is a necessary tool for the exercises proposed in the following discussion. The historical analysis of regional inequality within the framework provided by New Economic Geography can be undertaken using the newly constructed database. First, it provides the opportunity of examining in depth the impact of market access as a determinant of the patterns of industrial location in Spain in the second half of the 19th and the first third of the 20th centuries. The study of these determinants is the objective in Chapter 4. Moreover, the initial conclusions to be drawn from this chapter already link market potential and regional inequality. Examining in greater detail the relationship between market potential and the differences in provincial growth rates is the aim of Chapter 5.

Chapter 4. The determinants of industrial location in Spain, 1856-1929

4.1. 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[178].

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, Tirado and Pons, 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 regions 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, Paluzie, Pons, 2006). But what were the forces driving the location of different industries across Spain in that period?

On the theoretical side, two major explanations stand out in the literature. First, the classical 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. To briefly introduce this model, it can be assumed that there are two

[178] A comprehensive and detailed view of this process, including the market integration of goods and factors, can be found in chapter 1.

production factors (capital and labour). In addition, assuming absence of transport costs, commodities being produced under 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. Therefore, a capital-abundant location will specialise in the production of capital-intensive goods whereas a labour-abundant location will produce labour-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. However, differences in natural resources or factor endowments, especially within countries where the underlying characteristics of the different regions in terms of relative endowments are usually more similar than between countries, may not be large enough to fully explain the high degree of geographical concentration in economic activity observed in reality.

As regards the second explanation, New Economic Geography models highlight that the interaction between transport costs, increasing returns to scale and the size of market under a monopolistic competition framework can lead to the spatial 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*[179].

As shown by Krugman (1991), the concentration of economic activity is a result of the interaction of two 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 location (the ‘home market effect’ as in Helpman and Krugman (1985)), 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 decisions of location 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.

[179] A comprehensive survey of NEG theoretical models can be found in chapter 2. Here, the more relevant features are summarised.

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 final goods. In this case, the lower transport costs make inputs cheaper. When intermediate goods are considered (Krugman and Venables, 1995; Venables, 1996) the presence of input-output linkages emerges as an alternative mechanism favouring agglomeration. 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 also analysed by Puga (1999)[180]. This author showed that the relationship between the process of regional integration and the degree of concentration in the economic activity can describe a bell-shaped 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. Therefore, NEG models predict the existence of a bell shaped relationship between the process of market integration and the degree of concentration of industrial activity in the territory.

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[181]. The pioneering work of Kim (1995) examined the long term evolution in the location of manufacturing activities and its determinants in the US between 1860 and 1995, a country where such activities became more spatially concentrated until the end of the 1920s. Nevertheless, from that moment onwards, the trend was reversed and a new pattern of continuous reduction in the geographical concentration until 1987 made

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

[181] Here, the empirical exercises reviewed in section 2.2.c in chapter 2 that are directly related with the exercise to be undertaken are briefly summarised to put the debate in context.

manufacturing more spatially dispersed across the United States[182]. At the same time, regional specialisation increased from 1860 to World War I (in spite of the decrease in 1880 and 1890), flattened out in the interwar years and then decreased since the end of War World II until 1987[183]. As a result, the evolution of manufacturing concentration and regional specialisation in the US described a bell-shaped curve[184]. Therefore, in the first stages of development, from 1860 to 1930, a period in which regional economies in the United States were integrating into a national economy, industry became more localized as regions became more specialised. In this sense, the US experience, at least in its first stages, shows a similar pattern when compared to the Spanish case earlier described[185].

In order to explain the determinants of this evolution in the US, Kim (1995) estimated an equation where the endogenous variable, an index of industrial spatial concentration, was regressed on a measure of resource endowments (the intensity in the use of raw materials) and a measure of economies of scale (size of establishment) using panel data for the US states, including 5 years between 1880 and 1987, and 20 industries. 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[186].

As regards Spain, Tirado, Pons and Paluzie (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 (economies of scale and market access) were playing an increasing role in shaping the industrial map of Spain.

On another hand, the ongoing process of supranational integration in Europe has raised some concerns about the potential effects on the location of industrial activities,

[182] Hoover's coefficient of localisation for the US states had a value of 0.273 in 1860, reached a peak of 0.316 in 1927 and then decreased to 0.197 in 1987. Thus, in this last year, the level of geographical concentration of manufacturing was lower than in 1860.

[183] Again, the evolution of the Krugman specialisation index shows that the US states were less specialised in 1987 (0.43) than in 1860 (0.69).

[184] For France, Combes, Lafourcade, Thisse and Toutain (2008) found, as in the case of the US a non-monotonic evolution in the geographical concentration of manufacturing. The analysis, focused on the French *départements*, showed that spatial concentration of manufacturing increased considerably from 1860 to 1930 and then decreased in 2000, and the level of dispersion in this last year was higher than in 1860.

[185] Nevertheless, Marshallian externalities were more present in the US case before World War II (Kim, 1995), whereas in Spain Betrán (1999) has shown the relevance of externalities *à la Jacobs* in the interwar years.

[186] The same evidence was found in Kim (1999) where a model based on the Rybczynski theorem was estimated.

since this may imply unevenly adjustments for the territories involved in this process[187]. Midelfart-Knarvik, Overman, Redding and Venables (2002) studied the changes in the geographical distribution within the European Union between the 1970s and the 1990s. In this period, there was a slight decrease in the spatial concentration of manufacturing across EU-15. In addition, the industrial structures of these countries became more similar in the 1970s, but then diverged from the beginning of the 1980s until the mid-1990s. The contribution of Midelfart-Knarvik, Overman, Redding and Venables (2002) was not only to analyse the determinants behind the evolution of these industrial patterns but also to develop a new analytical framework. The standard model of Midelfart-Knarvik, Overman and Venables (2000) allowed the use of an alternative empirical strategy to explore the underlying forces that determine industrial location combining standard international trade theory and economic geography variables[188]. This approach presents several advantages over the exercise applied by Kim (1995) and Tirado, Pons and Paluzie (2002)[189]. In this case, the equation to be estimated is directly derived from a theoretical model. Furthermore, a much larger number of variables are included in the equation including region and industry characteristics, and more importantly, the interaction between them. In particular, the role of the market potential in industrial location decisions, a key element in NEG explanations, can be tested directly. Finally, the model not only nests both Heckscher-Ohlin comparative advantage and NEG factors but it also tests the relative strength of these forces as drivers of industrial location. Their results showed that between 1970 and 1997 in the EU the supply of skilled workers and researchers became increasingly important in attracting industries that used that factor intensively. On the NEG side, the capacity of attraction of central locations for industries with high increasing returns was significant but decreasing over time. Finally, industries with high shares of intermediate goods in production moved to regions with good access to markets.

Some studies in economic history have carried out a similar exercise implementing the empirical framework developed by Midelfart-Knarvik, Overman, Redding and Venables (2002). First, Wolf (2007) made use of this model in his analysis of the determinants of the industrial relocation in interwar Poland after the reunification

[187] In this regard, as argued below, studies in historical perspective might be useful to understand the dynamics and the effects of economic integration on industrial location.

[188] Other approaches in empirical exercises in this field are Davis and Weinstein (1999, 2003), Ellison and Glaeser (1999), Brülhart and Torstensson (1996), Amiti (1999) and Haaland, Kind, Midelfart-Knarvik and Torstensson (1999).

[189] See Combes, Mayer and Thisse (2008) for a more detailed description.

in 1918. The integration of the domestic market led to important changes in regional specialisation and the spatial concentration of manufacturing[190]. In this case, Wolf (2007) found that both Heckscher-Ohlin factor endowments and NEG mechanisms were economically relevant in explaining these changes in the location of industry in the new Polish state. Particularly, the most important factor for industrial location between 1926 and 1934 was skilled labour endowment and the availability of innovative activities, proxied by patent announcements, was also significant. As regards NEG forces, there was evidence of a forward linkage[191].

Likewise, Crafts and Mulatu (2005, 2006) also explored the determinants of industrial location in Britain estimating an equation based on Midelfart-Knarvik, Overman, Redding and Venables (2002). According to their results, again Heckscher-Ohlin and NEG forces played a relevant role in explaining the location of industry within Britain between the 1870s and the beginning of the 20th century. However, the British experience is somewhat different when compared to the other countries analysed and in particular, to the Spanish case. In the second half of the 19th century the process of industrialisation had developed further and both the geographical concentration of industry and regional specialisation revealed an evolution characterised by stability[192]. The econometric results confirmed the significance of the availability of natural resources and factor endowment variables in shaping the industrial location. The interaction capturing the endowment for agriculture was statistically significant prior to World War I and they also found a strong impact for the human capital. Nevertheless, the influence of market potential and economies of scale on location decisions since 1871, that is, in a period of intense fall in transport costs, was also significant albeit this interaction weakened over time and in 1931, it was no longer significant. Besides, no evidence of linkage effects was found by these authors. Therefore, in the British case, the pattern of industrial location would have responded to factor endowments and natural resources, although it was reinforced by NEG forces.

[190] Krugman's specialisation index for manufacturing increased from 0.71 in 1925 to 0.77 in 1937. The index of spatial concentration showed a much more stable pattern and on average there was a slight increase in concentration, although some industries became more spatially dispersed (*minerals*, *wood*).

[191] In a previous version (Wolf, 2004), the interaction of market potential and forward linkages was significant when a further interaction with economies of scale was considered. In both cases, the forward linkages capture the intensity in the use of intermediate inputs.

[192] The Krugman index for specialisation was 0.63 in 1841, increased to 0.66 in 1871 and then decreased to 0.61 in 1911. As for the localisation index in manufacturing, this indicator showed that spatial concentration did not increase in the same dates, and had a constant value of 0.23.

Following a similar approach, Klein and Crafts (2009) have recently questioned some of the conclusions obtained by Kim (1995) for the US. 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 US. In particular, among the variables which capture factor endowments, only agriculture abundance was significant, and prior to 1900. Yet, the availability of skilled labour force and coal were not statistically significant. However, NEG forces had a major impact on industrial location in the period considered. The interaction between market potential and economies of scale was a crucial element for industrial location and linkage effects eventually gained explanatory power and became the main driving force in 1920. Thus, Klein and Crafts (2009) offered an alternative explanation of the genesis of the manufacturing belt in the US, where NEG mechanisms were the major drivers of the spatial concentration of manufacturing activities. In sum, the picture that emerged from this study was that the relative market potential of the US states was a key element in explaining the location of manufacturing in the US between 1880 and 1920; it was more important than factor endowments; and its influence came from both scale and linkage effects, although the latter became increasingly important over time. Besides, these authors contributed to substantially improve the empirics of the original paper.

In this chapter the empirical strategy developed by Midelfart-Knarvik, Overman, Redding and Venables (2002) is applied in order to study the Spanish experience in a period when the integration of the domestic market was completed and a notable increase in the spatial concentration of manufacturing activities was recorded. Therefore, the Spanish case, with its particular characteristics, can be added to other relevant study cases in countries that have been analysed within this framework (the US, Britain and Poland) in order obtain a broader picture. Moreover, this exercise when compared to previous studies on the forces driving industrial location in Spain covers a larger period, going from mid-19th century until the outbreak of the Spanish Civil War[193]. In addition, a more thorough empirical approach, which is directly derived from a theoretical model, is used. This alternative approach may help to examine whether the new evidence confirms or not the results in previous studies[194]. Finally, looking back to the creation of the national

[193] As described in the next section, Rosés (2003) focused on 1861, Tirado, Paluzie and Pons (2002) analysed the second half of the 19th century, and Betrán (1999) studied the interwar years.
[194] For instance, the results of Klein and Crafts (2009) for the US are at odds with those of Kim (1995).

markets from a historical perspective can shed light on the potential outcomes that the current process of European integration can generate on the geographical changes in manufacturing activities.

Thus, in the next pages the determinants of industrial location in Spain are empirically examined. The analysis is focused on four benchmark years (1856, 1893, 1913, and 1929), seven industries and 43 Spanish provinces[195]. Then, an equation similar to the one originally proposed by Midelfart-Knarvik, Overman, Redding and Venables (2002) is estimated. This equation nests both factor endowments comparative advantage and NEG forces, whose combined effect is captured through a series of interactions between region and industry characteristics.

Then, the aim is to shed light on some appealing questions regarding the location of industry in Spain between 1856 and 1929: What were the underlying forces that determined the increase in the geographical concentration of industry in Spain in the years under study? Was market potential, as suggested by NEG theory, a relevant variable to explain the location of industry in a period of falling transport costs? What was the role of comparative advantage and geography in driving the changes in industrial location? Were Heckscher-Ohlin elements and/or NEG forces at work? If so, what was the relative strength of these two explanations? Did the relative strength change as the Spanish market became more integrated? And finally, if there is evidence in favour of the NEG hypotheses, did the impact on the location of industry come from scale and/or linkage effects?

The remainder of the chapter is structured as follows. In the next section, the main changes in industrial location in Spain during the second half of the 19th century and the first decades of the 20th century are surveyed. Then, in section 3, the empirical strategy is defined. First, the Midelfart-Knarvik, Overman, Redding and Venables (2002) model is introduced; second, the variables considered in the exercise and the sources consulted are detailed; third, the variables included are briefly described. In section 4, the estimation results are shown. At this point, different robustness tests are carried out and alternative specifications are presented in order to deal with some econometric issues incorporating the empirical improvements 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. Finally, in the last section the main results are discussed.

[195] From a total of 50 provinces (NUTS3). The exclusions are due to data restrictions as it will be discussed in section 4.3.2.

4.2. Transport costs and industrial location in Spain

As detailed in the first chapter, during the second half of the 19th century the Spanish market became increasingly integrated. In this period, the construction of the railway network and its expansion played a key role in the integration of the different regional economies. Traditionally, transport in Spain 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)[196]. Apart from the ruggedness of the land, the rivers are characterised by their poor and irregular flow and inland navigation has been very limited. 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, inadequately designed and in a poor state of conservation, and inland navigation, the cheapest means of transport before the railway era, was almost non-existent. The limits imposed by geography to these transport means were, however, partially counterbalanced by the availability of coastal navigation. Located in the Iberian Peninsula, the total length of the Spanish coastline is around 8,000 km.

In this context, the expansion of the railway network brought about important changes that favoured the process of market integration. The main effect of the new infrastructures was the decrease in transport costs[197]. The ratio between the unit price of railway transport for commodities and the alternative mean of transport was around 0.14% in 1878 (Herranz, 2005). In addition, coastal navigation, which represented around a 20% of the volume of commodities transported by railway over this period (Frax, 1981), also experienced a decrease in freight rates[198]. Besides, motorized road transport was at an infant stage at this time in Spain and the process of substitution of railways by trucks started in the 1930s (Herranz, 2005)[199].

---

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

[197] The fall in transport costs was one of the main features in the world economy during the First Globalization (O'Rourke and Williamson, 1999).

[198] Gómez Mendoza (1981), p. 57. See also Herranz (2004), p. 61-62.

[199] All these aspects have been analysed in depth in chapter 3, where the construction of the market potential estimates for the Spanish provinces in historical perspective has been detailed.

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, Pons and Tirado (2004) analysed the spatial distribution of industry across Spain in the long term. Measured through the Gini and Hirschman-Herfindhal indexes, the geographical concentration of industrial production increased from 1856 to 1929 in parallel with the deeper integration of the Spanish domestic market. 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. As a result, the Spanish experience shows a bell-shaped non-monotonic relationship between market integration and the spatial distribution of industry in the long run. In particular, it was in the first stages of the industrialisation process, at the time that market integration was completed, that industrial production became increasingly concentrated in a limited number of provinces.

Tirado, Pons and Paluzie (2006) further analysed the geographical pattern of this industrial concentration by computing industrial intensity indices in order to identify the provinces in which manufacturing production was concentrating[200]. Their results showed that the number of provinces with an index above one, that is, provinces with an industrial specialisation, decreased between 1856 (14 provinces) and 1893 (9 provinces), and also from 1913 (8 provinces) to 1929 (7 provinces). Therefore, in the mid-19th century manufacturing activities exhibited a high degree of dispersion in space. However, the integration of the Spanish market and the take-off of the industrialisation process led to radical changes in the localisation and the spatial concentration of industry. During the second half of the 19th century an industrial axis 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 Zaragoza emerged as industrial centres together with Catalonia, the Basque Country, Madrid and Valencia[201].

When individual industries were studied, the evidence provided by Paluzie, Pons and Tirado (2004) showed that most of the seven industries considered shared an increase

[200] This index is calculated as the ratio *share of industrial output / share of population*.

[201] In 1929, the seven provinces with an industrial intensity index above one were Barcelona, Madrid, Vizcaya, Guipúzcoa, Valencia, Zaragoza and Santander. These authors have linked this change to the protectionist turn of the Spanish trade policy in the 1890s (Tirado, Pons and Paluzie, 2009).

in the geographical concentration between 1856 and 1929. The only exceptions were *wood and 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*.

Likewise, Tirado, Pons and Paluzie (2006) examined the effects of the Spanish market integration on regional specialisation. These authors calculated a bunch of indicators in order to describe the main patterns of regional specialisation at a provincial level in four benchmark years (1856, 1893, 1913 and 1929), considering seven industries: *foodstuffs*, *textile and leather*, *metallurgy*, *chemicals*, *paper*, *glass and ceramics*, and *wood and furniture*[202]. On the basis of the weighted Krugman specialisation index, regional specialisation increased in Spain 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. The same conclusion was reached by Betrán (1999) in her study of the industrial location in Spain during the interwar years. Regional specialisation did not increase between 1913 and 1929 since most of the provinces showed a stable degree of specialisation and in some cases even a reduction[203].

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. In this sense, several empirical studies have tried to explain the forces driving the industrial location during the second half of the 19th century when a notable increase in the geographical concentration of industry was in motion.

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.

---

[202] This data, which comes from the *Estadísticas Administrativas de la Contribución Industrial y de Comercio*, will be used as well in the empirical exercise undertaken in this chapter. Interestingly, this source was also used in seminal studies like Nadal (1987) and in recent studies like Parejo (2001).

[203] This is based on the weighted average of the specialisation index. The value decreased from 0.439 in 1913 to 0.395 in 1929. However, when the simple average is taken the index went from 0.499 to 0.516 in the same dates. Betrán (1999), p. 679.

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, Paluzie and Pons (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 was 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, Overman, Redding and Venables (2002) allows an analysis of the determinants of industrial location in Spain and complement the studies described in the final part of this section. This approach presents some advantages: first, it is based on a sound empirical test in the sense that it is directly derived from a theoretical model; and second, the exercise covers a much longer period, the years between the mid-19th century and the Spanish Civil War and therefore, the entire period of increasing spatial concentration in manufacturing activities in Spain can be examined.

4.3. Empirical strategy

4.3.1. The model

The approach to be implemented in order to study the spatial distribution of industry in Spain between 1856 and 1929 is based on the model suggested by Midelfart-Knarvik, Overman, Redding and Venables (2002) in which, in equilibrium, the pattern of industrial location is determined both by factor endowments and economic geography. In addition, the equation to be estimated includes region and industry characteristics. This model departs from the idea that 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 provinces are relatively abundant in land; others are relatively abundant in labour; and some provinces 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 proposed by Midelfart-Knarvik, Overman, Redding and Venables (2002) nests both comparative advantage and New Economic Geography effects through a set of interactions capturing region and industry characteristics as determinants of industrial location. The specification derived by these authors can be expressed as:

$$\ln(s_i^k) = \alpha \ln(pop_i) + \ln(man_i) + \sum_j \beta[j](y[j]_i - \gamma[j])(z[j]^k - \kappa[j]) \quad (4.1)$$

where s_i^k is the share of industry *k* in province *i* (see data section); pop_i is the share of Spain's population living in province *i*; man_i is the share of the total of Spain's manufacturing located in province *i*; $y[j]_i$ is the level of the *j*th region (province) characteristic in province *i*; $z[j]^k$ is the industry *k* value of industry characteristic paired with region characteristic *j*. The remaining terms in the summation capture the interaction between region and industry characteristics. Finally, α, β, $\beta[j]$, $\gamma[j]$ and $\kappa[j]$ are coefficients.

To clarify the intuition behind this equation, the usual example is reproduced[204]. Consider one particular characteristic j = skilled labour, so $z[skilled\_labour]^k$ is the skilled labour intensity of industry k (in this case, white-collar worker intensity) and $y[skilled\_labour]_i$ is the skilled labour abundance of province i. First, there exists an industry with skilled labour intensity, $z[skilled\_labour]^k$, the location of which is independent of the skilled labour abundance of provinces. Second, there exists a level of skilled labour abundance, $y[skilled\_labour]_i$, such that the province's share of each industry is independent of the skilled labour intensity of the industry. Third, if $\beta[skilled\_labour] > 0$, then industries with skilled labour intensity greater than $z[skilled\_labour]^k$ will tend to locate in provinces with skill labour abundance greater than $y[skilled\_labour]_i$, and away from provinces with skilled labour abundance less than $y[skilled\_labour]_i$.

Equation (1) cannot be estimated in that form. Expanding the relationship gives the equation to be estimated:

$$\ln(s_i^k) = c + \alpha \ln(pop_i) + \beta \ln(man_i) + \sum_j (\beta[j] y[j]_i z[j]^k - \beta[j]\gamma[j] z[j]^k - \beta[j]\kappa[j] y[j]_i) \qquad (4.2)$$

The key parameters to be estimated are detailed below. The first two variables (pop_i, man_i) take into account size effects meaning that, all other things equal, we would expect larger provinces to have a larger industrial share in any given industry. Thus, the coefficients α and β are straightforward, and c is a constant term. The estimated coefficients for region (province) characteristics and industry intensities, $y[j]$ and $z[j]$, are estimates of $-\beta[j]\kappa[j]$ and $-\beta[j]\gamma[j]$, respectively. Thus, they are expected to have negative signs. If we divide these estimates by $\beta[j]$, we obtain an estimate of the cut-off point defining high and low abundance and intensity. However, the most relevant information comes from the interaction between region and industry characteristics since these interactions determine industrial location. The estimated coefficients of the

[204] Here, the discussion follows Midelfart-Knarvik, Overman, Redding and Venables (2002), p. 245; Crafts and Mulatu (2005), p. 504; Crafts and Mulatu (2006), p. 597; and Klein and Crafts (2009), p. 8.

interaction variables, $y[j]_i z[j]$, are estimates of $\beta[j]$, the sensitivity of industry location to the different region (province) characteristics. For this reason, we concentrate on the study of the parameters $\beta[j]$[205].

The analysis of the determinants of industrial location in Spain between 1856 and 1929 is based on the region and industry characteristics displayed in table 4.1. The first three variables in the column where region characteristics are included capture the 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 provinces[206]. The second factor endowment variable refers to the abundance of labour in each province measured by total active population per square kilometre[207]. Educated population, the third Heckscher-Ohlin variable, measures the relative endowment of human capital proxied 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 displayed in the second column in table 4.1, the first three variables are based on Heckscher-Ohlin theory. The first two variables reflect factor intensities (land and labour) and the third one captures the intensity in the use of 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.

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 three first interactions are based on comparative advantage theory. The Heckscher-Ohlin model predicts that industries using intensively a factor of production will tend to be located in provinces abundantly endowed with that factor. Therefore, these interactions take into account that industries in which agricultural inputs, labour, and skilled labour are used

[205] This approach has, nonetheless, some limitations linked to the existence of multiple equilibria in NEG models. 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, Overman and Venables (2000), p. 3.

[206] "*We take agricultural production as an exogenous measure of 'agriculture abundance' (rather than going back to an underlying endowment such as land)*". Midelfart-Knarvik, Overman, Redding and Venables (2002), p. 243.

[207] As usual in this type of exercise, capital is not included among the variables considered based on the assumption of capital mobility across regions.

intensively will tend to be established in locations with a better relative endowment of these factors. Thus, a positive sign for these interactions is expected.

Table 4.1. Region and industry characteristics

Region characteristics (provinces)	*Industry characteristics*
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

The next three interactions consider NEG-type mechanisms. The interaction between *market potential x size of establishment* captures the idea 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. Therefore, the sign of this interaction is expected to be positive. If transport costs were either very high or very low, then the pull of market potential may be weakened.

The interaction *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, in order to minimize transport costs, locate near central areas where a higher number of suppliers will be concentrated. Thus, a bigger market implies better access to suppliers, and again, a positive sign is expected for this term.

Finally, firms may also produce goods for other industrial users. The interaction *market potential x sales to industry* takes into account the presence of 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 minimize transport costs. However, a priori, the sign of this interaction is not clear. The backward linkages induce firms to be near the industrial customers, but firms may want to be located near final consumers as well[208].

4.3.2. Data

The analysis of the determinants of the spatial distribution of industry in Spain is carried out for a total of 43 provinces (NUTS3)[209], 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 i, defined as:

$$s_i^k(t) = \frac{x_i^k(t)}{\sum_k x_i^k(t)} , \qquad (4.3)$$

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*" [210]. 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.

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

[208] In this case, backward and forward linkages refer to the links between producers in the chain of production. Thus, this meaning is different from the definition presented in chapter 2, also abundant in the theoretical literature, where backward and forward linkages denote the agglomeration forces pulling to attract capital and labour, respectively.

[209] 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.

[210] Nadal and Tafunell (1992), p. 256. [Own translation].

had a special fiscal regime[211]. Second, this tax underwent major changes in 1907[212]. 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 the database constructed in chapter 3, where regional GDP estimates were estimated based on Geary and Stark's (2002) methodology[213]. This new database covers NUTS3 provinces and 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 Census of Population (1860, 1900, 1910 and 1930).

Industry characteristics are derived from various sources. A good number of the variables reported in table 4.1 are based on input-output relationships. Unfortunately, the first input-output table for the Spanish economy was published in 1958. Accordingly, 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, 1984) made use of this source in constructing the historical industrial production index for Spain, and Prados de la Escosura (2003) employed this information in his reconstruction of historical series for

[211] 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.

[212] "*The act of 03-08-1907 represents a break in the history of the taxation on manufacturer or industrial activities, since it establishes that joint stock companies and limited partnership by shares devoted to the production may pay the tax 'Impuesto de Utilidades', a tax on property passed in 1900. From 1921 this prescription was extended to every mercantile society [...]. The nature of this latter tax is much more different from the previous one: it rests on the (net) profits of societies, and it does not take into account the means of production nor the supposed income generated by them. This fact raises a serious problem because from the year 1907 onwards the information contained in the industrial contribution books presents an incomplete image of industrial activities. And, what is absent constitutes an important part of the biggest companies. Besides, it is a growing part of companies because of the conversion of many societies in joint stock companies in order to benefit from the benevolent tax treatment offered by the 'Impuesto de Sociedades' (Corporate Income Tax)*" [Own translation]. Nadal and Tafunell (1992), p. 259.

[213] See chapter 3 for a detailed description. See also Crafts (2005a).

the Spanish GDP starting in 1850[214]. Hence, agricultural input as a percentage of total costs, labour input as a percentage of Gross Value Added, intermediate goods as a percentage of output and sales to industry 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 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 1929[215]. Therefore, for this set of industry characteristics considered, the information is not time-varying; the only partially time-varying industry characteristic is the size of establishment[216].

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 following the methodology suggested by Geary and Stark (2002). Labour abundance is measured through the total active population in each province per square kilometre. In this case, information on the total active population by province has been compiled from the different Census 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 rate offered by Núñez (1992). Lastly, estimates of market potential for the Spanish provinces have been constructed[217]. Regional accessibility is measured through the Harris market

---

[214] The information relative to the Gross Value Added in different industries came from the input-output table of 1958 and it was used to weight the industrial production index for the period between 1850 and 1958. Carreras (1983, 1984).

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

[216] 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 again, we use information on intensities that is not time-varying*". Midelfart-Knarvik, Overman, Redding and Venables (2002), p. 34.

[217] Full information on the market potential estimates for the Spanish provinces can be found in chapter 3.

potential equation, following the recent work by Crafts (2005b) for Britain from 1871 to 1931[218]. Market potential estimates are provided for 47 NUTS3 provinces in different years in the 1860s, 1900, 1914 and 1930[219].

4.3.3. Comparative advantage and NEG variables: a description of the patterns

Before implementing the empirical strategy, in this subsection the main patterns and evolution of the variables considered in the exercise are briefly described. Following the order shown in table 4.1, region characteristics are examined first. As regards comparative advantage, the relative abundance of land across provinces is measured by agricultural production. In this sense, in 1860 the picture that emerges from the maps 4.1-4.4 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 a growing polarisation over the period considered (Núñez, 1992). In 1860, literacy rates were higher in the area going from Madrid, through Castilla-Leon (Burgos, Palencia, Segovia, Valladolid and Soria) and northern Spain (Cantabria and Álava)[220]. By 1900, literacy

[218] See also Keeble, Owens and Thompson (1982). For historical exercises, see Crafts (2005b) and Schulze (2007). In the case of Wolf (2007), estimates of both market access (*MA*) and supplier access (*SA*) 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.

[219] As mentioned above, the three insular provinces (Baleares, Las Palmas de Gran Canaria and Santa Cruz de Tenerife) are excluded from the market potential calculations.

[220] Interestingly, although Barcelona was the main industrial province, it was not among the high literacy provinces.

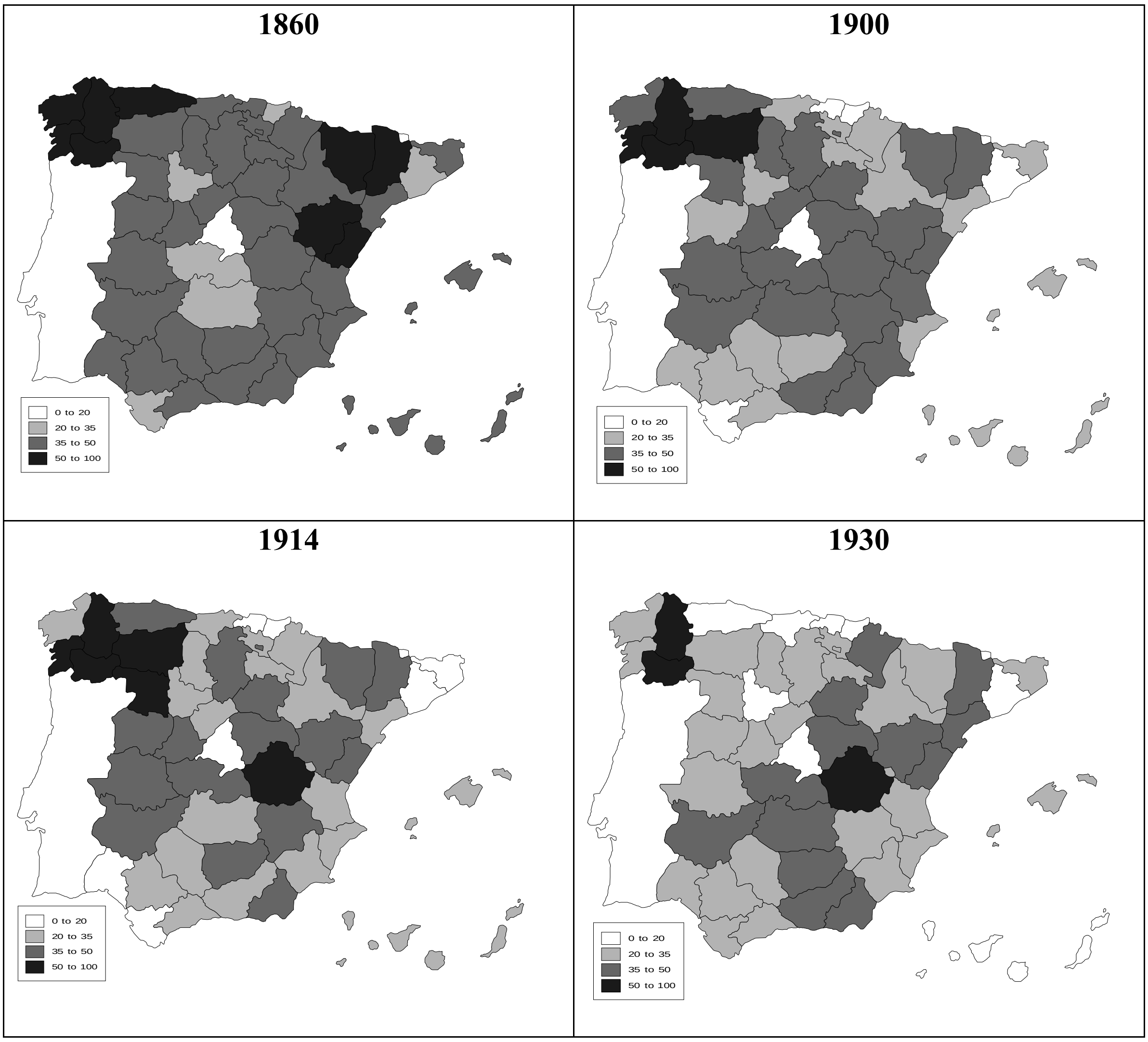
rates in Catalonia and Aragon had risen, leading to a marked division of the country between the north and the south in terms of education. The only exceptions were Galicia in the north-west with a lower literacy rate than the rest of the northern provinces, and south-western Andalusia (Cádiz, Huelva 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 its accessibility to markets. 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[221]. 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 persisted in the first three decades of the 20th century.

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*.

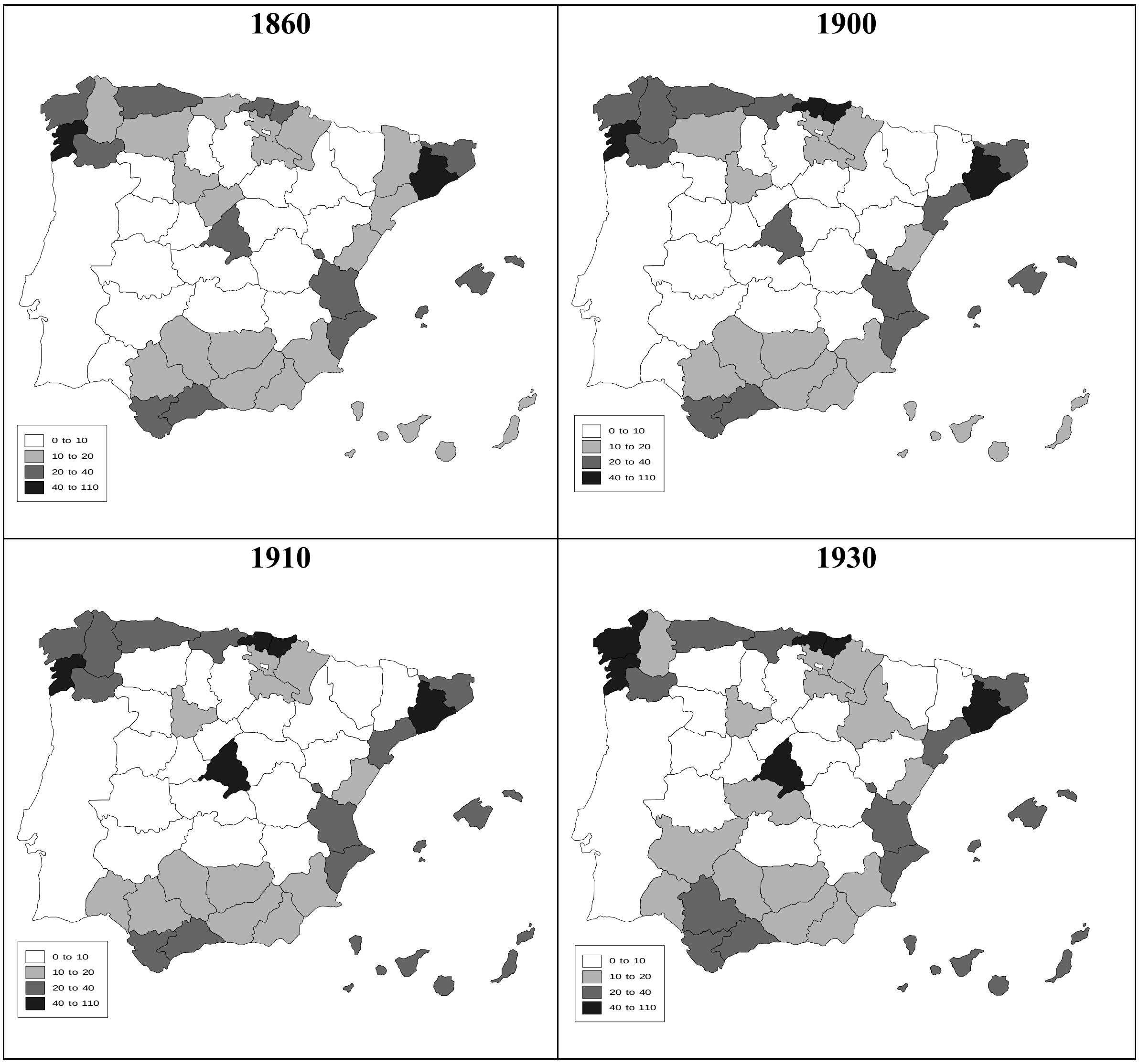
[221] Market potential estimates have been constructed and discussed in chapter 3. Here, the main results are summarised. The maps are expressed in relative terms, with Barcelona, the highest market potential province, being equal to 100.

Maps 4.1. Gross Value Added in Agriculture as a % of GDP, Spanish provinces



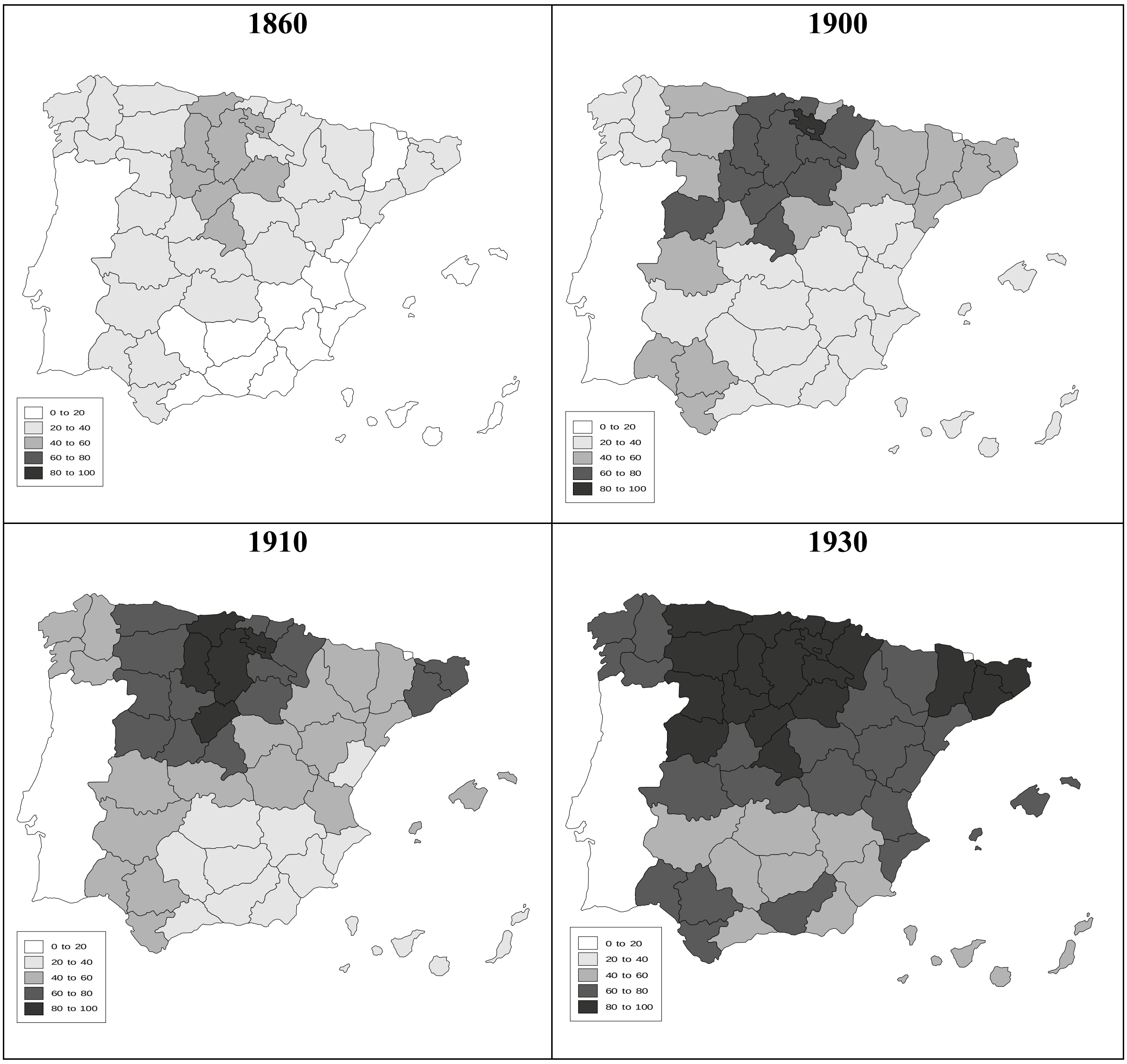
Source: see text.

Maps 4.2. Total active population per square km, Spanish provinces



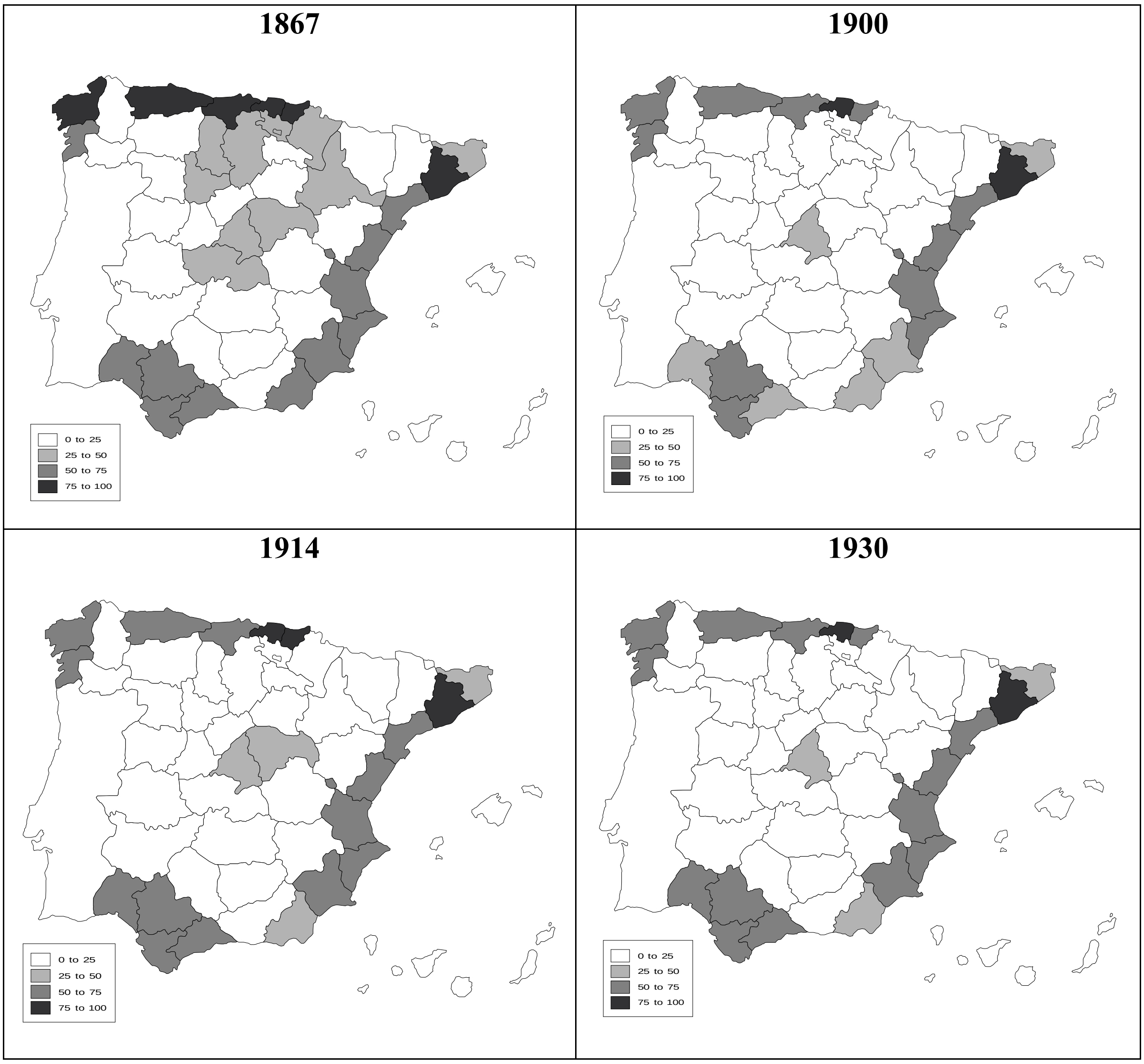
Source: see text.

Maps 4.3. Literacy rate, Spanish provinces



Source: Own elaboration based on Núñez (1992).

Maps 4.4. Market potential, Spanish provinces (Barcelona=100)



Source: see text and chapter 3.

4.4. Estimation results

Equation (4.2) 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 used[222]. On the other hand, the observations included in the regression are not completely independent since we have seven industries for each province, that is, seven observations corresponding to the same province. This may

[222] The Breusch-Pagan/Godfrey test (table 4.2) confirms the presence of heteroskedasticity since the null hypothesis of homoskedasticity is rejected at the 1% significance level in the years considered (5% in 1893).

generate problems with the standard errors obtained in the estimation[223]. The within group or intra-cluster dependence can be addressed using cluster-robust standard errors, 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 4.2 shows these first estimation results, including the constant term, the coefficients of the two size variables (pop_i, man_i), the four region characteristics, $y[j]$, the six industry characteristics, $z[j]$, and the six interaction variables, $\beta[j]$. The 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[224]. When the whole period is considered, two of the six interactions included are statistically significant. 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 expected positive sign. Therefore, these results suggest that both HO and NEG factors had a significant impact on the location of industry in Spain.

However, pooling data across years can generate problems since it implicitly assumes that the parameters of the equation are constant over time[225]. This concern is important in our 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 may be too strong and needs to be tested. In order to decide whether it is appropriate or not to pool the data, a Chow test is conducted[226]. The calculated value of the F-statistic F(19, 1119) is 12.75423 which allows rejecting the null hypothesis at the

[223] "*Research that ignores potential correlation between respondents sharing the same cluster may draw distorted inferences*". Pepper (2002), p. 342. A good example can be found in table 4.2, where the standard error, and therefore, the significance of the interaction capturing skilled labour in 1913 changes when cluster-robust standard errors are used.

[224] Albeit changes in the standard errors in columns 'a' and 'b', which take into account heteroskedasticy-robust and cluster-robust standard errors, respectively, the significance of the interaction variables is not affected.

[225] "*…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 β[j] changes*". Midelfart-Knarvik, Overman and Venables (2000), p. 16.

[226] 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).

1% significance level that there is no structural break. Hence, equation (4.2) 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 4.2. 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 0.56 to 0.66, is acceptable when compared to other similar exercises[227]. 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. 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 were higher in regions with a lower degree of industrialisation at the start of the industrialisation process[228]. 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

[227] In this case, the values of the adjusted-R^2 are close to the ones obtained for the British case (Crafts and Mulatu, 2005, 2006) and also for the United States (Klein and Crafts, 2009).

[228] Note that this interaction is significant in column 'a' in 1913, but it is not when the estimation is corrected taking into account cluster-effects.

market potential and the share of intermediate inputs used (forward linkages) and between market potential and sales to industry (backward linkages) are always statistically insignificant and in some years have the wrong sign. Therefore, 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.

Table 4.2. Regression results. Dependent variable ln (s_{it}^{k})

	Pooled		1856		1893	
	a	b	a	b	a	b
Constant	82.155	82.155	42.866	42.866	-267.011	-267.011
	(53.954)	(75.518)	(106.783)	(134.033)	(183.823)	(230.675)
Population	0.192	0.192	0.069	0.069	-0.019	-0.019
	(0.126)	(0.139)	(0.343)	(0.272)	(0.219)	(0.189)
Manufacturing	-0.106	-0.106	-0.109	-0.109	-0.061	-0.061
	(0.092)	(0.088)	(0.231)	(0.173)	(0.164)	(0.139)
Share of agricultural GVA	-0.099	-0.099	-0.829**	-0.829***	-0.180	-0.180
	(0.112)	(0.121)	(0.396)	(0.265)	(0.228)	(0.186)
Labour abundance	-0.604	-0.604	-0.149	-0.149	-2.139**	-2.139**
	(0.537)	(0.762)	(1.307)	(1.360)	(1.010)	(1.007)
% Educated population	0.755**	0.755*	2.246**	2.246**	-0.196	-0.196
	(0.374)	(0.389)	(0.888)	(1.012)	(0.786)	(0.870)
Market potential	-1.016	-1.016	-2.130	-2.130	14.508	14.508
	(8.836)	(13.081)	(20.405)	(25.881)	(30.311)	(38.407)
Agricultural input use	-0.741***	-0.741***	-2.094***	-2.094***	-1.155***	-1.155***
	(0.107)	(0.148)	(0.322)	(0.320)	(0.217)	(0.267)
Share of labour in GVA	-6.423***	-6.423***	-15.182***	-15.182***	-8.181***	-8.181***
	(0.943)	(1.153)	(2.633)	(2.807)	(1.516)	(1.491)
% White-collar workers	-2.814***	-2.814***	-6.508***	-6.508***	-2.070	-2.070
	(0.719)	(0.997)	(1.807)	(1.943)	(1.460)	(1.608)
Size of establishment	-1.734***	-1.734***	-1.256	-1.256	-6.733***	-6.733***
	(0.284)	(0.269)	(1.021)	(0.881)	(1.201)	(1.233)
Intermediate input use	-4.855	-4.855	8.953	8.953	56.059*	56.059
	(8.827)	(12.322)	(17.354)	(21.283)	(30.585)	(38.510)
Sales to industry	-6.005	-6.005	0.918	0.918	27.210*	27.210
	(4.649)	(6.442)	(9.250)	(11.643)	(15.931)	(19.604)
Share of agricultural GVA	0.116***	0.116***	0.212***	0.212***	0.066	0.066
** Agricultural input use*	(0.023)	(0.285)	(0.076)	(0.066)	(0.047)	(0.056)
Labour abundance	0.144	0.144	0.046	0.046	0.572**	0.572**
** Share of labour in GVA*	(0.146)	(0.210)	(0.353)	(0.370)	(0.277)	(0.279)
Educated population	-0.178	-0.178	-0.935**	-0.935**	0.117	0.117
** White-collar workers*	(0.152)	(0.158)	(0.373)	(0.417)	(0.317)	(0.360)
Market potential	0.223***	0.223***	-0.051	-0.051	0.602***	0.602***
** Size of establishment*	(0.047)	(0.048)	(0.185)	(0.156)	(0.183)	(0.187)
Market potential	-0.099	-0.099	0.385	0.385	-2.643	-2.643
** Intermediate input use*	(1.455)	(2.155)	(3.339)	(4.169)	(5.023)	(6.392)
Market potential	0.164	0.164	0.183	0.183	-1.617	-1.617
** Sales to industry*	(0.760)	(1.116)	(1.768)	(2.276)	(2.606)	(3.244)
Number of observations	1157	1157	261	261	295	295
Adjusted R^2	0.51	0.51	0.66	0.66	0.63	0.63
BPG test chi2(18)	124.225***		52.662***		33.576**	
F-test joint significance	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 4.2. (continued)

	1913		1929	
	a	b	a	b
Constant	-171.518	-171.518	90.443	90.443
	(190.523)	(226.947)	(164.906)	(182.720)
Population	0.132	0.132	0.052	0.052
	(0.254)	(0.204)	(0.248)	(0.156)
Manufacturing	0.013	0.013	-0.052	-0.052
	(0.201)	(0.160)	(0.180)	(0.958)
Share of agricultural GVA	0.344	0.344	-0.008	-0.008
	(0.273)	(0.267)	(0.133)	(0.905)
Labour abundance	-3.455***	-3.455***	-0.298	-0.298
	(0.940)	(0.994)	(0.924)	(0.112)
% Educated population	1.499**	1.499*	1.737	1.737
	(0.717)	(0.886)	(1.142)	(1.458)
Market potential	31.460	31.460	-6.642	-6.642
	(29.330)	(35.890)	(23.129)	(26.664)
Agricultural input use	-0.525**	-0.525**	-0.162	-0.162
	(0.251)	(0.253)	(0.160)	(0.162)
Share of labour in GVA	-7.489***	-7.489***	2.397**	2.397**
	(1.321)	(1.111)	(1.211)	(1.107)
% White-collar workers	-0.548	-0.548	4.221**	4.221
	(1.319)	(1.666)	(2.065)	(2.541)
Size of establishment	-6.060***	-6.060***	-2.932***	-2.932***
	(1.103)	(1.084)	(0.958)	(0.893)
Intermediate input use	37.315	37.315	-13.141	-13.141
	(31.809)	(37.628)	(27.363)	(30.082)
Sales to industry	19.238	19.238	-10.716	-10.716
	(16.254)	(19.396)	(13.993)	(15.689)
Share of agricultural GVA	0.063	0.063	0.106***	0.106***
** Agricultural input use*	(0.062)	(0.062)	(0.032)	(0.030)
Labour abundance	0.960***	0.960***	0.118	0.118
** Share of labour in GVA*	(0.257)	(0.272)	(0.252)	(0.300)
Educated population	-0.505*	-0.505	-0.504	-0.504
** White-collar workers*	(0.289)	(0.359)	(0.467)	(0.569)
Market potential	0.826***	0.826***	0.299**	0.299**
** Size of establishment*	(0.157)	(0.153)	(0.123)	(0.118)
Market potential	-5.558	-5.558	0.698	0.698
** Intermediate input use*	(4.864)	(5.927)	(3.825)	(4.380)
Market potential	-3.179	-3.179	0.603	0.603
** Sales to industry*	(2.484)	(3.064)	(1.952)	(2.278)
Number of observations	300	300	301	301
Adjusted R^2	0.56	0.56	0.56	0.56
BPG test chi2(18)	55.187***		45.400***	
F-test joint significance	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.

4.4.1. Robustness, alternative specifications and standardised coefficients

In this section, different robustness tests are carried out in order to confirm the previous findings, since some econometric issues may 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 province. To test this potential measurement error, an alternative specification including region and industry fixed effects is estimated. 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 coefficients and significance of the interaction variables are virtually unchanged (table 4.3) and therefore the results previously shown in table 4.2 are robust when region and industry fixed effects are included.

Table 4.3. 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***		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 4.3. (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***		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%.

At this point, another potential problem needs to be addressed. The 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).

When some regressors are endogenous, OLS estimation generally results in inconsistent estimators and so the method of instrumental variables (IV) provides a general solution to the problem of endogeneity (Wooldridge, 2002; Cameron and Trivedi, 2005). Therefore, an endogeneity test on the potentially endogenous regressors is performed. In this context it is difficult to find an appropriate instrument, that is, an observable variable which is correlated with the endogenous explanatory variables but not with the residuals. As in previous exercises (Klein and Crafts, 2009), the IV estimation relies on lagged variables for the market potential. However, in this case, the complexity of obtaining homogeneous estimates for market potential for the Spanish provinces before the 1860s restricts the IV 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 and the IV estimates are appropriate, a weak instrument test is conducted. Then, equation (2) is estimated using 2-step GMM which is preferred to the IV/2SLS[229].

First, the Durbin-Wu-Hausman endogeneity test rejects the null hypothesis that market potential and its interactions are exogenous in 1893 and 1913 but exogeneity is not rejected in 1929[230]. In addition, the weak instruments test suggested by Stock and Yogo is based on the Cragg-Donald F-statistic whose values in all the years considered exceed the critical values in the tables provided by these authors (Stock and Yogo, 2005) leading to the rejection of the null hypothesis of weak instruments. The results of these tests are reported in table 4.4 together with the 2-step GMM estimation of equation (2) 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 4.2), confirming the validity of the previous results[231].

[229] "*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 4.2).

[230] The statistic test follows a chi-squared distribution with degrees of freedom equal to the number of explanatory variables tested for endogeneity.

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

Table 4.4. Estimation results: 2-step GMM/IV. Dependent variable ln (s_{it}^{k})

	1893	1913	1929
Constant	-355.315* (-1.823)	-139.919 (-0.724)	120.155 (0.721)
Population	-0.002 (-0.008)	0.145 (0.569)	0.032 (0.132)
Manufacturing	-0.077 (-0.464)	0.002 (0.008)	-0.038 (-0.215)
Share of agricultural GVA	-0.179 (0.787)	0.339 (1.237)	-0.007 (-0.052)
Labour abundance	-2.097** (-2.024)	-3.424*** (-3.664)	-0.251 (-0.270)
% Educated population	-0.214 (-0.273)	1.497** (2.085)	1.737 (1.518)
Market potential	28.453 (0.896)	26.683 (0.903)	-10.733 (-0.457)
Agricultural input use	-1.133*** (-5.248)	-0.529** (-2.089)	-0.170 (-1.058)
Share of labour in GVA	-8.126*** (-5.355)	-7.455*** (-5.645)	2.419** (1.994)
% White-collar workers	-2.093 (-1.431)	-0.547 (-0.413)	4.240** (2.052)
Size of establishment	-6.670*** (-5.223)	-6.053*** (-5.431)	-2.882*** (-2.918)
Intermediate input use	70.683** (2.174)	32.240 (0.998)	-18.007 (-0.651)
Sales to industry	34.195** (2.027)	16.438 (0.999)	-13.233 (-0.934)
*Share of agricultural GVA * Agricultural input use*	0.060 (1.289)	0.065 (1.032)	0.108*** (3.428)
*Labour abundance * Share of labour in GVA*	0.554* (1.967)	0.948*** (3.710)	0.110 (0.434)
*Educated population * White-collar workers*	0.124 (0.390)	-0.505* (-1.741)	-0.508 (-1.088)
*Market potential * Size of establishment*	0.592*** (3.029)	0.825*** (5.195)	0.292** (2.286)
*Market potential * Intermediate input use*	-4.955 (0.939)	-4.793 (-0.977)	1.365 (0.352)
*Market potential * Sales to industry*	-2.721 (0.997)	-2.757 (-1.103)	0.948 (0.478)
Number of observations	295	300	301
Adjusted R^2	0.63	0.56	0.56
DWH Endogeneity Test: Chi-square (4)	7.868*	8.607*	5.932
Cragg-Donald F-statistic	359.281	4209.759	4818.758

Note: White heteroskedasticity-robust standard error in brackets. * Significant at 10%; ** significant at 5%; *** significant at 1%.

On the other hand, 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 or iron ore were located in the provinces of Vizcaya and Santander in the north, and in Málaga in the south[232]; lead ore mines existed in southern Spain (Murcia, Jaén, Almería, Córdoba, Granada, Badajoz and Ciudad Real); copper was abundant in the south (Huelva)[233]; there were mercury mines in Almadén (Ciudad Real), in the south western area of Castilla, next to Andalusia; and finally, coal was mainly concentrated in the north (Asturias and León) and in some southern provinces (Ciudad Real and Córdoba)[234].

Hence, a new interaction variable (*mineral resources abundance x mineral resources intensity*) is included in the equation[235]. Mining production by province is obtained from the Spanish Statistical Yearbooks (*Anuarios Estadísticos de España - AEE*) for the years 1860, 1915 and 1931[236]. 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 are interesting on two respects when compared to the previous estimates (table 4.5). First, with the new specification the significance of the interaction variables and the magnitude of the coefficients are not qualitatively altered[237]. Therefore, the results are robust to this alternative specification.

232 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.

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

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

235 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.

236 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).

237 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%).

Table 4.5. Robustness Check II: Alternative specification. Dependent variable $\ln(s_{it}^{k})$

	1856		1893	
	a	b	a	b
Constant	-202.203*	-202.203	-238.636	-238.636
	(114.858)	(138.389)	(191.766)	(233.786)
Population	0.007	0.007	-0.017	-0.017
	(0.358)	(0.288)	(0.218)	(0.201)
Manufacturing	-0.056	-0.056	-0.128	-0.128
	(0.232)	(0.185)	(0.184)	(0.174)
Share of agricultural GVA	-0.748**	-0.748**	-0.266	-0.266
	(0.357)	(0.280)	(0.229)	(0.217)
Mineral resources abundance	-0.045	-0.045	-0.0004	-0.0004
	(0.045)	(0.040)	(0.023)	(0.022)
Labour abundance	-0.248	-0.248	-2.165**	-2.165**
	(1.426)	(1.473)	(1.094)	(1.117)
Market potential	-1.315	-1.315	14.501	14.501
	(20.845)	(26.809)	(31.188)	(38.413)
Agricultural input use	-0.190	-0.190	-0.467	-0.467
	(0.314)	(0.244)	(0.315)	(0.287)
Mineral resources intensity	1.613***	1.613***	0.525**	0.525**
	(0.266)	(0.255)	(0.233)	(0.220)
Share of labour in GVA	4.014***	4.014***	-2.796*	-2.796*
	(1.129)	(1.171)	(1.552)	(1.570)
Size of establishment	-0.162	-0.162	-5.213***	-5.213***
	(1.075)	(0.940)	(1.432)	(1.494)
Intermediate input use	36.504*	36.504	46.507	46.507
	(18.641)	(22.266)	(31.984)	(31.916)
Sales to industry	10.661	10.661	21.671	21.671
	(9.795)	(12.098)	(16.811)	(20.056)
Share of agricultural GVA	0.159*	0.159**	0.077	0.077
** Agricultural input use*	(0.084)	(0.072)	(0.047)	(0.055)
Mineral resources abundance	0.012	0.012	0.007	0.007
** Mineral resources intensity*	(0.020)	(0.020)	(0.011)	(0.014)
Labour abundance	0.065	0.065	0.588**	0.588*
** Share of labour in GVA*	(0.385)	(0.400)	(0.299)	(0.308)
Market potential	-0.051	-0.051	0.583***	0.583***
** Size of establishment*	(0.189)	(0.157)	(0.191)	(0.192)
Market potential	0.256	0.256	-2.594	-2.594
** Intermediate input use*	(3.401)	(4.312)	(5.164)	(6.375)
Market potential	0.116	0.116	-1.643	-1.643
** Sales to industry*	(1.813)	(2.359)	(2.695)	(3.262)
Number of observations	241	241	274	274
Adjusted R^2	0.64	0.64	0.61	0.61
F-test joint significance	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 4.5. (continued)

	1913		1929	
	a	b	a	b
Constant	-101.289	-101.289	167.797	167.797
	(202.800)	(229.298)	(181.548)	(205.651)
Population	0.141	0.141	-0.139	-0.139
	(0.270)	(0.249)	(0.228)	(0.140)
Manufacturing	-0.054	-0.054	-0.014	-0.014
	(0.212)	(0.199)	(0.188)	(0.115)
Share of agricultural GVA	0.229	0.229	-0.129	-0.129
	(0.303)	(0.330)	(0.130)	(0.090)
Mineral resources abundance	-0.0002	-0.0002	0.026	0.026
	(0.022)	(0.024)	(0.038)	(0.038)
Labour abundance	-3.319***	-3.319***	0.164	0.164
	(0.933)	(1.025)	(0.912)	(1.020)
Market potential	26.601	26.601	-19.440	-19.440
	(30.563)	(36.089)	(25.183)	(29.585)
Agricultural input use	0.481	0.481	-1.022***	-1.022***
	(0.339)	(0.328)	(0.230)	(0.191)
Mineral resources intensity	0.744***	0.744***	-0.435*	-0.435**
	(0.194)	(0.171)	(0.222)	(0.185)
Share of labour in GVA	0.540	0.540	-3.791***	-3.791***
	(1.351)	(1.158)	(1.300)	(1.095)
Size of establishment	-3.983***	-3.983***	-3.949***	-3.949***
	(1.226)	(1.024)	(1.139)	(1.003)
Intermediate input use	19.723	19.723	-19.574	-19.574
	(33.845)	(37.703)	(30.253)	(33.995)
Sales to industry	9.342	9.342	-12.930	-12.930
	(17.268)	(19.603)	(15.338)	(17.552)
Share of agricultural GVA	0.066	0.066	0.117***	0.117***
** Agricultural input use*	(0.066)	(0.068)	(0.028)	(0.025)
Mineral resources abundance	0.032***	0.032**	0.022	0.022
** Mineral resources intensity*	(0.012)	(0.014)	(0.019)	(0.021)
Labour abundance	0.925***	0.925***	0.009	0.009
** Share of labour in GVA*	(0.255)	(0.282)	(0.249)	(0.270)
Market potential	0.800***	0.800***	0.235*	0.235*
** Size of establishment*	(0.159)	(0.154)	(0.131)	(0.121)
Market potential	-4.676	-4.676	2.823	2.823
** Intermediate input use*	(5.072)	(5.935)	(4.182)	(4.875)
Market potential	-2.830	-2.830	1.702	1.702
** Sales to industry*	(2.588)	(3.097)	(2.114)	(2.505)
Number of observations	259	259	280	280
Adjusted R^2	0.57	0.57	0.55	0.55
F-test joint significance	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%.

Second, the estimation adds a role for another comparative advantage variable. In 1913, the interaction capturing the relevance of mineral resources is significant and shows a positive sign. Thus, industries that used intensively mineral resources were responsive to the provincial endowment of such 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 located[238]. Nevertheless, in 1929, this effect is no longer present.

These results can shed some light on a more general debate. Sachs and Warner (2001) argued that developing countries in the second half of the 20th century did not benefit from the abundance of natural resources. However, a positive relationship between natural resources and industrialisation 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 clear[239]: Klein and Crafts (2009) did not find evidence of such relationship in the period 1880-1920[240]. As regards Spain, the benefits of being endowed with mineral resources have been questioned in view of the low linkages effects that mining activities produced on the industrial sector, with the exception of iron ore in Vizcaya[241]. In this regard, the results show that over the period considered, with the only exception of 1913, industries that made intensive use of mineral resources did not tend to locate in mineral resource abundant provinces.

Finally, the relative strength of the forces shaping the location of industry in Spain needs to be analysed. So far, the analysis has focused on the significance and the signs of the interaction variables capturing Heckscher-Ohlin and NEG mechanisms. The natural

[238] However, coal is only a part of the total mineral resources in Spain.

[239] 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).

[240] "*Both coal and skilled-labor interactions change signs and are insignificant for most of the time*". Klein and Crafts (2009), p. 20.

[241] This pessimistic view has been stressed by authors like Vicens Vives (1959), Sánchez Albornoz (1968), Nadal (1975) and Chastagnaret (2000). However, a more optimistic view has been defended by Tortella (1981), Coll (1985) and Prados de la Escosura (1988). 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).

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 4.6). 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 expressed in the same unit (standard deviations)[242]. The higher value of the beta coefficient for an independent variable indicates a higher impact 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.

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) th

at of comparative advantage.

Table 4.6. Standardised beta coefficients. Interaction variables

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

Source: see text. Statistically significant variables in bold.

[242] Following Klein and Crafts (2009), beta coefficients are calculated as follows: 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*. These coefficients are calculated from the regressions in table 4.2 (OLS). This specification has been selected on the basis of the higher goodness of fit. In addition, Schwarz and Akaike Info criterions show lower values and therefore favour this selection.

4.5. Discussion

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 the country 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 foodstuffs[243], and thus spatial distribution of industry was determined by comparative advantage. More precisely, according to the results, at that time, the location of industry was driven by the relative endowment of land, measured through agricultural production. The interaction that relates agricultural abundance and agriculture input use is the only positive significant interaction variable and thus, it shows that manufacturing industries that used intensively agricultural inputs were located in provinces well endowed for agriculture.

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 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 (Venables, 1996; Puga, 1999). 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. The first great impulse in the construction of the railways started after the Railway Act of 1855 was passed and lasted until the 1870s (Herranz, 2005). By 1893, the main economic centres and a large number of provincial capitals were connected to the network. The result was an important decrease in transport costs in a country where geographical conditions had

[243] In 1856, the provinces of Barcelona (11.04%) and Girona (28.86%), in Catalonia, were the only ones where *foodstuffs* accounted for less than 50% of the total manufacturing production. 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.

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[244].

In addition, at the end of the century, factor endowments were still relevant, but in this case, through the interaction considering labour abundance. Hence, 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 products[245]. 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 this 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 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, Pons and Paluzie, 2006) in a context where the crop of cereals was declining and the Mediterranean regions gradually increased their production[246].

244 NEG models show that when transport costs are high, agglomeration forces are low and firms tend to be dispersed across space in order to save transport costs. When transport costs are intermediate, centripetal forces intensify agglomeration when workers are mobile (Puga, 1999).

245 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.

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

Besides, the strength of the scale effects in the interwar years declined 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[247].

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[248]. 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.

Similarly, no impact for the human capital is found. The comparative advantage interaction capturing the availability of skilled labour appears to be significant only in 1856 and 1913 but with a negative sign. Therefore, education, measured through the literacy rate, was not an important factor during this stage of the industrialisation process. As described above, the geographical pattern of literacy in Spain shows that it was higher in provinces where manufacturing activities were not predominant, as has been noted in the literature (Núñez, 1992)[249].

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 market

[247] The Gini coefficient for the geographical concentration of manufacturing went from 0.44 in 1856 to 0.60 in 1893, and then, from 0.68 in 1913 to 0.78 in 1929. Tirado, Pons and Paluzie (2006), p. 49.

[248] 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.

[249] 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 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).

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 chapter, the analysis of the interactions suggests that NEG forces were not strong enough in 1856 to exert a pull on industrial location[250].

In turn, Tirado, Paluzie and Pons (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 integrated[251]. 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 extended in this chapter until the 1930s using an empirical approach which is more theoretically sound to analyse industrial location. In this case, the agglomeration force driving such changes in the industrial location in Spain is captured by the interaction between market potential and economies of scale[252].

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

[250] However, Rosés (2003) applied an alternative empirical strategy and a different regional aggregation, where historical Spanish regions were considered. Nevertheless, in this chapter it has been 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.

[251] 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.

[252] 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 that she analysed was internal to provinces.

location of industry in Victorian Britain[253], 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 than comparative advantage. This could be linked to the fact that in Spain (1856-1929) changes in the regional specialisation and the concentration of manufacturing were much more profound than in Britain (1871-1931). 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 created, were, to a large extent, similar to the mechanisms present in the manufacturing sector in the European Union in recent times (Midelfart-Knarvik, Overman, Redding and Venables, 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 time[254].

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 skilled labour, and when this 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.

---

[253] In the Spanish case agriculture is also significant in 1856 and 1929. In addition, in the alternative specification where mineral resources abundance is considered, this variable is significant in 1913.

[254] Innovation activities have not been considered in the Spanish case.

As seen, several economic history studies have applied the empirical strategy developed by Midelfart-Knarvik, Overman, Redding and Venables (2002). Overall, this approach has proved to be very useful to explore the forces driving industrial location in different historical cases. Likewise, these studies (especially Klein and Crafts, 2009) have 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.

To sum up and conclude, the results can shed some light on the questions raised in the introduction. 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 as determinants of the industrial map of Spain; the evidence suggests that both Heckscher-Ohlin and NEG forces were 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 explained the spatial distribution of industry in Spain, which at that time showed a high level of dispersion. As the domestic market became more integrated, the impact of NEG forces increased and in 1893 NEG scale effects were already the main driving force behind of the geographical concentration of industry in Spain. Then, there was an intensification of the influence of increasing returns until World War I, which nonetheless declined in the interwar years, although at the end of the 1920s still was the main driver of industrial location. The presence of this agglomeration force (Krugman, 1991) could therefore explain the notable spatial concentration of industrial activity in Spain in the 19th century and prior to the outbreak of the Spanish Civil War. Finally, no evidence of linkage effects is found.

Thus, the exercise carried out in this chapter has shown that market potential was a relevant variable to explain the location of industry in Spain in the first stages of the industrialisation process as the domestic market became more integrated. As stressed by NEG models, the combined effect of regional access to demand and the gradual increase in the average size of plants (economies of scale) was a key aspect to understand the profound changes in the spatial distribution of industry in Spain, i.e., the increase in the geographical concentration of industry. In this regard, the exercise reinforces, with a theoretically sound empirical approach, the view that access to markets played a role in determining the industrial map of Spain in the second half of the 19th century and it expands the influence of this NEG-type mechanism until the interwar years. But was the impact of market potential limited to the industrial sector? Did market potential have a similar effect at a more aggregate level when income per capita is considered? In other words, did geography also have an impact on regional inequality in the period under study? Was market potential not only responsible for the uneven spatial distribution of industry but also for the differences in regional income per capita growth rates? These questions are addressed in the next chapter.

Chapter 5. Market potential and economic growth in Spain, 1860-1930

5.1. Introduction

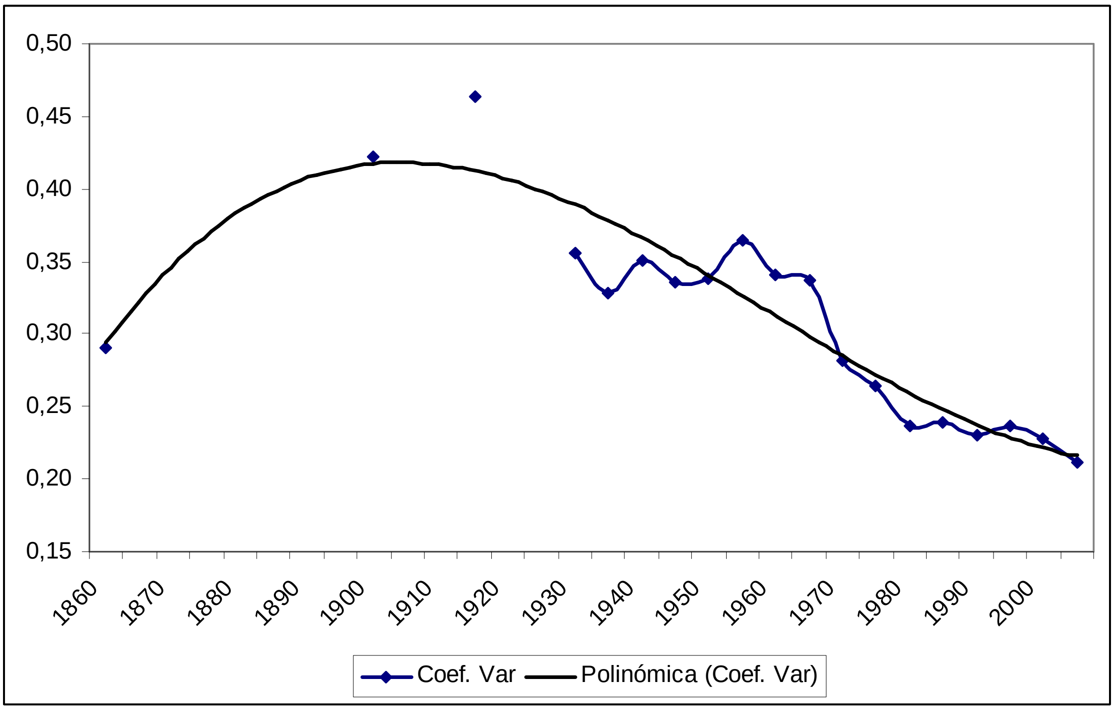
In the opening pages of this thesis, a first approach to regional inequality in income per capita in Spain in recent times was undertaken. As in many other countries in the European context, the economic growth experienced in the last decades by the Spanish economy did not come along with a reduction of regional inequalities within Spain (Puga, 2002; Cuadrado-Roura, 2010). In short, at the end of the first decade of the 21st century, regional inequality is still a striking and persistent feature of the Spanish economy. But when did regional inequality begin and what has been its evolution over time? In Spain, regional income per capita is well documented since 1955, when the Banco Bilbao Vizcaya first published the series of regional and provincial GDP (BBV, 1999, 2000). With the new database constructed in chapter 3 for the period going from 1860 to 1930 following Geary and Stark's (2002) methodology, it is possible to provide a first tentative answer to the previous question, going back to the mid-19th century[255].

Taking the coefficient of variation as a measure of inequality, figure 5.1 shows the long term evolution of regional income per capita disparities at a NUTS3 level, which corresponds to the Spanish provinces. As it can be observed, figure 5.1 illustrates the existence of a bell-shaped curve in the evolution of regional inequality over time. This result confirms the evidence found by Williamson (1965) who, in an empirical exercise using an international sample of countries, proved that in the first stages of development regional disparities may arise, whereas in more mature stages of growth a convergence pattern is found. In the Spanish case, the second half of the 19th century witnessed a

[255] Although regional disparities existed already at the end of the 18th century (Llopis, 2001), it was in the second half of the 19th century, when the Spanish market became integrated, that the contemporary pattern of inequality in Spain was created (Carreras, 1990a). For the following description, the series are completed with Alcaide (2003) for the period 1935-1950 and Funcas (2006) for 2000 and 2005.

remarkable increase in regional income inequality. Then, in the first decades of the 20th century, this process came to a halt and a tendency towards the reduction of income per capita inequality is observed. Nevertheless, this convergence was temporary; in the first years under the Franco regime the coefficient of variation remains almost stable and still above its initial value of 1860. It was at the end of the 1960s that a period of convergence started. However, in the beginning of the 1980s this process came to a halt and from that moment onwards, over the last two decades of the 20th century, the pattern of regional convergence was again interrupted.

Figure 5.1. Long term regional per capita GDP inequality. Spain's NUTS3



Source: see text.

What are the determinants behind this evolution of regional income inequality? Neoclassical growth models predict the existence of convergence in the long run. From a methodological point of view, most of the empirical research within the growth literature has relied on the estimation of growth regressions where the existence of β-convergence and σ-convergence has been tested (Barro, 1991; Barro and Sala-i-Martin, 1991, 1992, 1995). As regards the first concept, it assumes that there is an inverse relation between the growth rate and the initial income per capita, so for a set of economies the growth rate generates a tendency towards convergence, i.e., the initially poorer countries will grow faster than the richer ones. Therefore, convergence occurs as poorer countries exhibit higher rates of growth over time than the most prosperous ones. The driving force of such convergence is the presence of diminishing returns to physical and human capital implicit in the neoclassical production function. But economies differ, among other things, in their levels of technology, their savings rate or their population growth rates, and therefore they have different steady states[256]. Thus, it might be the case that β-convergence only holds when it is conditioned to a set of variables as a proxy of the different steady states (Mankiw, Romer and Weil, 1992; Barro and Sala-i-Martin, 1995; Sala-i-Martin, 1996)[257]. This is the concept of conditional β-convergence as opposed to the absolute β-convergence defined above. Besides, the existence of σ-convergence implies a reduction of the income per capita dispersion over time for a sample of countries[258]. If instead of countries the study focuses on regions within the same state, convergence will take place at a faster pace, as they are more homogeneous units which have historically shared the same institutions, culture and economic policy, and therefore, a more similar steady state across the regional economies can be expected. The more homogeneous are the regions considered, the more likely will be that convergence occurs.

Alternative theoretical approaches have emphasised the role of geography as a driving force behind income inequality. So far, in the previous chapters, the analysis has focused on New Economic Geography but at this point two different views regarding geography need to be introduced: 'first nature' and 'second nature' geography, as labelled by Krugman (1993). The first concept takes into account pure geography elements such as the environmental, ecological or physical conditions of countries. Some authors have argued that geographical conditions have represented overwhelming obstacles for economic growth in developing countries (Gallup, Sachs and Mellinger, 1999; Gallup and Sachs, 2001; Sachs and Warner, 2001; Rappaport and Sachs, 2003). According to these studies, geographical location and climate have significant effects on income growth through different channels. First, a mountainous topography hinders transport and imposes limits on agricultural activities. Climatic conditions have to be taken into account as well because they have a direct effect on agricultural productivity and the population settlement. Likewise, in tropical areas, diseases like malaria and yellow fever have a

[256] Hence, "*…the Solow model is perfectly compatible with income divergence*". Temple (1999), p.123.
[257] An alternative option in the empirical tests for the existence of β-convergence is to restrict the analysis to a subset of countries where the assumption of a similar steady state is not unrealistic.
[258] Figure 5.1 is an example of measuring σ-convergence where the coefficient of variation captures the level of dispersion in income per capita across Spanish provinces.

negative impact on the economic outcome of the countries. In this regard, when looking at the cross-country differences in income levels most of the richest countries in the world are located in tempered areas. As for the geographical location, access to sea and navigable rivers has been stressed to be positive for growth through lower transport costs which, in turn, enhance trade. By contrast, hinterlands and landlocked countries have to face a locational disadvantage that may hamper their economic growth.

Geography was considered to play a crucial role for economic development by classical economists[259]. However, geography variables are seldom included in cross-country growth studies in the line of Barro and Sala-i-Martin (1995). Sachs and his co-authors, by including geography in their growth regressions, concluded that the differences in income per capita across regions around the world can be largely attributable (albeit not exclusively) to these 'first nature' geography elements.

This strand of the literature stresses the direct impact of nature on economic outcomes. There is, however, an indirect effect of pure geography through the interaction with past historical events. Acemoglu, Johnson and Robinson (2001) claimed that geography influenced the institutions created by the settlers during the colonization period. In high mortality areas, where tropical diseases were present, there was a higher risk for colonizers. The lower settlement of Europeans in these territories then resulted in weak institutions, which are at the root of the poor institutional and economic performance of these countries today[260]. Engerman and Sokoloff (2002) argued, in turn, that institutions are determined by factor endowments, soil conditions and climate. That was the case of the English colonies where there was a relationship between crops (for instance, cotton or sugar), slavery and institutions. In a recent paper, Nunn and Puga (2009) suggest that although the ruggedness of land is seen as negative for economic development, in the case of Africa, it had a positive effect from a historical point of view. In the period of slave trade, rugged areas provided protection to the local inhabitants, and thus reduced slave exports. Since slave trade also negatively affected the quality of the

[259] Adam Smith stated: "*As by means of water carriage a more extensive market is opened to every sort of industry than what land carriage alone can afford it, so it is upon the sea-coast, and along the banks of navigable rivers that industry of every kind begins to sub-divide and improve itself, and it is frequently not till a long time after that those improvements extend themselves to the inland part of the country*". Smith (1776). Quoted in Rappaport and Sachs (2003), p. 6. Also Gunnar Myrdal suggested in 1968 that "*…serious study of the problems of underdevelopment [...] should take into account the climate and its impacts on soil, vegetation, animals, humans and physical assets –in short, on living conditions in economic development*". Quoted in Acemoglu (2009), p. 118.

[260] In Acemoglu, Johnson and Robinson (2002), the argument is based on the impact of population density and urbanization in the colonized areas on institutions.

institutions established, ruggedness prevented the development of these low quality institutions and hence, created long-run benefits in Africa.

On the other hand, 'second nature' geography as developed in the New Economic Geography models stresses the spatial interaction between economic agents. In this framework, more theoretically founded, falling trade costs and economies of scale interact in a cumulative process shaping the distribution of economic activity across space. As a consequence of this agglomeration process geographic disparities can initially increase. Nonetheless, these two views regarding geography (first and second nature) are usually considered to be complementary and not opposed to each other. It might perfectly be the case that first nature geography may give a region an initial advantage which then becomes amplified by second nature agglomeration forces (Krugman, 1992, 1993) [261].

New Economic Geography models show that the interaction between transport costs, increasing returns to scale and the size of market under a monopolistic competition framework may lead to the spatial concentration of economic activity[262]. In Krugman (1991), firms tend to locate close to large markets in order to have better access to customers ('*demand* or *backward linkages*') and suppliers ('*cost* or *forward linkages*')[263]. Thus, when transport costs decline, locations with a large market will attract firms and manufacturing activities will increase more than proportionally in such regions. Then, higher nominal wages and a reduction in the local price index as a result of both a greater variety of local goods and a reduction in transport costs will increase real wages. Therefore, if labour is mobile new workers will be attracted to these high market potential locations ('*cost-of-living* or *amenity linkages*')[264]. These agglomeration forces generate a cumulative process which favours the spatial concentration of economic activities and an increase in regional inequality.

[261] This view is also shared by the defenders of first nature geography: "*The two approaches can, of course, be complementary: a city might emerge because of cost advantages arising from differentiated geography but continue to thrive because of agglomeration economies even when the cost advantages have disappeared*". Gallup, Sachs and Mellinger (1999), p. 184.

[262] A more comprehensive survey of NEG models can be found in chapter 2. Here, only the basic ideas that will be used in the model to be presented in section 5.3.a in this chapter are summarized.

[263] This notation differs to the one presented in the previous chapters where *forward linkages* were associated, on the one hand, to the mobility of workers (chapter 2), and on the other, to the firms' suppliers of intermediate goods (chapter 4). However, the terminology used here aims to be homogeneous with the model developed by Ottaviano and Pinelli (2006), which is the basis for the empirical exercise to be carried out in this chapter.

[264] In addition to the spatial distribution of final consumers, an alternative agglomeration force appears when intermediate goods are considered. Again, larger markets would be preferred by firms as they can purchase cheaper intermediate goods and they have a larger demand to sell their output to other producers as intermediate goods. Hence, input-output linkages between firms may also induce agglomeration when labour is not mobile (Venables, 1996).

However, when transport costs leave the 'intermediate level' and they are sufficiently low, dispersion forces may reverse this situation. Wage differentials across locations, congestion costs, fragmentation of firms or personal decisions regarding the decision to migrate act as centrifugal forces leading to a bell-shaped evolution in the concentration of economic activity (Puga, 1999)[265]. Therefore, a further reduction in transport costs (more integration) may eventually lead to a phase characterised by the reduction of regional inequality. As in the case of 'first nature geography' empirical exercises, a positive relationship between market potential and economic growth has been found within a NEG framework, both at an international level (Redding and Venables, 2004; Mayer, 2008) and at a regional scale with wages (Hanson, 2005).

In Chapter 4, evidence in favour of the presence of NEG effects in the location of industrial activities in Spain between 1856 and 1929 has been provided. Making use of a theoretically based approach, the results obtained have shown that market potential was a key element explaining the distribution of industry across Spanish provinces. As the process of industrialisation progressed, industries characterised by larger economies of scale tended to be located in the provinces with a higher market potential. This *scale effect* was the main driving force that shaped the industrial map of Spain from the last decades of the 19th century, when the integration of the Spanish market was completed, until the 1930s. But what happened at a larger level of aggregation when not only industry but the whole economic structure is considered? How much did geography matter for economic growth? What geography did matter, first nature, second nature or both? Can market potential as stated by NEG models explain to some extent the different growth rates in regional income per capita? Did accessibility play a role in the significant increase in regional inequality in Spain from the mid-19th century onwards? Did all these patterns change in the period under study where both divergence and convergence across Spanish provinces is found?

In order to shed light on these questions, an empirical exercise based on Ottaviano and Pinelli (2006) is undertaken in this chapter. Starting from a NEG model, these authors derived an empirical strategy based on the estimation of growth regressions where pure geography elements and market potential were included as explanatory variables of the disparities in regional income per capita. They focused on the evolution of Finnish

[265] The model suggested in Puga (1999) is specially suited for a within country analysis. It combines labour mobility (Krugman, 1991) with input-output linkages (Venables, 1996) and mobility between the two sectors in the economy (agriculture and manufacturing).

regions in the periods 1977-1990 and 1994-2002, with the aim of analysing the changes in regional economic growth in Finland after the collapse of the Soviet block. Here, a similar strategy is followed in order to study whether or not geography had an impact on the regional income per capita growth rates between the mid-19th century and up to the Spanish Civil War (1936-1939). As argued before, the origins of the current regional income disparities in Spain can be found in this period, especially in the remarkable increase in regional income inequality recorded in the second half of the 19th century.

Therefore, following Ottaviano and Pinelli (2006), it is possible to combine the two approaches described above: growth literature and geography. In this sense, growth regressions similar to the ones that can be found in the empirical growth literature to test for the existence of conditional β-convergence can be derived from a NEG model. Then, the impact of geography on economic growth can be assessed by looking at the geography variables included in the regression. Besides, the different concepts relating geographic factors can be considered by taking into account first nature and second nature geography variables[266].

The analysis, as in the previous chapters is focused on Spanish provinces. The analysis covers the years going from 1860 to 1930 and therefore, the period under study is one of particular interest since it corresponds to the upward side of the bell-shaped curve observed in the trend of regional inequality and the period when this tendency was reversed. Finally, the exercise aims to explore regional inequality determinants paying special attention to first and second nature geography variables, which are included as control variables in the growth regressions.

5.2. Related empirical literature: first nature vs. second nature geography

The development in the last decades of the New Economic Geography has opened up new possibilities to analyse the spatial distribution of economic activity and regional disparities in income. Nonetheless, most of the NEG empirical studies have been focused on the industrial sector. The focus on industry is a natural one, since it is in this sector

[266] "*Empirical work should aim to disentangle the forces of differential geography and self-organizing agglomeration economies*". Gallup, Sachs and Mellinger (1999), p. 184.

where economies of scale and increasing returns tend to operate[267]. In addition, analyses from a long term perspective are also relevant, as industrial reallocation processes need time in order to be completed. The process of industrialisation and the integration of national markets began in many cases back in the 19th century and as NEG models show, initial conditions may confer advantages to some locations that later on are reinforced by a cumulative agglomeration process.

At an international level, some studies have tried to disentangle the driving forces behind the location of industry in different historical periods, as seen in the previous chapter. Following the standard model developed by Midelfart-Knarvik, Overman, Redding and Venables (2002) the relative strength of resource endowments as stated by Traditional Trade Theory models and NEG mechanisms has been tested in order to determine the forces shaping the spatial distribution of industry. The evidence collected shows that both type of mechanisms were driving industrial location in interwar Poland (Wolf, 2007), in Victorian Britain (Crafts and Mulatu, 2005, 2006), in the US at the time of the emergence of the manufacturing belt (Klein and Crafts, 2009) and in Spain between 1856 and 1929 (chapter 4).

As regards the Spanish economy, NEG empirical studies from a historical perspective have been profuse[268]. Paluzie, Pons and Tirado (2004) showed the existence of a bell-shaped curve in the evolution of the manufacturing sector in Spain, and recorded the increase in the geographical concentration from the mid-19th century until the 1970s[269]. Rosés (2003), following Davis and Weinstein (1999, 2003) verified the existence of a 'home market effect' in the first stages of the industrialisation process in Spain. Tirado, Pons and Paluzie (2002), based on the proposal of Kim (1995) concluded that NEG factors played a key role in the distribution of industry across Spain in the second half of the 19th century. Moreover, the influence of the NEG forces increased as the integration of the Spanish market progressed. In turn, Betrán (1999) analysed the interwar period, suggesting that the relative industrial rise of provinces like Biscay, Guipúzcoa, Madrid or Zaragoza during this period could be related to the presence of agglomeration economies derived from the size of market.

[267] Services can also show an agglomeration pattern, as shown by Kolko (2007) for the US. In the Spanish case, evidence of strong agglomeration effects in the services sector in the period 1965-1999 were found in Paluzie, Pons and Tirado (2007).

[268] See section 2.2.3 in chapter 2 for a more detailed description of these studies. Here, a brief review of the main contributions is presented, again, to put the debate in context.

[269] A non-monotonic evolution in the manufacturing sector in the long term was also found for the US (Kim, 1995) and France (Combes, Mayer and Thisse, 2008).

In NEG models, the size of a specific location and its access to markets are relevant aspects in the location decisions taken by firms and workers since it may lead to the emergence of *agglomeration* or *dispersion forces*. In this regard, a number of studies tested for the presence of the different agglomeration forces described in the previous section. According to Krugman's (1991) wage equation, relative wages are higher in regions with better access to markets. Following the influential work by Hanson (1998)[270], the existence of a spatial structure in industrial nominal wages in the 1920s was examined by Tirado, Pons and Paluzie (2009). The results verified the existence of a wage gradient centred in Barcelona, the main industrial centre in interwar Spain. On the other hand, Pons, Paluzie, Silvestre and Tirado (2007) established, based on Crozet (2004), a direct relationship between migration decisions and the market potential of the host regions during the 1920s. Martínez-Galarraga, Paluzie, Pons and Tirado (2008) found evidence of an 'agglomeration effect' linking the spatial density of economic activity and the interregional differences in labour productivity in the industrial sector in Spain for the period 1860-1999. Following Ciccone and Hall (1996) and Ciccone (2002), these authors showed that this effect, measured by the estimated elasticity of labour productivity with respect to employment density, was present in the beginning of the industrialisation process in the mid-19th century although its evolution described a decreasing pattern over time[271].

The above studies have offered evidence showing that the forces stressed in the NEG models were present in the first stages of the industrialisation process in Spain. But in what way does geography matter? In the last years an interesting debate at an international level has focused on the relevance of first and second nature geography and their role in explaining the uneven distribution of economic activity across space. In short, first nature geography refers to natural features which are exogenous to the economy such as climate, location or resource endowments. In this sense, Sachs and his co-authors suggest that these pure geography elements had an important effect on income levels, growth rates and population density across countries. Consequently, development is to a large extent determined by physical geography. In contrast, second nature denotes economic man-made geography as suggested by the NEG (Fujita, Krugman and Venables, 1999) which takes into account the location decisions arising from the interaction between economic agents. Nevertheless, as already mentioned, these two

[270] This is the working paper version of Hanson (2005).
[271] A similar analysis has been carried out for France. Combes, Mayer and Thisse (2008).

visions should not be considered mutually exclusive but complementary, although a different role is assigned to economic policy in its effectiveness to correct regional disparities.

As regards the Spanish economic history, some contributions have been made in the terms of this debate. On the one hand, Dobado (2004, 2006) following the line of Sachs argued that the differences in the geographical characteristics across Spanish provinces determined at the end of the 18th century the demographic and economic disparities observed in that period and its persistence over the next two centuries. This author analysed and confirmed the existence of a statistical relationship between provincial economic density both in terms of GDP per square km (for the 20th century) and population density (for the 19-20th century) and a set of pure geography variables. As a result, Dobado (2004, 2006) conferred to pure geography elements a preferential role among the determinants of regional inequality in Spain. The diverse fortunes of Spanish territories would be therefore linked to their geographical conditions.

Pons and Tirado (2008) approached this topic in an attempt to quantify the relative contribution of first and second nature factors to explain the distribution of economic activity throughout the 20th century in Spain. To this aim, they performed an ANOVA analysis based on Roos (2005). With this methodology, the total variance can be decomposed in order to distinguish to what extent such variance is linked first, to pure geography elements; second, to new economic geography factors; and third, to the interaction between first and second nature aspects. Their results showed that pure geography elements explained around 20% of the variance in the relative GDP density between Spanish provinces in 1920. From that moment onwards, the relevance of first nature variables decreased to 6% in 2003. Thus, the relevance of pure geography elements would have weakened over time. The effect of second nature geography was lower in 1920 (10.7%) but increased to a peak of 20% in 1975 and a 14% in 2003. However, the major factor behind economic inter-provincial disparities would be the interaction between both types of variables. Thus, initial differences in terms of first and second nature were amplified by economies of agglomeration[272]. The evolution of

[272] As the authors state, "*...the principal factor behind the regional inequality of relative economic density in Spain is in fact the interaction between both types of variables. This effect explains a growing proportion of the variance throughout the study, rising from a minimum value of 59% in 1920 and a maximum of more than 68% in 2003. The general conclusion that can be drawn from these figures is that inter-provincial economic discrepancies are related to the existence of initial geographical differences of both first and second nature, subsequently amplified by economies of agglomeration in production processes.*

regional income inequality has eventually been shaped by human input into economic activity, and the impact of the initial geographic conditions has decreased over time.

Similarly, Ayuda, Collantes and Pinilla (2010) explored the general patterns in the distribution of population in Spanish provinces and its determinants. Their study went back to the end of the 18th century, when the origins of the present-day population density distribution are to be found. In the pre-industrial period, when agriculture was the predominant activity, first nature advantages determined the distribution of population across Spain since climatic and topographic conditions have a direct impact on the agrarian productivity. As a result, natural conditions provided some locations with an initial advantage. However, the process of industrialisation strengthened this pattern. From 1900 onwards, second nature geography, linked to increasing returns and access to markets, reinforced the cumulative process in the spatial concentration of population as suggested by Krugman (1991) and pure geography elements lost explanatory power as determinants in the spatial concentration of population[273]. To complete the picture, the ANOVA analysis of variance between 1787 and 2000 confirmed the results previously obtained by Pons and Tirado (2008)[274].

Recently, Rosés, Martínez-Galarraga and Tirado (2010) analysed the upswing of regional income inequality in Spain between 1860 and 1930 based on a new regional GDP dataset. The decomposition of the Theil Index showed that Heckscher-Ohlin forces (*between-sector*) were the main driver behind the divergence in regional incomes between 1860 and 1910 in Spain. Therefore, the limited expansion of industry to a small number of regions during the second half of the 19th century increased regional inequality. Then, with the expansion of industry to a larger number of regions in the first decades of the 20th century a convergent pattern in regional specialisation was found, although NEG forces (*within-sector*) were strengthened after 1910.

In a good number of the empirical studies surveyed in this section, the focus on industry has been predominant. In short, the evidence gathered in these studies shows that for the Spanish case, the presence of increasing returns and access to markets played an

Furthermore, the net effect of factors referred to as 'Krugman' geography has increased throughout the 20th Century". Pons and Tirado (2006), p. 17.

[273] Studies on the evolution of the Spanish population and urban system in the long term can be found in Lanaspa, Pueyo and Sanz (2003), Goerlich and Mas (2009).

[274] "*From the beginning of the XX century onwards, the interaction between the two types of variables is the main factor, although it exceeds the first nature effects by very little. Lastly, in the final period, 1950-2000, while first nature effects continue to lose their relative importance, their impact via those of second effect now reaches 49% of explanatory power*". Ayuda, Collantes and Pinilla (2010), p. 43

important role in determining the geography of industry in Spain from the second half of the 19th century onwards when the industrialisation process began and the integration of the domestic market was completed. Market potential was also a relevant variable for the spatial distribution of nominal wages and migration decisions in the 1920s in Spain, showing the existence of the agglomeration forces described by Krugman (1991).

As for the debate between first and second nature geography, the studies have focused on the impact of these two concepts regarding geography on economic density over the 20th century and the concentration of population in the last two centuries. In some cases, a key role for first nature geography has been found (Dobado, 2004, 2006). Other studies have concluded that the increasing impact of second nature geography or agglomeration forces should be added to explain the discrepancies in regional income inequality and population trends. As regards the spatial concentration of population in Spain, second nature or NEG effects became relevant only after 1900 when industry had spread sufficiently across Spain in parallel with the decline of agricultural population (Ayuda, Collantes and Pinilla, 2010). In addition, over the 20th century, cumulative processes in the line of NEG arguments were strengthened as drivers of the differences in the economic density of Spanish provinces (Pons and Tirado, 2008).

In the next pages, the aim is to analyse the impact of geography on economic growth considering not only the industrial sector but provincial GDP per capita. Within the NEG framework, empirical studies on regional economic growth are yet scant in contrast with some well-known cross-country studies (Redding and Venables, 2004; Mayer, 2008). In this context, the estimation of growth regressions allows exploring the causes of the differences in the growth rates across provinces and the focus will be on the impact of geography. The empirical strategy adopted follows Ottaviano and Pinelli (2006) and thus, in the next section, the NEG model used by these authors to derive the growth regressions is detailed.

5.3. Empirical strategy

5.3.1. The model

This section reproduces Ottaviano and Pinelli (2006). Their NEG model is obtained by extending the set-up of Redding and Venables (2004) by introducing labour mobility and land *à la Hanson* (1998) and Helpman (1998). The economy consists of *i=1,…, R* regions. On the demand side, in region *j*, the representative worker consumes a set of horizontally differentiated varieties and land services (‘*housing*’). Her utility function is:

$$U_j = (X_j)^{\mu}(L_j)^{1-\mu}\text{, } 0 < \mu < 1$$

where L_j is land consumption and

$$X_j = \sum_{i=1}^{R}\left\{\int_0^{n_i}\left[x_{ij}(z)\right]^{\frac{\sigma-1}{\sigma}}dz\right\}^{\frac{\sigma}{\sigma-1}} = \sum_{i=1}^{R}\left(n_i x_{ij}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}}$$

is a CES quantity index of the $\sum_{i=1}^{R} n_i$ varieties available in region *j* with x_{ij} labelling the consumption in region *j* of a typical variety produced in region *i*. The associated exact CES price index is:

$$P_j = \sum_{i=1}^{R}\left\{\int_0^{n_i}\left[p_{ij}(z)\right]^{1-\sigma}dz\right\}^{\frac{1}{1-\sigma}} = \sum_{i=1}^{R}\left(n_i p_{ij}^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$$

where p_{ij} is the delivered price in region *j* of a typical variety produced in region *i*. In the above expressions, the second equality exploits the fact that in equilibrium quantities and prices are the same for all varieties produced in country *i* and consumed by country *j*. Utility maximization gives the demand in *j* for a typical variety produced in *i*:

$$x_{ij} = p_{ij}^{-\sigma} E_j P_j^{\sigma-1} \qquad (5.1)$$

where E_j is expenditures on X_j, which is a fraction μ of income I_j, while σ>1 is both the own and the cross price elasticity of demand.

On the supply side, each variety is produced by one and only one firm under increasing returns to scale and monopolistic competition. In so doing, the firm employs labour, land and, as intermediate input, the same bundle of differentiated varieties that workers demand for consumption. Specifically, in region *i*, the total production cost of a typical variety is:

$$TC_i = P_i^{\alpha} r_i^{\beta} w_i^{\gamma} c_i (F + x_i)\text{, α, β, γ >0, α + β + γ = 1}$$

where x_i is total output, r_i and w_i are land rent and wage, while c_i and c_iF are marginal and fixed input requirements respectively[275]. Trade faces iceberg frictions: for one unit of any variety to reach destination when shipped from region *i* to region *j*, $\tau_{ij}>1$ units have to be shipped. Hence, $x_i = \sum_{j=1}^{R} x_{ij}\tau_{ij}$.

Firm profit maximization yields the standard CES mark-up pricing rule:

$$p_i = \frac{\sigma}{\sigma - 1} P_i^{\alpha} r_i^{\beta} w_i^{\gamma} c_i \text{, } p_{ij} = \tau_{ij} p_i \tag{5.2}$$

Free entry then implies that in equilibrium firms are just able to break even, which happens when they operate at scale $\bar{x} = (\sigma - 1)F$. Together with (1) and (2), that allows writing the free entry condition in region *i* as:

$$(FE) \bar{x} \left(\frac{\sigma}{\sigma - 1} r_i^{\beta} w_i^{\gamma} c_i \right)^{\sigma} = MA_i SA_i^{\frac{\alpha\sigma}{\sigma - 1}}$$

where $MA_i = \sum_{j=1}^{R} \tau_{ij}^{1-\sigma} E_j P_j^{\sigma-1}$ is the 'market access' of region *i*. This is a measure of customer competitor proximity ('*demand linkages*') that predicts the quantity a firm sells given its production costs. The term $SA_i = P_i^{1-\sigma} = \sum_{j=1}^{R} n_j p_j^{1-\sigma} \tau_{ji}^{1-\sigma}$ is, instead, the 'supplier access' of region *i*, a measure of supplier proximity. This inversely predicts the prices a firm pays for its intermediate inputs ('*cost linkages*') and a worker pays for her consumption bundle ('*cost-of-living linkages*') when located in a certain region.

Workers work and consume in the region where they reside and can pick their residence freely. This implies that in equilibrium they are indifferent about location as

[275] In the cross-country study by Redding and Venables (2004), the parameter c_i is allowed to vary to capture Ricardian productivity advantages across countries. This interpretation is hard to defend within the same country, so its variation across regions will be interpreted as the outcome of localized technological externalities. These will be introduced as controls in the empirical analysis.

they would achieve the same level of indirect utility *V* wherever located. Given the chosen utility, if it is further assumed that the land of a region is owned by locally resident landlords, free mobility then gives[276]:

$$(FM)\frac{w_i}{SA_i^{\frac{\mu}{1-\sigma}} r_i^{1-\mu}} = V$$

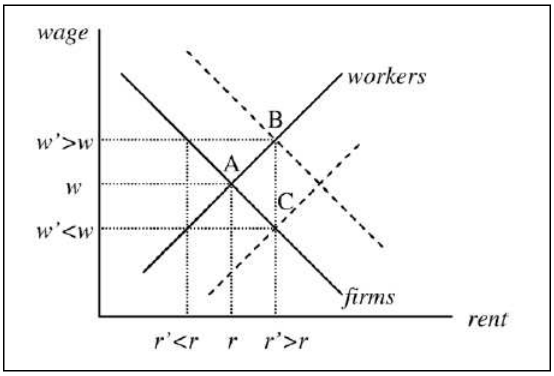
After log-linearization, conditions (*FE*) and (*FM*) are depicted in Fig. 1, which measures the logarithm of regional nominal wages (*w*) along the vertical axis and the logarithm of regional land rents (*r*) along the horizontal one. Downward sloping lines are derived from (*FE*) and depict the combinations of wages and rents that make firms indifferent about regions. Their downward slope reflects the fact that firms can break even in different regions provided that higher wages correspond to lower rents and vice versa. Upward sloping lines are derived from (*FM*) and depict the combinations of wages and rents that make workers indifferent about regions. Their upward slope reflects the fact that workers can achieve the same utility (‘*real wage*’) in different regions provided that higher rents correspond to higher wages and vice versa.

The exact positions of the two lines depend on regional market access and supplier access. Better market access (larger *MA*) shifts *FE* up, increasing both wages and land rents. Better supplier access (larger *SA*) shifts both *FE* and *FM* up, also increasing rents. The effect on wages is, instead, ambiguous: they increase (decrease) if the shift in *FE* dominates (is dominated by) the shift in *FM*. This theoretical ambiguity makes it pointless to try to disentangle the effects of *MA* and *SA* on equilibrium wages and rents. What we can do, instead, is to check whether their combined effect is indeed positive on rents as predicted by the model. In addition, we can use information about migration flows. Since land values capitalise the attractiveness of a place, land rents rise also because immigration increases the demand for land.

More interestingly, we can also check whether the combined effect of *MA* and *SA* is positive or negative on wages, which would point at a dominant impact on firms (point B) or on workers (point C), respectively. ‘*Demand linkages*’ and ‘*cost linkages*’ would dominate in the former case, ‘*cost-of-living linkages*’ in the latter.

276 This assumption is made only for analytical convenience. What is crucial for what follows is that the rental income of workers, if any, is independent of locations and, thus, it does not affect the migration choice. The alternative assumptions of absentee landlords or balanced ownership of land across all cities would also serve that purpose.

Figure 5.2. The geographical equilibrium



Source: Ottaviano and Pinelli (2006), p. 640.

Growth regressions

The discussion in the previous section suggests identifying the combined effects of *MA* and *SA* on productivity and amenity through their impacts on the levels of wages, rents and migration flows using panel techniques. However, in the exercise to be carried out for the Spanish case, only the combined effects of *MA* and *SA* on the level of wages are going to be studied. Under the assumption that regions have been fluctuating around a balanced growth path (BGP) during the observed period, the panel estimation of those impacts can be interpreted as their long-run effects along the BGP. This interpretation allows using growth regressions instead of panel regressions with a double advantage. First, endogeneity would potentially affect the panel estimates since higher productivity and amenity could be the causes rather than the effects of better market and supplier access. For example, if booming regions attracted firm's land workers, then the positive correlation between access and immigration could arise due to reverse causation from the latter to the former. Second, the focus on levels would obscure the dynamic evolution of productivity patterns across regions, which is an interesting issue in itself as NEG stresses the possibility of cumulative agglomeration.

Both issues can be dealt with by estimating standard growth regressions over a set of explanatory variables including some measure of market and supplier access:

$$\ln(w_t) - \ln(w_{t-1}) = \alpha + \beta \ln(w_{t-1}) + \gamma \ln(access_{t-1}) + \delta \ln(controls_{t-1}) + \varepsilon_t \qquad (5.3)$$

where the growth rate of regional wages on the left hand side is regressed on its initial value and other 'initial conditions' including some measure of market and supplier access[277].

The idea is that along a BGP productivity grows at a constant rate across regions so that these may differ only in terms of wage levels. Then, under the assumption that the economy fluctuates around its BGP, the growth equation captures transitional growth: if a certain province exhibits a higher growth rate than the other, then the former has a higher level of wage in BGP than the latter and it is converging to that level, given its initial conditions. As anticipated, while modelling the dynamics of the economy, the above equation also allows to partially tackling the endogeneity problem. The reason is that, whereas market and supplier access is measured at the beginning of period (at time $t-1$), the growth of wage is measured during the period of observation (from times $t-1$ to t). In other words, the independent variables are predetermined relative to the dependent one.

5.3.2. Data

As seen in previous sections, the study of the forces that have driven the regional economic performance in Spain is carried out through the estimation of standard growth regressions. The analysis is based on the Spanish provinces[278] and the period under study, which goes from 1860 to 1930 is, in turn, divided into two different periods[279]. Thus, the impact of geography variables on economic growth can be analysed for the second half of the 19th century (1860-1900) and the first decades of the 20th century (1900-1930) prior to the break out of the Spanish Civil War.

The growth equation derived from the NEG theoretical model by Ottaviano and Pinelli (2006) is used to explain differences in the provincial economic growth rates. The equation includes a set of explanatory variables, which is usually divided in the growth literature into two alternative groups[280]: first, the *proximate sources of growth* include those variables related with the production factors that directly affect regional

[277] To fully exploit the model and disentangle productivity from amenity effects, the above equation has to be matched by similar regressions for land values and migration flows.

[278] Now, the Basque Country provinces' and Navarre are included in the sample. Nevertheless, due to their geographic peculiarities, the Balearic and the Canary Islands, as well as the autonomous cities of Ceuta and Melilla are excluded. As a result, the study includes 47 Spanish provinces.

[279] First, the period has been divided into two subperiods of a similar size; second, this division is based on the divergence/convergence observed patterns; third, the division takes into account the change in the trade policy implemented by the Spanish governments. The protectionist turn in the 1890s (Canovas' 1891 tariff) was expanded and consolidated in the first decades of the 20th century, with the Salvador Tariff (1906) and the Cambó Tariff (1922).

[280] See, e.g. Temple (1999).

performance. In the context of the Solow growth model income disparities are explained by differences in technology, physical capital and human capital. Second, the *wider influences* take account of other variables that might have an indirect impact on regional performance, including policy, geography or institutions. However, in contrast with cross-country studies, regional analyses do not usually include the last set of variables since regions within the same boundaries usually share the same institutional framework.

Performance measures. Regional economic performance is captured by the growth in the income per capita, measured by GDP per capita. Therefore, this variable is considered to proxy wages in equation (5.3). Data on Spanish GDP at a NUTS3 level of aggregation have been constructed in chapter 3, where new estimates of provincial GDP were calculated applying the methodology proposed by Geary and Stark (2002). On the other hand, data on each province population are collected from the Population Censuses of 1860, 1900 and 1930, respectively. Then, per capita GDP growth rates are calculated using a simple log growth rate.

Explanatory variables. Two sets of explanatory variables that have traditionally been considered by the growth literature are included in the regressions: *proximate sources of growth* and *wider influences*. One of the most interesting contributions of the empirical strategy developed by Ottaviano and Pinelli (2006) is that the wider influences for growth can be enlarged in order to include geography variables and therefore, assess their impact on regional economic growth.

Proximate sources of growth

a) Physical capital. The regressions include the initial level of GDP per capita to control for decreasing returns to capital accumulation. Moreover, the initial level of GDP per capita is also going to provide relevant information about the existence of conditional convergence.

b) The stock of human capital in each province is proxied by the literacy rates. In this case, data is taken from Núñez (1992).

c) Knowledge capital. The stock of knowledge capital is measured by the number of patents per capita. Unfortunately, data on the number of patents registered is only available at a NUTS2 level of aggregation (Autonomous Communities). In this case, the information comes from Sáiz (2005). Therefore, NUTS2 data have been applied to the provinces within each NUTS2 region. Population figures are once again collected from the Population Censuses.

Wider influences

d) Policies. The provision of infrastructures is one of the most important policies undertaken by governments (Aschauer, 1989). In order to capture the provincial availability of infrastructures, two alternative measures are used: the total stock of infrastructures and the infrastructure density of the Spanish provinces[281]. This information can be found in Herranz (2008).

e) First Nature Geography. First nature variables stress the relevance of location and climate for economic development. Spain shows profound interregional differences in terms of climatic conditions and the orography. In this case, alternative variables have been compiled in order to capture the effects of pure geography:

- *Location*. The importance of the geographical location is measured in two ways. First, the variable *altitude* is considered. This variable refers to the altitude of the provincial capital above sea level in meters. The altitude of provincial capitals has been selected since it captures the altitude of the population settlement[282]. This information can be found in the Spanish Statistical Yearbook of 1930 (*Anuario Estadístico de España - AEE*). Differences in altitude may potentially have a significant effect on agriculture, on transport costs, and on population settlement, especially in a mountainous country such as Spain[283]. Therefore, a negative sign for the coefficient associated to altitude is expected. When looking at the regional patterns, huge differences are found in the average altitude of Spanish provincial capitals ranging from Alicante (3 m above sea level) to Ávila, the highest provincial capital placed 1131 m above the sea level, and a total of 23 provincial capitals are above 400 m. The geographical patterns of the variable *altitude*, as well as other first nature variables comprised in the exercise, are displayed in maps 5.1.

The second variable considered is the coastal or inland position of the different provinces (*coast*). Spain, located in the Iberian Peninsula, is a costal country. The Spanish

[281] Infrastructures density is measured as the provincial stock of infrastructures per square km. Data on the provinces' area comes from www.ine.es.

[282] "*This choice stresses the agricultural potential of the core economic and demographic areas of our provinces and, since agronomic work has found that altitude is the crucial factor in diminishing agricultural yields in upland environments, it penalises those provinces whose core areas had weak agricultural potential*". Ayuda, Collantes and Pinilla (2010), p. 33.

[283] "*According to IGN data, 39.3% of Spain's land area lies between 600 and 1000 metres above sea level, and 18.5% is above that height*". Goerlich and Mas (2008), p. 7.

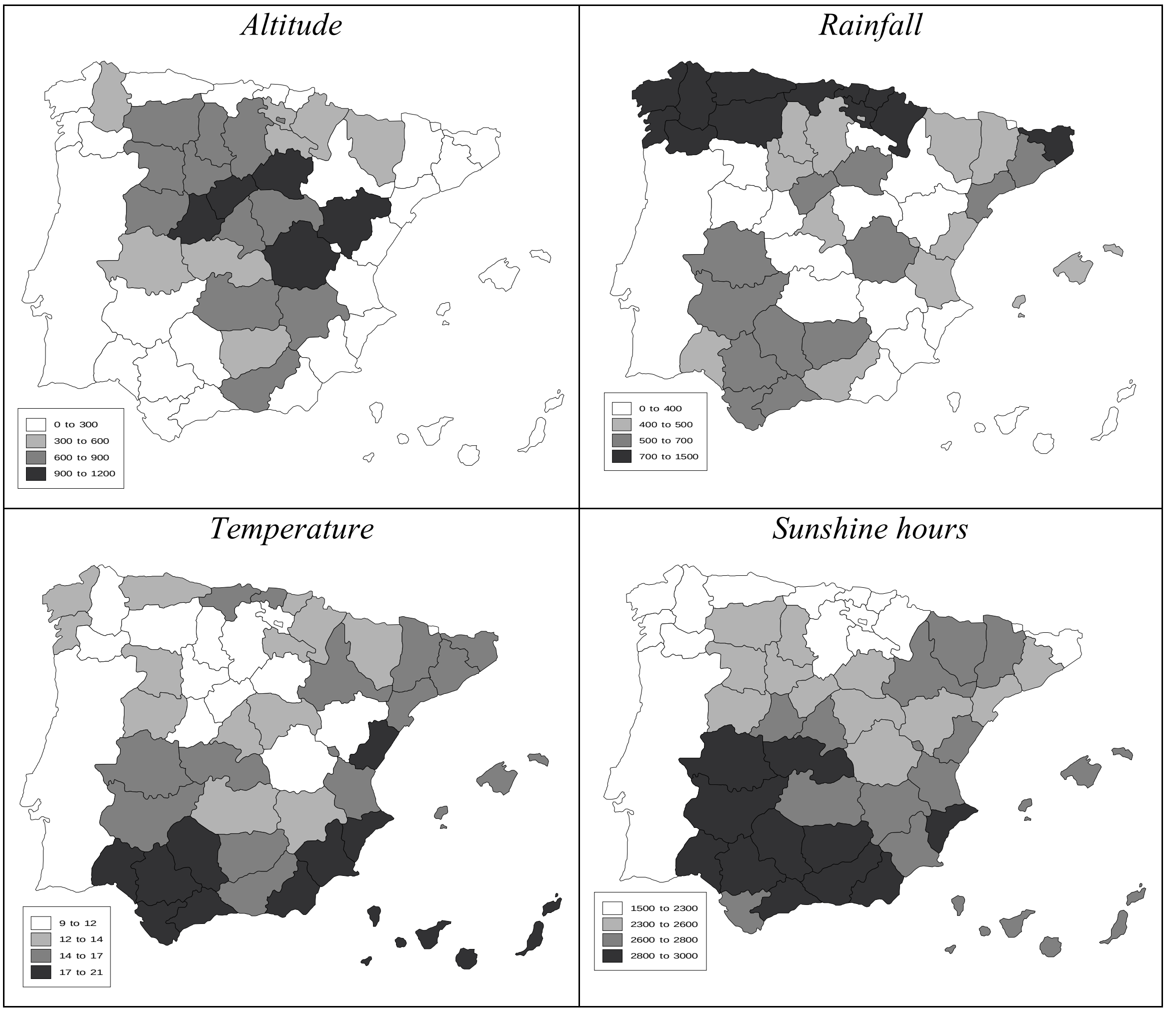
peninsular coastline has a total length of 4865 kilometres[284]. Moreover, 19 provinces out of the 47 continental provinces included have direct sea access. In this case, a dummy variable taking the value of 1 has been assigned to those provinces with direct access to sea. As earlier mentioned, it is expected that coastal provinces will have to face lower transport costs for trade, and therefore, a positive sign for this variable is expected.

- *Climate*. The impact of climatic conditions in Spain, a southern European country, is studied through three alternative measures. First, the variable *rainfall* is computed as the annual average precipitation in mm in the provincial capitals during the period 1901-1930. Data come from the *Spanish Statistical Yearbook* of 1960. The highest rainfall areas are located in the northern coast where average rainfall is above 700 mm per year. Below this area, in the inland provinces and along the Mediterranean coast rainfall is scarce. The driest province is Almería, in the south-east of Spain, with an average of 219 mm per year. By contrast, Pontevedra, in the north-west, is the province with a higher annual rainfall average of 1455 mm. A positive relationship between rainfall and agricultural productivity is expected, especially in a country such as Spain, characterised by a dry climate and a high proportion of the agricultural land devoted to dry-farmed crops like cereals.

The second variable, *temperature*, is measured through the annual average temperature in ºC for the period 1901-1930. The source is again the *Spanish Statistical Yearbook* of 1960. Finally, data on the annual number of sunshine hours for a more recent period (1971-2000) have also been collected to construct the third variable: *sunshine*[285]. In this case, Almería, the driest province, is also the one with a higher average temperature (19.8ºC per year). The coldest province is León with an average temperature slightly below 10ºC. As regards sunshine hours, Huelva, with almost 3000 hours of sunshine per year almost doubles the province at the other extreme, Vizcaya (1554 hours/year). These climatic conditions have an influence on agrarian activities and in the case of the last two (*temperature* and *sunshine*) a negative sign is expected for the coefficients, while a positive sign is expected for *rainfall*.

[284] If the insular territories (Balearic and Canary Islands) and the Autonomous cities in Northern Africa, which are excluded of our exercise, are considered, the total coastal length of Spain adds up to 7905 km.
[285] www.aemet.es/es/elclima/datosclimatologicos/valoresclimatologicos.

Maps 5.1. 'First nature' variables



Sources: see text. Altitude: altitude of the provincial capital above sea level (m); Rainfall: annual average rainfall in the provincial capitals, 1901-1930 (mm); Temperature: annual average temperature, 1901-1930 (ºC); Sunshine hours: annual number of sunshine hours, 1971-2000. See table A8 in the Appendix.

f) Second Nature Geography. According to the NEG models, the interaction of increasing returns and transport costs, can make firms and workers to be attracted to high market potential areas in a process of cumulative causation that reinforces the spatial concentration of activity and therefore, regional income disparities. At an international level, Redding and Venables (2004) used data on bilateral trade flows to estimate both market access (*MA*) and supplier access (*SA*). The lack of this information for the Spanish regions in the period under study does not allow exploiting this approach. However, as Ottaviano and Pinelli (2006) stated, the separate effects of *MA* and *SA* cannot be disentangled with labour mobility. For that reason they used a joint measure of market and supplier access which corresponds to the Harris' (1954) market potential equation, defined as:

$$MP_r = \sum_s \frac{M_s}{d_{rs}} \qquad (5.4)$$

where MP_r is a measure of the size of province *r* (usually GDP) and d_{rs} is the distance, or as in this case, bilateral transport costs between *r* and *s*. On the basis of the Harris' (1954) expression, market potential estimates for the Spanish provinces have been constructed in chapter 3 following the work by Crafts (2005b).

5.4. Estimation results

Table 5.1 reports the estimation results of the growth regressions for the two subperiods analysed: 1860-1900 and 1900-1930. Considering the whole information described in the previous section, only the benchmark regressions are presented. These regressions are selected on the basis of the explanatory power and the significance of the variables considered. Equation (5.3) is estimated using OLS corrected for heteroskedasticity using White's method. An important component of the Harris market potential measure is the contribution to the potential of region *r* of its own GDP, also known as self-potential. Therefore, by construction, the explanatory variable market potential and the dependent variable (GDP per capita growth) influence each other at the same time. In order to avoid simultaneity problems market potential has alternatively been calculated purging the self-potential (columns 2 and 4)[286].

<u>1860-1900</u>

Growth regressions for the second half of the 19th century are depicted in columns (1) and (2) in table 5.1. First, in terms of the R^2, the goodness of fit is acceptable: over 46% of the variation in provincial per capita income growth rates is explained. In addition, the exclusion of the self-potential in the market potential calculations does not

[286] An interesting expansion of the exercise could be to test for spatial correlation *à la Anselin* to examine whether or not spatial externalities beyond the limits of the provincial boundaries exist. The presence of such externalities would confirm that income per capita growth rate of the neighbouring provinces have an effect on local growth rates.

alter the results obtained. The signs of the coefficients and their significance, and the explanatory power of the regressions remain unchanged.

Table 5.1. Growth regressions. Dependent variable: GDP per capita growth

years→	1860-1900		1900-1930	
explanatory variables ↓	(1)	(2)	(3)	(4)
Constant	7.609*** (4.77)	7.996*** (4.80)	8.080*** (5.43)	8.025*** (5.11)
GDP per capita	-0.372** (-2.40)	-0.407** (-2.66)	-0.764*** (10.58)	-0.670*** (-8.10)
Literacy	0.158 (1.54)	0.159 (1.53)	-0.090 (-0.81)	-0.107 (-0.87)
Patents per capita	0.101*** (3.24)	0.099*** (3.39)	0.106*** (4.68)	0.114*** (4.46)
Infrastructures	0.025 (0.46)	0.017 (0.35)	0.149** (2.37)	0.188*** (2.86)
Altitude	-0.063* (-1.86)	-0.068* (-2.01)		
Temperature			-0.157 (-0.82)	-0.205 (-0.94)
Sunshine	-0.551*** (-2.88)	-0.569*** (-2.87)	-0.457*** (-2.99)	-0.439** (-2.63)
Coast			-0.188** (-2.51)	-0.134* (-1.77)
Market potential	-0.121 (-1.29)		0.278*** (3.79)	
Market potential (self-potential excluded)		-0.128 (-1.55)		0.187*** (2.88)
Number of observations	47	47	47	47
R-squared	46.4	47.5	63.4	60.1

All explanatory variables are in log terms; t-statistics are in parentheses (based on robust standard errors).
* Significant at 10%; ** significant at 5%; *** significant at 1%.

The negative and significant coefficient on the initial level of income per capita reveals the existence of conditional β-convergence among the Spanish provinces' in this period. However, when the initial level of income per capita is included alone in the regression its sign is positive although not significant. Therefore, only after controlling for other regional specific variables this coefficient turns negative and significant which is indicative of conditional β-convergence. As to other proximate sources of growth, evidence of a positive relationship between the stock of knowledge capital and per capita GDP growth is found. The sign of the coefficient on the number of patents is positive and highly significant. There is, instead, no evidence of a significant effect of the human capital stock.

As regards the geography variables considered, it has to be stressed first the fact that the inclusion of such variables in a standard growth regression equation does not modify the expected sign of the other variables included. More importantly, columns 1 and 2 confirm the relevance of first nature geography in explaining growth differentials across the Spanish provinces in the second half of the 19th century. The coefficient associated to the variable sunshine is highly significant and shows the expected negative sign, meaning that climatic conditions were important: provinces with a higher number of annual sunshine hours experienced lower per capita income growth rates. The geographical location is also important: the higher the altitude, the lower the growth rate of income per capita. A negative sign for the coefficient associated to the altitude of the provincial capital is found[287]. Finally, the results show that NEG-related effects measured by the market potential at the beginning of the period did not have a significant impact on the provincial income per capita growth rates in this period[288].

<u>1900-1930</u>

The results for the second period are shown in columns (3) and (4) in table 5.1. Again, growth regressions are estimated on the basis of two alternative measures for the market potential. Some differences appear in the fit of the model when compared to the previous period: the explanatory power in terms of the R^2 has increased. Now, the regression explains 63.4% of the provincial per capita income growth. However, the goodness of the fit decreases when the self-potential is purged from the market potential measure and the value for the R^2 is a bit lower (60.1%).

The coefficient on the initial per capita income is negative and highly significant. The significant negative relationship between the initial income levels and subsequent growth implies a conditional β-convergence pattern in the first decades of the 20th century. Furthermore, when this variable is included alone in the regression, the coefficient remains negative and significant which points to the existence of unconditional β-convergence in this period. However, decreasing returns to capital accumulation were not the only force at work.

[287] In this case, the variable is significant at 10%.
[288] Besides, market potential shows a negative sign on its coefficient.

The number of patents per capita remains significant and the magnitude of the coefficient is almost maintained. Hence, a positive effect of the stock of knowledge capital on provincial per capita income growth is found. Yet, human capital is statistically insignificant as in the preceding period. Conversely, the coefficient on the total stock of infrastructures becomes significant in these years. Thus, a positive relationship is established between the stock of infrastructures in 1900, when all the provincial capitals were finally connected to the railway network, and the subsequent GDP per capita growth rate in Spanish provinces. This result confirms that the infrastructure endowment had a positive impact on regional growth as economic historians have often emphasised (Herranz, 2007b).

As regards geography, two 'first nature' variables are found to be relevant: sunshine and coast. Firstly, the coefficient on the annual number of sunshine hours is again significant and negative, as expected. Nonetheless, in this case the slight reduction in the magnitude of this coefficient reveals a weakening tendency in the impact of weather conditions. Secondly, geographical location is now represented by the coastal situation of the Spanish provinces. A positive relationship between coastal location and GDP per capita growth is expected since provinces placed nearby the sea have an advantage in terms of lower transport costs, and it is very likely that they also have a lower altitude and a more temperate climate (Rappaport and Sachs, 2003). However, the coefficient on the coastal provinces is significant but negative. This result is counterintuitive. Some possible explanations to this outcome can be suggested as hypotheses. From a methodological point of view, the coast variable is implicitly taking into account distances to other markets. Since market potential is by definition also considering distances, the coastal statistical effect could be partially captured by the market potential variable. In fact, when market potential is excluded from the equation, the variable coast shows a positive sign, as expected, although the coefficient is not statistically significant. There might be an economic reason as well. In this case, the explanation could be linked to the change in the trade policy implemented by the Spanish governments after the Canovas Tariff was established in 1891. As Spain was progressively becoming a more closed economy the domestic market also became more economically relevant for the Spanish provinces[289]. In such context, coastal provinces had

[289] For the debate on the impact of trade policy in the internal geography of countries: Krugman and Livas Elizondo (1996), Monfort and Nicolini (2000), Paluzie (2001), Alonso-Villar (2001), or Crozet and Koenig (2004a). See section 2.1.3 in chapter 2.

to face relatively longer distances to reach interior markets. The geographical characteristics of Spain therefore conferred an advantage to the inland provinces especially in this protectionist period since the beginning of the 20th century when all the provincial capitals were connected to the railway network.

More interestingly, in the period 1900-1930 there is evidence in favour of NEG-related effects. The coefficients for the market potential in columns 3 and 4 in table 5.1 are positive, as expected, and highly significant. Hence, provinces with a higher market potential in 1900 also experienced higher income per capita growth rates in the subsequent decades. Thus, the role of agglomeration economies in explaining income per capita growth differentials among the Spanish provinces is confirmed for this period. The point estimate of the market potential in table 5.1 implies that a 10 percent increase in a province's market potential would increase her GDP per capita growth rate by 2.8 percent. Finally, even when the self-potential of the provinces is excluded from the accessibility measure there is evidence that the market potential of neighbouring provinces had a positive effect on the economic growth rate although in this case the effect is lower (a bit below 2 per cent).

5.5. Discussion

As suggested by Rosés, Martínez-Galarraga and Tirado (2010), Heckscher-Ohlin forces were the main driver behind the divergence in regional incomes in the second half of the 19th century. Regional specialisation and structural change, in line with the predictions of Williamson (1965), can explain the increase observed in regional inequality. In this regard, no evidence that NEG-type mechanisms were in operation in the industrial sector in mid-19th century was found in chapter 4. However, as the process of industrialisation arrived in a larger number of regions in the first decades of the 20th century, the convergence in the regional economic structures favoured the reduction of regional income inequality, although NEG forces were strengthened as the industrial sector was increasing its weight in the Spanish economy. The results obtained in this chapter can shed light on this issue and can also help to complete the picture described by Rosés, Martínez-Galarraga and Tirado (2010). In what follows, the results are examined paying special attention to the geography variables included in the growth regressions. Overall, the results indicate that geography matters in explaining regional income asymmetries in Spain in the first stages of economic development.

In the second half of the 19th century the impact of geography on the provincial GDP per capita growth rates came from first nature geography elements. First, the provinces whose capital, the main population settlement, was characterised by a higher altitude, experienced lower growth rates. On the one hand, this is reflecting the influence of geographic conditions on transport costs. The ruggedness of land makes communications more difficult and expensive and thus, altitude becomes a clear economic disadvantage for these locations. In the case of Spain, in mid-19th century the basic railway network was under construction but still at an infant stage and road transport, the main inland alternative traditional transport, was relatively more costly and time-consuming[290]. Moreover, some of the provincial capitals with a higher altitude above the sea level were among the last ones to be connected to the railway network[291]. These mountainous provincial capitals are often located at a relatively longer distance from the coast, making it more difficult to have access to coastal shipping or to reach foreign markets. Second, altitude also had an impact on agriculture. The work on rough terrain is more complicated, land may be less suited for the cultivation of crops, the cost of irrigation may increase, and it is more likely that high altitude locations have to face more severe climatic conditions, being lower yields for agrarian activities the outcome of all these difficulties.

The influence of climatic conditions on agriculture is even more evident through the annual number of sunshine hours, a highly significant variable in this period. In the mid-19th century, the agricultural sector was predominant in the Spanish economy. At an aggregate level, the workers enrolled in agrarian activities represented around two-thirds of the total Spanish active population[292]. Therefore, through its direct impact on agrarian productivity and transport costs, first nature variables stand as the main geographic forces driving GDP per capita growth differentials within Spain in the second half of the 19th century.

The predominance of the agricultural sector may also help to explain the lack of second nature geography effects in mid-19th century, i.e., the fact that the market

[290] The problems that road transport and inland navigation had to face in mid-19th century Spain have been described in the first chapter.

[291] That would be the case of Segovia (1884) and Cuenca (1885), and especially Soria (1892) and Teruel (1901). There are exceptions, however. Ávila, the highest altitude provincial capital in Spain enjoyed an early connection to the rail network. See map 3.1 in chapter 3, where the expansion of the railway network is shown, and Wais (1987).

[292] In terms of the GVA generated in the agrarian sector, its contribution to total Spain's GDP was 39.5% in 1860 and 20.9% in 1900. Prados de la Escosura (2003).

potential of the Spanish provinces in the 1860s did not have an influence on subsequent economic growth. The spatial pattern of market potential at that time, as shown in chapter 3, was characterised by a more equal distribution when compared to the following period. In addition, although in the 1860s industrialisation had already started in some regions like Catalonia, the process in Spain was still at its first stages and most of the country remained basically agrarian, and thus, the presence of increasing returns was very limited. Moreover, industry was mainly oriented towards the production of consumption goods where economies of scale are less important. Besides, as already mentioned, the railway network was starting to be constructed and market integration was not consolidated yet. This outcome reinforces the results obtained in the previous chapter, where no evidence of NEG effects in industry in 1856 was found and land endowment for agriculture was the main driving force determining industrial location.

Yet, the effect of first nature geography is still present in the second period studied, from 1900 to 1930. In this case, the impact comes again from the annual number of sunshine hours, a variable related to the dryness of the climate and therefore linked to the disadvantages for the attainment of high levels of productivity in the agrarian activities. Hence, first nature elements had an influence in explaining income per capita growth rate differences as suggested by Dobado (2004, 2006).

More prominently, in the beginning of the 20th century NEG forces had a positive influence on the subsequent provincial growth differentials. During the second half of the 19th century, there were fundamental changes in the geographical pattern of provincial market accessibility. As argued in chapter 3, coastal provinces improved its relative market potential when compared to inland provinces (with the exception of Madrid) and in 1900 a polarised structure in the spatial distribution of market potential had been shaped. The results indicate that market accessibility in 1900 had a positive impact on provincial GDP per capita growth rates in the next three decades. In a NEG framework, the emergence of agglomeration forces is the outcome of the interaction between transport costs and increasing returns to scale. On the one hand, it was during the second half of the 19th century that the railway network was completed, transport costs experienced a marked fall and market integration progressed decisively. On the other, in the same period, the process of industrialisation expanded to a larger number of provinces, and thus, the presence of increasing returns was amplified. The first decades of the 20th century witnessed the development of new industrial centres in inland provinces like Madrid and Zaragoza[293]. Likewise, activities where economies of scale were more prominent achieved a higher development within the industrial sector. In this regard, the presence of agglomeration forces at an aggregate income level is related with the increasing weight of the industrial sector within the Spanish economy[294]. After 1900, when NEG-type mechanisms were already in operation in the industrial sector (chapter 4), agglomeration effects also had an impact on the economic performance of Spanish provinces in terms of the GDP per capita growth rates.

This result is very close to the one obtained by Ayuda, Collantes and Pinilla (2010) in their analysis of the determinants of the population patterns in Spain at a regional level. These authors found that from 1900 onwards the combined effect of first and second nature elements reinforced the ongoing concentration of the Spanish population. Similarly, Pons and Tirado (2008) argued that the interaction between first nature and second nature was a key element in the explanation of the differences in the relative provincial density of GDP in Spain in 1920 and onwards. In our case, in a period of regional convergence, the emergence of NEG effects at a regional income level pushing in favour of an increase in regional inequality was more than offset by the effect of diminishing returns to capital and the convergent pattern in the economic structures across provinces as the industrialisation process reached a larger number of provinces. Nevertheless, geography variables are also behind the evolution of provincial income per capita growth rates. Thus, the geographic conditions of the Spanish provinces and the presence of NEG effects need to be borne in mind in the analysis of the forces driving regional inequality in the long term.

[293] The emergence of these new industrial centres has been linked with the protectionist policy implemented by the Spanish governments after the 1890s (Tirado, Pons and Paluzie, 2006, 2009).

[294] The industrial GVA as a percentage of total GDP in Spain increased from 15.5% in 1860 to 27.2% in 1900 (Prados de la Escosura, 2003). In addition, the share of the active population working in the industrial sector in 1900 was remarkable in some provinces such as Barcelona (35.5%), Guipúzcoa (31%) or Vizcaya (27%).

Chapter 6. Conclusions

The origins of the notable regional disparities in terms of income per capita that characterise the Spanish economy today can be found in the second half of the 19th century. At that time, 'modern economic growth' as defined by Kuznets (1966) arrived to the core European countries, and the process of structural change, i.e., the reallocation of resources from the agrarian to the industrial and services sectors, favoured the achievement of higher and self-sustained income per capita growth rates in these countries. Spain, in the periphery of Europe, could not join this process from its beginning. The outcome was the lack of convergence between Spain and the more developed economies in Europe in the period going from the mid-19th century until the Spanish Civil War. Therefore, when a long-term perspective is adopted, a positive relationship between industrialisation and economic growth emerges, at least from the 19th century onwards, although this link has weakened in the last decades.

As economic historians have argued, one of the main reasons that explains why the Spanish economy lagged behind the major economic powers in Europe before World War I was the 'failure' of the Industrial Revolution. However, the process of industrialisation was in motion in Spain and although this process did not spread out to the whole country, there were some regions where the industrial sector achieved a remarkable development. The second half of the 19th century witnessed a gradual concentration of industrial activities in a limited number of territories: Catalonia consolidated its predominant position as the 'factory of Spain', the Basque Country emerged as a powerful industrial centre, the western provinces in Andalusia lost ground, and the relative weight of industry declined in most of the inner regions in Spain. The higher spatial concentration of industry shows that structural change took place at different speed in the Spanish regions as illustrated by the increase in regional specialisation up to World War I. As a consequence, the pattern of geographical distribution of industry, the most dynamic sector of the economy, led to the creation of significant economic disparities across regions, in the period that the Spanish market became integrated.

In the last decades, the theoretical framework developed by the New Economic Geography appears to be particularly suitable for the study of the distribution of economic activities across space. The New Economic Geography is not based on the neoclassical assumptions of perfect competition and markets operating under constant returns to scale. Instead, NEG models assume monopolistic competition and increasing returns to scale in production. The interaction between economies of scale and falling transport costs may lead to the agglomeration of economic activity in a cumulative process in which the spatial concentration is reinforced once the process has started, and therefore a pattern of regional divergence is expected in the first stages of economic integration. Thus, this framework is also particularly suitable for undertaking historical studies.

In the Spanish case, the increase in regional inequality recorded during the second half of the 19th century occurred in parallel with the integration of the domestic market and the expansion of the process of industrialisation. First, market integration was favoured by the fall in transport costs. Second, as industrialisation advanced the presence of activities exploiting economies of scale increased. These two processes are well documented in the Spanish historiography as the survey of the initial chapter has shown. Taken together, the integration of the domestic market favoured the spatial concentration of industrial production recorded in the long period going from the mid-19th century to the 1930s, as NEG models predict. In this framework, a fundamental driver of agglomeration is market accessibility in the sense that market potential raises the price of the production factors and therefore, firms (capital) and workers (labour) will be attracted to high market potential locations.

Although the empirical research within the NEG is still lagging behind the more abundant theoretical developments, market potential has become a key variable to empirically test the theoretical predictions that can be derived from the NEG models, as the review of the NEG literature in chapter 2 has emphasised. Therefore, the research has been focused on the construction of market potential estimates for the Spanish provinces between the 1860s and 1930. In this case, market accessibility is measured with the Harris (1954) market potential equation following the recent application of this methodology carried out by Crafts (2005b) for Britain in the period 1871-1931. Albeit the Harris equation is an *ad hoc* measure developed by geographers, it is possible to establish a close relationship between this indicator and the measures of market accessibility structurally derived from the NEG models. According to this equation, the potential of a province is computed as the sum of the economic size of other provinces and export markets (usually GDP) discounted by the proximity, in terms of the distance and transport costs, to those markets. Hence, provincial GDP estimates are necessary for the calculation of market potential. This database has been constructed following the methodology proposed by Geary and Stark (2002).

The results show that the integration of the domestic market brought about significant changes in the relative accessibility of the Spanish provinces. In particular, it has been stressed that the construction and the progressive expansion of the railway network from the mid-19th century onwards, and the subsequent reduction in transport costs, explains these changes. In the 1860s, the geographical pattern of the market potential was characterised by a certain degree of stratification with three distinct groups of provinces in terms of their market potential. A first group of low market potential corresponds to those inland provinces that remained unconnected to the rail network in 1867, after the first wave in the construction of the railways had concluded. A second group was formed by the inland provinces connected to the network, with the sole exception of Madrid, a province that presented a market potential similar to that of the third group. This last group includes the coastal provinces, where the levels of market potential were higher.

Yet, in 1900 these three groups became two clearly differentiated groups and the geographical structure evolved towards a more polarised distribution with a division between inland and coastal provinces, with the latter showing a higher market potential than their inland counterparts, again with the exception of Madrid. It has been argued that the greater proximity of these provinces to foreign markets and the possibility to trade directly via coastal shipping with other Spanish ports would appear to be the most reasonable explanations for this difference. In 1900, when all the provincial capitals were connected to the railway network, this dual pattern was already established and once established, this division showed a persistent structure in the first three decades of the 20th century. Hence, the major changes in the market potential of Spanish provinces occurred during the second half of the 19th century when the basic rail network was constructed, transport costs (for railways and coastal shipping) were falling more intensely, and the Spanish market became integrated.

The availability of market potential estimates represents a very useful tool to undertake empirical exercises within the NEG. On the one hand, the presence of the economic forces stressed by the NEG models behind the remarkable increase in the concentration of manufacturing activities across the Spanish provinces between 1856 and 1929 has been tested. Then, in a second stage, the research has focused on analysing the impact of market potential on regional inequality in the same period.

The study of the industrial sector has been carried out using a standard NEG model which nests both comparative advantage in line with the Heckscher-Ohlin theorem and NEG-type mechanisms, as determinants of the location of industry. This model has been widely applied in economic history research in an attempt to explore with a theoretically-based exercise how the location of industry has responded to the forces stressed by these two explanations. The results of the exercise show that both comparative advantage and NEG forces were at work simultaneously in Spain, although the relative strength of these mechanisms varied over time.

In the first benchmark year considered in the mid-19th century, the spatial distribution of industry in Spain was determined by comparative advantage. At that time, the process of industrialisation in Spain was in its first stages, the domestic market was far from being fully integrated, and the intense process of spatial concentration in the location of industry had not started yet. Thus, the location of industry was driven by the relative endowment of land of the Spanish provinces. The predominance of the agrarian sector in the Spanish economy and the high share that the production of foodstuffs had in the total industrial production would explain this result.

From that moment onwards, when industry began to concentrate in a limited number of provinces, evidence of NEG effects is found. The main driving force behind this concentration of industry from the end of the 19th century until the 1930s was the interaction between market potential and economies of scale. Therefore, industrial activities producing under increasing returns tended to be located in high market potential provinces, in a context where transport costs were falling, the domestic market became gradually integrated, and the industrial sector was going through significant changes: industrialisation progressed, technical advances were incorporated to the production processes, and economies of scale achieved a higher development within the industrial sector. Yet, factor endowments were still relevant. The specialisation of the Spanish industry in the production of labour-intensive consumption goods offered an advantage to the provinces where labour was relatively abundant, although this advantage disappeared in the interwar years with the increasing production of capital goods.

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 and the abundance of natural resources. 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. Likewise, industrial activities were not attracted to the provinces where mineral resources were abundant. In this case, however, a positive relationship between the endowment of mineral resources and industry location is found in 1913, probably reflecting the change in comparative advantage that the protection on coal imports after 1891 conferred to the provinces where this mineral was produced. As regards the NEG variables, the scale effect was not amplified by the linkage effects that may appear when intermediate goods are considered. Here, it has been suggested that transport costs were not low enough at that time to exert a pull of centrality on industries with linkage effects.

In sum, although no evidence of NEG effects in mid-19th century is found, the research confirms the results previously obtained by other authors in the analysis of the determinants of industrial location in Spain: as the integration of the domestic market proceeded in the second half of the 19th century, increasing returns and market access became the main forces shaping the Spanish industrial map. Here, these results have been complemented with the examination of a larger number of variables that capture the relative strength of comparative advantage and NEG mechanisms making use of a theoretically sound empirical exercise, and expanded over time until the 1930s, thus covering the whole period in which the spatial concentration of industry in Spain took place.

In addition, the availability for other countries of similar studies to the one undertaken in these pages makes it possible to evaluate the Spanish experience in a comparative perspective at an international level. First, these studies confirm the relevance of both comparative advantage and NEG forces as determinants of industrial location across space. 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.

In the case of Poland in the interwar years (Wolf, 2007) the forces determining the location of industry were similar to the ones operating in the European Union between the 1970s and the 1990s (Midelfart-Knarvik, Overman, Redding and Venables, 2002). The comparative advantage of the regions in terms of the endowment of skilled workers and innovation activities shaped the industrial location although a relevant role was assigned to market potential through the interaction with intermediate goods (forward linkages). In turn, the British experience has been analysed in the period between 1871 and 1931 (Crafts and Mulatu, 2005, 2006), when the Industrial Revolution of the late 18th century was well under way, and the concentration of industry and regional specialisation showed, differently to the Spanish case, a stable evolution. In Britain, traditional factor endowments (skilled workers, coal abundance and agriculture) played the central role in the industrial location decisions, although NEG scale effects also had a remarkable influence. When compared to Spain, as regards factor endowments, skilled workers and mineral resources did not have the impact recorded for Britain. However, the evolution shown by NEG forces is rather similar: the interaction of market potential and economies of scale was the main NEG mechanism at work and no linkage effects were found up to the 1930s.

Finally, the US experience at the time that the manufacturing belt was created (Klein and Crafts, 2009) is the most comparable one to the Spanish case. In both cases, there was a remarkable concentration of industry across space and NEG effects exceeded comparative advantage as drivers of industrial location. In the US, as regards comparative advantage, only the agriculture endowment was significant and industries were not attracted to the states where skilled workers and coal were abundant. This result shows a strong resemblance with the Spanish case. However, although initially scale effects were the main NEG force in operation, eventually linkage effects increased their significance and by 1920 they had become the main determinants of industrial location in the US.

Once the impact of the mechanisms enhanced by the New Economic Geography on the spatial distribution of industry in Spain has been confirmed through the combined effect of regional access to demand and economies of scale, the aim of the final chapter in the dissertation is to examine whether geography also had an influence at a more aggregate level, when regional income per capita is considered. In so doing, the empirical strategy developed by Ottaviano and Pinelli (2006), in which growth literature and economic geography are combined, is applied to the Spanish case. These authors departed from a NEG model to derive standard growth regressions, where income per capita growth rates are regressed on the customary proximate sources of growth and a set of explanatory variables which consider wider influences on growth. The interesting contribution of this empirical strategy is that geography variables can be included among the wider influences in order to test if geography had an impact on provincial income per capita growth rates in the second half of the 19th century (1860-1900) and the first decades of the 20th century (1900-1930).

In addition, two alternative views of geography can be considered: 'first nature' and 'second nature' geography, as labelled by Krugman (1993). 'First nature' geography takes into account pure geography elements such as physical and climatic conditions of the territories in line with the work by Sachs and his co-authors. Likewise, 'second nature' geography refers to the mechanisms emphasised by Krugman and the New Economic Geography, and in this case, are captured by the variable market potential. The analysis of the results obtained is primarily focused on the effects of the geography variables on the provincial GDP per capita growth rates.

On the one hand, pure geography or 'first nature' geography influenced the economic growth attained by the Spanish provinces in the period under study. In the second half of the 19th century, the impact came from both the location of the provincial capitals and the climatic conditions. First, the altitude of the provincial capitals, the main population settlements where economic activity tends to concentrate, exerted a negative influence in the sense that higher altitude meant lower income per capita growth rates. A mountainous relief makes transport more costly, especially in the period when the basic railway network was under construction and traditional road transport was relatively more expensive. Second, the work on rough terrain is more complicated and typically, agrarian yields are negatively affected by altitude. On the other hand, the impact of climatic conditions on agrarian productivity is even more obvious. Here, the total number of sunshine hours per year had a strong and significant influence on provincial growth rates. The dryness of the climate directly affects the quality of land and therefore, under these conditions, it becomes more difficult to obtain high levels of productivity in agrarian activities. In this period, the agricultural sector was predominant in the Spanish economy, as shown by the fact that at an aggregate level, the workers enrolled in agrarian activities during the second half of the 1800s represented around two-thirds of the total Spanish active population. Therefore, through its direct impact on transport costs, and especially, on agrarian productivity first nature variables were the main geographic forces driving provincial GDP per capita growth differentials in the second half of the 19th century in Spain.

Nevertheless, in the first decades of the 20th century some remarkable changes are observed. Although pure geography elements are still significant, their impact which now comes only from the number of sunshine hours had weakened. More importantly, for the period 1900-1930 evidence of a positive and significant relationship between market potential and economic growth is found, as some recent cross-country studies within the NEG framework have demonstrated (Redding and Venables, 2004; Mayer, 2008). By 1900, the basic railway network had been finished and the integration of the domestic market was completed. Throughout this process, the relative market accessibility of the Spanish provinces had changed and at the end of the 19th century a polarised pattern in the geographical distribution of market potential had emerged. In addition, the advance in the process of industrialisation amplified the presence of increasing returns in the Spanish economy once the industrial sector achieved a higher weight within the productive structure of Spain. Hence, by 1900, when NEG-type mechanisms were already at work in the industrial sector, market accessibility also had an impact on the economic performance of the Spanish provinces in the subsequent decades up to the Spanish Civil War.

The research confirms the results reached in previous studies for Spain, in which 'first nature' geography variables shaped the spatial distribution of the population over the long term, as suggested by Dobado (2004, 2006). However, man-made economic geography is also relevant as shown by Ayuda, Collantes and Pinilla (2010). These authors proved that from 1900 onwards, increasing returns favoured the concentration of population in Spain and they argued that this process was the outcome of the combined effect between 'first nature' and 'second nature' geography. This result corroborated the findings of Pons and Tirado (2008), who concluded that in 1920 (and onwards) the differences in the relative economic density of Spanish provinces were explained by the interaction between 'first nature' and 'second nature' geography. The evidence provided in this thesis, where provincial GDP per capita growth rates are taken into account, goes in line with the findings of these studies.

Overall, it can be concluded that in the first stages of development in Spain, as the domestic market became integrated, the forces stressed by the New Economic Geography shaped the location of industry across space, through the interaction between market access and increasing returns in a context characterised by an intense fall in transport costs. In addition, market access played a significant role in explaining not only the location decisions in the industrial sector but also the differences in income per capita growth rates across provinces. In the light of these results, it seems clear that the role of geography needs to be considered in the analysis of regional inequality in the long term. In brief, geography matters.

Appendix

Table A1. Geographical distribution of Spanish exports (%)

	1865/69	1895/99	1910/13	1931/35
France	27,2	28,7	24,9	17,1
UK	29,9	26,1	21,4	23,6
Cuba	18,5	16,8	5,3	2,1
US	3,4	1,3	5,8	8,1
Belgium	0,6	2,5	3,9	4,6
Italy	1,0	1,4	3,5	4,5
Argentina	3,5	1,2	6,1	4,8
Philippines	0,4	4,0	0,7	-
Portugal	3,4	4,0	4,8	1,1
Puerto Rico	0,9	3,2	0,3	-
Uruguay	1,4	0,6	1,0	0,7
Germany	1,9	1,7	5,9	10,1
Nederland	0,6	2,2	5,4	4,7
Mexico	0,5	1,0	1,3	0,7

Source: 1865/69-1910/13, Prados de la Escosura (1982), p. 48; 1931/35, Tena (2005), p. 616. In bold, the markets included in the external market potential calculations.

Table A2. Provincial 'nodes' connected to the railway network in 1867

With connection to the railway network: Albacete, Alicante, Ávila, Badajoz, Barcelona, Bilbao, Burgos, Cádiz, Castellón, Ciudad Real, Córdoba, Girona, Guadalajara, Huesca, León, Lleida, Logroño, Madrid, Málaga, Murcia (Cartagena), Palencia, Pamplona, San Sebastián, Santander, Sevilla, Tarragona, Toledo, Valencia, Valladolid, Vitoria, Zamora y Zaragoza.
Without connection to the railway network: Almería, Cáceres, Coruña, Cuenca, Granada, Huelva, Jaén, Lugo, Ourense, Oviedo (Gijón), Pontevedra (Vigo), Salamanca, Segovia, Soria y Teruel.

Source: Wais (1987). In parenthesis the 'nodes' which do not coincide with the provincial capital.

Table A3. Agrarian and industrial wages in Spanish provinces, 1860-1930

	1860		1900		1914		1930	
	Wagr	Wind	Wagr	Wind	Wagr	Wind	Wagr	Wind
Álava	1	2,33	1,5	3,25	2,6	3,3	4,5	6,64
Albacete	1,05	1,77	1,6	2,75	2	3,2	4,5	5,60
Alicante	1,13	2,04	1,2	2,5	1,8	3,5	5,2	5,76
Almería	1,25	1,83	1,3	3,25	1,8	3,5	5,2	5,12
Ávila	1,13	1,83	1,2	1,75	2	3,8	2,56	6,40
Badajoz	1	1,88	2,3	2,62	1,8	3,1	4,35	5,04
Baleares	0,75	1,88	1,1	2,43	2,1	3,1	5	5,36
Barcelona	1,75	2,62	2,3	4,5	2,8	4,1	8,5	7,92
Burgos	0,96	1,94	1,7	2,18	2,3	3,4	5	5,76
Cáceres	0,88	1,64	1,1	2,12	1,5	2,8	3,2	4,40
Cádiz	1,25	2,39	1,2	3,25	1,8	4,5	5,4	7,04
Canarias	0,88	2,27	1,3	3,75	1,7	2,8	5	4,98
Castellón	1	2,20	1,3	2,87	1,6	3,1	5,67	5,76
Ciudad Real	0,78	2,38	1,4	2,31	1,8	3,7	5,5	5,12
Córdoba	1	1,57	1,2	3,14	1,8	3,7	5	6,48
Coruña	0,81	1,54	1,8	3	2	3,9	5,5	6,64
Cuenca	1,25	2,01	1,9	2,87	1,8	3,1	4,5	5,76
Girona	1,06	1,90	2	4,15	2,6	4,2	6,5	6,40
Granada	1	2,73	1,8	3	1,3	3,1	4,2	4,64
Guadalajara	1,16	2,01	1,9	3,43	1,7	3,4	4,2	6,00
Guipúzcoa	1,13	2,21	2,1	3,25	2,7	4,2	5,7	7,12
Huelva	1,25	2,05	2,5	3,56	2	3,7	4,7	6,24
Huesca	1,13	1,96	1,7	2,81	2,4	3,3	5,2	6,08
Jaén	1,13	2,45	1,3	3,15	1,8	3,9	5,2	5,84
León	0,63	1,70	0,8	2,75	1,9	4	4,2	7,60
Lleida	1,25	2,30	1,9	2,81	2,4	3,7	7,04	6,32
Logroño	0,94	2,13	1,7	3,31	2	3,4	5,5	5,12
Lugo	1	1,31	1,2	2,37	2,1	3,4	5	5,20
Madrid	1,25	2,32	1,5	3,56	2,1	4,5	6,08	7,68
Málaga	1,25	2,25	1,3	3,18	1,7	4	4,5	6,08
Murcia	1	2,14	1,3	2,74	1,7	3,2	4,7	5,92
Navarra	1,5	2,16	1,8	3,5	2,3	3,9	7,7	6,48
Ourense	0,88	1,37	1,7	2,12	2,1	3,3	5	5,76
Oviedo	1	1,92	1,5	3,06	2,5	4,1	6,7	8,64
Palencia	0,95	1,92	1,4	2,78	1,7	3,7	5	6,48
Pontevedra	0,63	1,08	1,6	2,5	2	3,4	4	5,92
Salamanca	0,88	1,64	1,1	4,68	2	3,2	2,5	5,92
Santander	1,13	2,10	1,8	4,25	2,3	4	5,5	8,08
Segovia	1,07	2,33	1,9	3,75	1,8	3,2	5,2	5,36
Sevilla	1,25	2,09	1,6	3,5	2,1	4	7	7,12
Soria	1	1,63	1,8	2,24	1,7	3,4	5,5	6,00
Tarragona	1,13	2,13	1,6	3,88	2	4	7,5	6,72
Teruel	1,13	1,76	2,1	2,13	1,9	3,1	5	5,04
Toledo	1	2,17	1,6	2,5	1,8	3,3	4,1	5,84
Valencia	1	2,04	2,5	2,75	1,9	2,9	4,5	5,68
Valladolid	0,88	2,03	2,2	2,63	1,8	3,5	4	6,00
Vizcaya	1,25	2,54	2	3	2,6	4	6,5	8,72
Zamora	0,8	1,83	1,5	3,75	1,8	3,1	3	5,60
Zaragoza	1,25	1,98	1,6	3,94	2,3	4,2	7,5	7,36

Sources: see text.

Table A4. Provincial Gross Domestic Product, 1860-1930 (current millions pesetas)

	1860	1900	1914	1930
Álava	36,7	64,9	85,4	141,4
Albacete	58,9	82,7	119,0	356,7
Alicante	152,9	166,3	292,0	750,6
Almería	100,6	113,0	149,9	317,7
Ávila	54,1	48,6	86,6	178,7
Badajoz	131,4	272,2	274,5	641,2
Baleares	93,9	135,7	287,5	538,0
Barcelona	516,8	1.391,5	1.806,6	5.206,3
Burgos	99,0	124,2	181,5	397,7
Cáceres	72,9	87,6	124,8	350,5
Cádiz	236,0	251,1	536,5	739,9
Canarias	76,3	158,7	232,8	644,3
Castellón	72,8	118,5	149,6	361,8
Ciudad Real	76,7	109,6	191,0	516,8
Córdoba	108,1	170,6	260,6	753,9
Coruña	177,8	304,3	388,8	1.170,3
Cuenca	77,8	99,4	105,8	241,1
Girona	116,2	229,0	366,9	530,1
Granada	160,2	214,7	188,4	499,1
Guadalajara	75,9	85,8	79,8	197,6
Guipúzcoa	73,1	192,6	313,1	598,8
Huelva	69,4	168,3	227,0	421,8
Huesca	96,4	101,1	142,4	311,1
Jaén	125,8	192,2	281,9	670,5
León	73,2	94,2	157,0	444,9
Lleida	104,1	122,6	163,8	460,6
Logroño	54,5	101,0	113,4	249,6
Lugo	83,8	148,4	223,6	472,0
Madrid	343,2	711,8	1.143,4	3.375,0
Málaga	190,7	184,9	264,4	625,3
Murcia	123,6	174,1	292,3	731,0
Navarra	142,1	156,5	199,9	541,4
Ourense	96,2	153,0	172,8	426,5
Oviedo	166,1	301,4	390,2	1.397,0
Palencia	60,9	68,0	92,5	234,0
Pontevedra	107,5	222,8	315,7	632,2
Salamanca	85,0	127,1	154,5	228,8
Santander	82,9	182,2	209,7	580,2
Segovia	62,5	73,1	81,5	185,2
Sevilla	209,8	313,0	427,8	1.314,7
Soria	39,7	57,4	55,0	169,3
Tarragona	108,6	198,3	220,2	570,5
Teruel	72,1	105,4	93,6	261,9
Toledo	111,0	140,1	178,6	387,8
Valencia	230,3	495,7	481,8	1.330,4
Valladolid	86,2	160,7	166,2	346,6
Vizcaya	88,2	298,0	368,2	1.143,2
Zamora	56,0	93,8	95,7	192,4
Zaragoza	153,7	238,0	286,1	959,2
Spain	**5.791,2**	**9.803,8**	**13.220,2**	**33.795,3**

Source: see text. For Spain's figures, Prados de la Escosura (2003).

Table A5. Market potential, 1867-1930 (millions pesetas, current prices)

	1867	1900	1914	1930
Álava	168	446	583	1.014
Albacete	143	353	474	925
Alicante	443	947	1.381	2.882
Almería	394	819	1.183	2.467
Ávila	160	386	531	1.000
Badajoz	103	338	411	758
Barcelona	657	1.883	2.506	5.278
Burgos	165	401	531	965
Cáceres	62	300	396	757
Cádiz	488	990	1535	2.840
Castellón	426	987	1336	2.786
Ciudad Real	130	343	478	933
Córdoba	161	398	553	1.067
Coruña	512	1.215	1.666	3.511
Cuenca	75	289	372	709
Girona	202	590	822	1.335
Granada	81	332	429	808
Guadalajara	187	470	631	1.264
Guipúzcoa	553	1.358	1.936	3.691
Huelva	392	938	1.322	2.711
Huesca	154	363	473	913
Jaén	103	366	506	942
León	123	366	493	944
Lleida	159	446	581	1.129
Logroño	158	427	538	984
Lugo	78	388	527	931
Madrid	307	842	1232	2.586
Málaga	457	925	1.334	2.727
Murcia	401	871	1.272	2.663
Navarra	184	441	573	1.078
Ourense	76	368	465	853
Oviedo	531	1.240	1.701	3.652
Palencia	174	424	550	1.021
Pontevedra	480	1.149	1.608	3.199
Salamanca	82	377	482	837
Santander	574	1.265	1.709	3.509
Segovia	103	419	550	1.051
Sevilla	444	1.002	1.416	3.012
Soria	71	306	388	762
Tarragona	463	1.128	1.516	3.151
Teruel	74	370	455	887
Toledo	183	444	603	1.167
Valencia	464	1.129	1.449	3.045
Valladolid	179	468	582	1.053
Vizcaya	570	1.517	2.026	4.278
Zamora	129	366	461	832
Zaragoza	175	440	560	1.142

Source: see text.

Table A6. Market potential relative to Barcelona and to 1867

	Barcelona = 100				1867 = 100			
	1867	1900	1914	1930	1867	1900	1914	1930
Álava	26	24	23	19	100	266	348	604
Albacete	22	19	19	18	100	247	331	647
Alicante	67	50	55	55	100	214	312	651
Almería	60	44	47	47	100	208	300	626
Ávila	24	21	21	19	100	241	331	624
Badajoz	16	18	16	14	100	327	398	734
Barcelona	100	100	100	100	100	287	382	804
Burgos	25	21	21	18	100	242	321	584
Cáceres	9	16	16	14	100	487	644	1.231
Cádiz	74	53	61	54	100	203	314	582
Castellón	65	52	53	53	100	232	314	655
Ciudad Real	20	18	19	18	100	263	367	717
Córdoba	24	21	22	20	100	248	344	664
Coruña	78	65	66	67	100	237	325	685
Cuenca	11	15	15	13	100	383	495	942
Girona	31	31	33	25	100	292	407	660
Granada	12	18	17	15	100	408	528	995
Guadalajara	29	25	25	24	100	251	337	675
Guipúzcoa	84	72	77	70	100	246	350	668
Huelva	60	50	53	51	100	239	337	691
Huesca	23	19	19	17	100	236	308	594
Jaén	16	19	20	18	100	357	493	918
León	19	19	20	18	100	297	401	767
Lleida	24	24	23	21	100	280	365	709
Logroño	24	23	21	19	100	270	341	623
Lugo	12	21	21	18	100	499	677	1.196
Madrid	47	45	49	49	100	274	401	841
Málaga	70	49	53	52	100	202	292	596
Murcia	61	46	51	50	100	217	317	664
Navarra	28	23	23	20	100	240	311	586
Ourense	12	20	19	16	100	483	610	1.120
Oviedo	81	66	68	69	100	234	320	688
Palencia	27	23	22	19	100	243	316	586
Pontevedra	73	61	64	61	100	239	335	666
Salamanca	12	20	19	16	100	461	589	1.022
Santander	87	67	68	66	100	220	298	611
Segovia	16	22	22	20	100	406	533	1.019
Sevilla	68	53	57	57	100	226	319	678
Soria	11	16	15	14	100	429	544	1.068
Tarragona	71	60	60	60	100	244	327	681
Teruel	11	20	18	17	100	500	614	1.199
Toledo	28	24	24	22	100	242	329	638
Valencia	71	60	58	58	100	243	312	656
Valladolid	27	25	23	20	100	261	325	588
Vizcaya	87	81	81	81	100	266	355	750
Zamora	20	19	18	16	100	284	357	644
Zaragoza	27	23	22	22	100	252	320	652

Source: see text.

Table A7. Market potential, self-potential excluded (millions pesetas, current prices)

	1867	1900	1914	1930
Álava	136	365	480	886
Albacete	120	307	409	779
Alicante	347	799	1.126	2.390
Almería	378	737	1.076	2.297
Ávila	132	349	466	901
Badajoz	61	212	287	541
Barcelona	376	804	1.135	2.316
Burgos	126	330	430	799
Cáceres	54	257	337	633
Cádiz	358	791	1.119	2.411
Castellón	383	888	1.214	2.564
Ciudad Real	104	290	387	749
Córdoba	117	299	405	746
Coruña	483	983	1.375	2.854
Cuenca	67	237	318	617
Girona	130	387	504	990
Granada	60	201	317	586
Guadalajara	154	417	583	1.174
Guipúzcoa	474	1.063	1.467	3.018
Huelva	382	824	1.172	2.501
Huesca	117	308	397	788
Jaén	87	253	344	653
León	95	314	410	765
Lleida	114	371	482	920
Logroño	121	330	432	808
Lugo	65	286	377	693
Madrid	124	301	381	701
Málaga	351	778	1.127	2.361
Murcia	346	760	1.088	2.319
Navarra	117	336	442	812
Ourense	59	246	329	602
Oviedo	507	1.040	1.448	2.973
Palencia	142	373	482	890
Pontevedra	457	923	1.294	2.727
Salamanca	71	300	390	734
Santander	520	1.095	1.517	3.111
Segovia	92	359	485	940
Sevilla	359	822	1.176	2.457
Soria	66	268	352	679
Tarragona	398	958	1.331	2.792
Teruel	65	311	403	780
Toledo	140	367	507	1.011
Valencia	358	804	1.140	2.405
Valladolid	133	346	459	861
Vizcaya	481	1.085	1.504	3.064
Zamora	103	304	399	739
Zaragoza	119	317	414	776

Source: see text.

Table A8. 'First nature' variables

	Altitude (*m*)	Rainfall (*mm*)	Temperature (*ºC*)	Sunshine hours	Coast
Álava	527,2	828	11	1.830	0
Albacete	686,0	336	14	2.730	0
Alicante	3,4	335	18	2.864	1
Almería	17,1	219	19	2.965	1
Ávila	1.131,5	369	10	2.644	0
Badajoz	184,8	538	16	2.830	0
Baleares	23,0	481	17	2.736	1
Barcelona	5,3	578	16	2.311	1
Burgos	856,2	486	10	2.183	0
Cáceres	439,3	562	16	2.890	0
Cádiz	4,1	546	18	2.752	1
Canarias	4,2	242	21	2.713	1
Castellón	28,2	405	17	2.689	1
Ciudad Real	635,1	377	14	2.656	0
Córdoba	122,2	631	17	2.800	0
Coruña	5,6	792	13	1.977	1
Cuenca	922,6	523	12	2.572	0
Girona	68,6	763	15	2.290	1
Granada	689,1	439	15	2.843	1
Guadalajara	679,1	384	13	2.440	0
Guipúzcoa	5,1	1.334	14	1.710	1
Huelva	3,7	444	17	2.998	1
Huesca	466,2	487	14	2.682	0
Jaén	573,8	628	16	2.800	0
León	822,8	965	10	2.369	0
Lleida	150,8	463	14	2.685	0
Logroño	384,4	392	13	2.242	0
Lugo	465,3	1.155	11	1.821	1
Madrid	688,3	420	13	2.622	0
Málaga	9,6	509	18	2.815	1
Murcia	43,5	289	18	2.649	1
Navarra	449,8	788	12	2.201	0
Ourense	126,0	1.155	11	2.043	0
Oviedo	228,7	966	13	1.711	1
Palencia	739,2	430	11	2.369	0
Pontevedra	19,5	1.455	14	2.218	1
Salamanca	798,2	396	12	2.586	0
Santander	5,0	1.191	14	1.638	1
Segovia	999,5	545	11	2.480	0
Sevilla	10,6	559	19	2.900	1
Soria	1.055,3	566	11	2.511	0
Tarragona	48,6	522	16	2.551	1
Teruel	915,7	381	12	2.596	0
Toledo	548,1	357	15	2.847	0
Valencia	13,3	416	17	2.683	1
Valladolid	692,6	407	12	2.590	0
Vizcaya	8,8	1.142	14	1.584	1
Zamora	651,0	255	13	2.587	0
Zaragoza	208,8	305	14	2.614	0

Sources: see text.

Annex 1. The calculation of internal transport costs between Spanish provinces

a) 1867

Transport costs are computed taking into account both the distances between the provincial 'nodes' and the average price for the transport of commodities obtained for the various modes of transport, as explained in chapter 3. In this year, three alternative transport modes have to be considered: transport by road, by coastal shipping and by railway. According to the map of the transport network in 1867, four different types of provinces can be distinguished on the basis of their geographical location and whether or not they had access to the railway network:

a) Coastal provinces with access to the railway (12 provinces).
b) Inland provinces with access to the railway (20 provinces).
c) Coastal provinces without access to the railway (5 provinces).
d) Inland provinces without access to the railway (10 provinces).

Based on this division, the matrix of internal bilateral transport costs between the provincial nodes has been constructed (table A9). This matrix [47x47], which has been particularly laborious to produce, contains 2,162 observations. The details for its calculation are detailed next:

- *Between two inland provinces with access to the railway*, the railway distances are taken from Wais (1987), obtaining a matrix [20x20].

- *Between inland provinces with access to the railway and coastal provinces with access to the railway*, distances are again calculated based on railway distances generating a matrix [20x12]. Then, for the distances *between coastal provinces with access to the railway and inland provinces with access to the railway* the previous matrix is transposed [12x20].

Table A9. Transport costs matrix for 1867

	Inland with railway	Coastal with railway	Coastal without railway	Inland without railway
Inland with railway	railway [20x20]	railway [20x12]	coastal shipping + railway [20x5]	road + railway [10x20]
Coastal with railway	railway [12x20]	0,7927 railway + 0,2073 coastal shipping [12x12]	0,7927[coastal shipping + railway] + 0,2073 coastal shipping [5x12]	Road + railway [10x12]
Coastal without railway	railway + coastal shipping [20x5]	0,7927[railway + coastal hipping] + 0,2073 coastal shipping [12x5]	*ATL: coastal shipping *MED: 0,7927 [railway + coastal shipping] + 0,2073 coastal shipping *ATL-MED: 0,7927 [railway + coastal shipping] + 0,2073 coastal shipping [5x5]	*ATL: road + railway + coastal shipping *MED: road + railway + coastal shipping [10x5]
Inland without railway	railway + road[295] [20x10]	railway + road [12x10]	*ATL: coastal shipping + railway + road[296] *MED: coastal shipping + railway + road [297] [5x10]	Road + railway + road[298] [10x10]

- *Between two coastal provinces with access to the railway*, the transport of goods can be made both by railway and by coastal shipping and therefore both means of transport have to be considered. The procedure goes in two steps. First, transport costs by railway and transport costs by coastal shipping are calculated. Second, these costs have to be weighted based on the total volume of commodities transported by each one of the modes of transport considered using the data provided by Frax (1981). Thus, in 1867, it is assumed that 20.73% of the transport of commodities was made by coastal shipping and the rest (79.27%) was made by railway (see table 3.1 in chapter 3). A matrix [12x12] is obtained.

- *Between inland provinces with access to the railway and coastal provinces without access to the railway*, the five coastal provinces without access to the network are divided in those located in the Atlantic seaboard and in the Mediterranean seaboard. In the first case, for Coruña, Oviedo (Gijón) and Pontevedra (Vigo) the cost by railway to the closest

[295] For the inland provinces of Galicia, coastal shipping has to be added. Details below.
[296] With the exception of Lugo and Ourense. Detils below.
[297] The routes Almería-Granada and Almería-Jaén, directly by road. Details below.
[298] For the inland provinces of Galicia, coastal shipping has to be added. Details below.

port to these provinces (Santander) is computed and then, the cost of coastal shipping to the ports of destination is added. As regards the Mediterranean ports, the provinces included are Almería and Huelva. Although the latter is on the Atlantic Ocean, for proximity and in order to simplify the calculations it is considered as a Mediterranean province. For this province, transport costs are obtained taking into account railway costs to the closest port (Cádiz) plus coastal shipping costs to the port of Huelva. In turn, for Almería, there are some differences in the computation to capture the locational advantage that represents the fact that this port can be connected to the railway network through two alternative ports which are at a similar distance from Almería (Cartagena in the east -176 km-, and Málaga in the west -185 km). From the 20 inland provinces with access to the railway considered in this case, a total of 19 gain access through Cartagena and only in one province the connection is made via Málaga. Hence, the absence of railway in these coastal provinces does not penalise them in excess as coastal shipping, which was a cheaper mean of transport than road transport, is available. Finally, a matrix [20x5] is obtained. *Between coastal provinces without access to the railway and inland provinces with access to the railway*, the previous results are once again transposed resulting in a matrix [5x20].

- *Between coastal provinces with access to the railway and coastal provinces without access to the railway*, transport is assumed to be both by railway and coastal shipping in a three-step process. First, the cost of railway transport from the coastal province connected to the network to the closest port to the coastal province of destination is computed and then coastal shipping is added to these costs. Second, there is the possibility of direct trade via coastal shipping. Third, these costs are weighted again based on the volume of goods transported by railway and coastal shipping calculated by Frax (1981) and displayed in table 3.1 in chapter 3. By proceeding in this way, it is possible to partially capture the disadvantage that the provinces excluded from the railway network had in this year, as coastal shipping was cheaper than railway transport especially for long distances. The calculations reproduce the procedure in the previous section, distinguishing between the Atlantic and the Mediterranean provinces. Some differences have to be stressed, though. First, between Huelva and Cádiz, on the one hand, and between Almería and Murcia, transport costs only include coastal shipping. In addition, in the case of Almería the connection by rail with the western provinces of Andalusia (Cádiz, Sevilla and Málaga) is made through the port of Málaga while the rest of the provinces are connected via the port of Cartagena. Then, a matrix [12x5] is obtained. *Between coastal provinces without access to the railway and coastal provinces with access to the railway*, the matrix is transposed [5x12].

- *Between two coastal provinces without access to the railway*, three alternative cases have to be considered. First, between the Atlantic nodes (Coruña, Gijón and Vigo) transport is assumed to be by coastal shipping. Next, between the Mediterranean ports (Cádiz and Huelva), the calculation is similar to the one in the previous section. Once the distances by both railway and coastal shipping (through Málaga), and coastal shipping from Cádiz and Huelva are computed, transport costs are obtained weighting the results once again according to the figures in Frax (1981). This procedure, although somewhat unrealistic, allows avoiding an underestimation in the computation of transport costs. Finally, between the Atlantic and the Mediterranean provinces, an analogous procedure is employed. From the Atlantic to Huelva: 79.23% by railway (coastal shipping to Santander, railway to Cádiz, and coastal shipping to Huelva) and 20.73% by coastal shipping directly. In the case of Almería, the connection is made through the port of Cartagena. This gives a matrix [5x5].

- *Between inland provinces with connection to the railways and inland provinces without connection to the railways*, transport costs are calculated combining rail and road transport. Since road transport is thrice as expensive as rail transport and more time-consuming, the connection to the railway network is assumed to be from the closest point in order to minimize road transport. A matrix [20x10] is obtained. *Between inland provinces without connection to the railways and inland provinces with connection to the railways* the results are transposed generating a matrix [10x20]. The routes to the ten inland provinces are detailed next:

- Cáceres. Railway to Mérida and Mérida-Cáceres by road.
- Cuenca. It can be connected to the network via Guadalajara (14 provinces) or Albacete (4 provinces). From these two nodes to Cuenca directly by road.
- Granada. Railway to Espeluy taking the distances from Córdoba, Albacete, Ciudad Real and Madrid according to Wais (1987). From Espeluy to Granada by road via Jaén.

- Jaén. Railway to Espeluy, and Espeluy-Jaén by road.
- Lugo. Railway to Santander, coastal shipping to Coruña and road to Lugo.
- Ourense. Railway to Santander, coastal shipping to Vigo and road to Ourense.
- Salamanca. Railway to Zamora and Zamora-Salamanca by road.
- Segovia. Railway to Collado Villalba and to Segovia by road.
- Soria. Since the road between Soria and Medinaceli had not been constructed yet, the connection is made through Catalayud. The central position of this town makes it accessible by rail via different locations in order to minimize transport costs: Guadalajara (10 provinces), Logroño (5 provinces), Zaragoza (4 provinces), and Navarra (1 province). Then, to Soria by road.
- Teruel. As in the previous case, it is accessible by rail from two alternative locations: Calatayud (15 provinces) and Sagunto (5 provinces). Then, to Teruel by road.

- *Between coastal provinces with access to the railway and inland provinces without access to the railway*, as in the previous case transport is assumed to be by railway and road, minimizing road distances. The only exception is the route Málaga-Granada were only road transport is considered. A matrix [12x10] is obtained. Then, *between inland provinces without access to the railway and coastal provinces with access to the railway*, the results are transposed giving a matrix [10x12].

- *Between coastal provinces without access to the railway and inland provinces without access to the railway*, a matrix [5x10] is obtained. And *between inland provinces without access to the railway and coastal provinces without access to the railway*, the transport cost matrix is transposed [10x5]. In this instance, the calculations differ for the provinces along the Atlantic and the Mediterranean seaboards:

- Atlantic provincial nodes (Coruña, Gijón, Vigo): coastal shipping to the closest port with access to the railway (Santander), then by railway, and by road. There are, however, some exceptions:
  - Lugo-Coruña: by road.
  - Lugo-Gijón: by road to Coruña and coastal shipping to Gijón.

  - Lugo-Vigo: by road via Santiago de Compostela as the road through Ourense did not exist at that time (Ministerio de Fomento, 1856).
  - Ourense-Coruña: by road via Santiago de Compostela.
  - Ourense-Gijón: by road to Vigo and coastal shipping to Gijón.
  - Ourense-Vigo: by road.

- Mediterranean provincial nodes:
  - Huelva: by coastal shipping to Cádiz, then railway and finally, by road.
  - Almería: by coastal shipping to Murcia (Cartagena), then railway and finally, by road. There are no provinces connected via Málaga. Instead, the routes Almería-Granada and Almería-Jaén are connected directly by road.

- *Between two inland provinces without access to the railway*, bilateral distances need to be calculated in order to complete the transport costs matrix [10x10]:

* From Cáceres to…:

Cuenca: road to Mérida, railway to Albacete, road to Cuenca.

Granada: road to Mérida, railway to Espeluy, road to Granada.

Jaén: road to Mérida, railway to Espeluy, road to Jaén.

Lugo: road to Mérida, railway to Santander, coastal shipping to Coruña, road to Lugo.

Ourense: road to Mérida, railway to Santander, coastal shipping to Vigo, road to Ourense.

Salamanca: road to Mérida, railway to Zamora, road to Salamanca.

Segovia: road to Mérida, railway to Villalba, road to Segovia.

Soria: road to Mérida, railway to Calatayud (via Guadalajara), road to Soria.

Teruel: road to Mérida, railway to Sagunto, road to Teruel.

* From Cuenca to…:

Granada: road to Albacete, railway to Espeluy, road to Granada.

Jaén: road to Albacete, railway to Espeluy, road to Jaén.

Lugo: road to Guadalajara, railway to Santander, coastal shipping to Coruña, road to Lugo.

Ourense: road to Guadalajara, railway to Santander, coastal shipping to Vigo, road to Ourense.

Salamanca: road to Guadalajara, railway to Zamora, road to Salamanca.

Segovia: road to Guadalajara, railway to Villalba, road to Segovia.

Soria: road to Guadalajara, railway to Calatayud, road to Soria.

Teruel: road Cuenca-Teruel.

* From Granada to…:

Jaén: road Granada-Jaén.

Lugo: road to Espeluy, railway to Santander, coastal shipping to Coruña, road to Lugo.

Ourense: road to Espeluy, railway to Santander, coastal shipping to Vigo, road to Ourense.

Salamanca: road to Espeluy, railway to Zamora, road to Salamanca.

Segovia: road to Espeluy, railway to Villalba, road to Segovia.

Soria: road to Espeluy, railway to Calatayud (via Guadalajara), road to Soria.

Teruel: road to Espeluy, railway to Sagunto, road to Teruel.

* From Jaén to…:

Lugo: road to Espeluy, railway to Santander, coastal shipping to Coruña, road to Lugo.

Ourense: road to Espeluy, railway to Santander, coastal shipping to Vigo, road to Ourense.

Salamanca: road to Espeluy, railway to Zamora, road to Salamanca.

Segovia: road to Espeluy, railway to Villalba, road to Segovia.

Soria: road to Espeluy, railway to Calatayud (via Guadalajara), road to Soria.

Teruel: road to Espeluy, railway to Sagunto, road to Teruel.

* From Lugo to…:

Ourense: road Lugo-Ourense.

Salamanca: road to Coruña, coastal shipping to Santander, railway to Zamora, road to Salamanca.

Segovia: road to Coruña, coastal shipping to Santander, railway to Villalba, road to Segovia.

Soria: road to Coruña, coastal shipping to Santander, railway to Calatayud (via Logroño), road to Soria.

Teruel: road to Coruña, coastal shipping to Santander, railway to Calatayud (via Logroño), road to Teruel.

* From Ourense to…:

Salamanca: road to Vigo, coastal shipping to Santander, railway to Zamora, road to Salamanca.

Segovia: road to Vigo, coastal shipping to Santander, railway to Villalba, road to Segovia.

Soria: road to Vigo, coastal shipping to Santander, railway to Calatayud (via Logroño), road to Soria.

Teruel: road to Vigo, coastal shipping to Santander, railway to Calatayud (via Logroño), road to Teruel.

* From Salamanca to…:

Segovia: road to Zamora, railway to Villalba, road to Segovia.

Soria: road to Zamora, railway to Calatayud (via Guadalajara), road to Soria.

Teruel: road to Zamora, railway to Calatayud (via Guadalajara), road to Teruel.

* From Segovia to…:

Soria: road to Villalba, railway to Calatayud (via Guadalajara), road to Soria.

Teruel: road to Villalba, railway to Calatayud (via Guadalajara), road to Teruel.

* From Soria to…:

Teruel: road Soria-Teruel (via Calatayud).

b) 1900-1930

In 1900, as mentioned in chapter 3, all the provincial nodes selected were connected to the railway network. Thus, from that moment onwards, internal transportation of goods is assumed to be by railway and it is no longer necessary to include road haulage in the calculations. Then, coastal shipping has to be added in the coastal provinces. The methodology is the same for the three benchmark years analysed: 1900, 1914 and 1930. In this case, two types of provinces have to be distinguished according to their geographical location:

a) Coastal provinces (17 provinces).

b) Inland provinces (30 provinces).

Based on this division, the matrix of internal bilateral transport costs between the provincial nodes has been constructed. The details for its calculation are detailed next:

- *Between two inland provinces*, transport is assumed to be by railway, where distances are obtained from Ministerio de Obras Públicas (1902). Matrix [30x30].

- *Between inland provinces and coastal provinces*, transport is assumed to be by railway. Matrix [30x17].

- *Between coastal provinces and inland provinces*, the previous results are transposed giving a matrix [17x30].

- *Between coastal provinces*, both railway transport and coastal shipping are considered. These transport costs are weighted on the basis of the volume of trade information provided by Frax (1981) and shown in table 3.1 in chapter 3. A matrix [17x17] is obtained.

Annex 2. Transport costs and the calculation of the foreign market potential

a) 1867

The computation of the foreign market potential in 1867 is based on a reduced version of the gravity equation in Estevadeordal, Frantz and Taylor (2003), considering the GDP of the foreign markets, maritime distances and average tariffs as detailed in chapter 3. For this year, it is necessary to distinguish between the 17 coastal provinces and the 30 inland provinces:

Coastal provinces. Transport costs between the Spanish coastal 'nodes' to the foreign ports of destination are calculated reducing the size of the foreign markets measured by the GDP on the basis of the distances and tariffs in force in such markets following the next equation:

$$\varphi_{rs} = GDP_s \left(dista\eta ce\right)_{rs}^{-0.8} (tariffs)_s^{-1.0}$$

Hence, the calculation of foreign market potential differs from the methodology applied to obtain the internal market potential, where provincial GDPs were weighted by transport costs assuming an elasticity of *-1*. In this case, foreign GDP is reduced taking distances (with an elasticity of -0.8 as suggested by Estevadeordal, Frantz and Taylor (2003)), and the average tariffs applied in the four countries considered (elasticity *-1*).

Inland provinces. In this case, a two-step procedure is used. First, it is necessary to compute transport costs from the inland province to the closest Spanish port taken as a provincial 'node'. The foreign GDP is therefore reduced according to the internal transport costs in the same way as it was done in the calculation of the internal market potential, and not using the elasticities derived from foreign trade. Second, the share of the foreign GDP which is still available in the Spanish port of origin, is decreased according to the gravity equation and the elasticities provided by Estevadeordal, Frantz and Taylor (2003), reproducing the methodology described for the coastal provinces, but now, departing from a lower GDP once the internal transport costs have been discounted.

In order to implement this procedure, it is first required to decide which is the closest Spanish port for each inland node distinguishing between the provinces whose products would be exported through the Atlantic seaboard and/or through a Mediterranean port. The decision is based on the transport costs obtained in the calculation of the internal market potential.

Some provinces (Albacete, Ciudad Real, Córdoba, Cuenca, Girona, Granada, Jaén, Lleida and Teruel) have the closest port in the **Mediterranean**. Transport costs are computed as internal transport costs to reach the Mediterranean port and then, from that port to the foreign port of destination. However, most of the foreign nodes considered are located in the Atlantic Ocean. Therefore, it is necessary to check whether or not is convenient for these provinces to export through the Mediterranean coast since they might have a higher market potential if they export from the Atlantic seaboard even though they have to face higher internal transport costs. That would be the case of the inland provinces located in the centre of the Peninsula, where the differential between exporting from Alicante (in the Mediterranean) or do it from Santander (in the Atlantic) would be very small. The two options have been computed for each provincial node and in the case of these nine provinces the Mediterranean option implies lower transport costs (table A10).

Nevertheless, there are some provinces (Badajoz, Cáceres, Guadalajara, Huesca, Madrid and Toledo) whose transport costs are lower to the Mediterranean ports, but overall this option generates a lower foreign market potential. For that reason, the Atlantic ports which result in a higher foreign market potential are preferred (see table A11). In these six provinces, exporting through the Atlantic allows them to be closer to the foreign markets, but there is an exception: the east of France. In order to reach the port of Marseille, it is assumed that exports leave through the Mediterranean port. This option sheds a higher market potential for the market of France (East).

On the other hand, there are fifteen provinces (Álava, Ávila, Burgos, León, Logroño, Lugo, Navarra, Ourense, Palencia, Salamanca, Segovia, Soria, Valladolid, Zamora and Zaragoza) whose location in relation to the railway network makes it cheaper to export from the **Atlantic** ports, i.e., with this option they have to face lower transport costs. As in the previous case, the only exception is Marseille, the node of the France (East) market. This route goes along the Mediterranean ports.

Table A10. Inland provinces (Mediterranean) and Mediterranean ports, 1867

	Mediterranean ports
Albacete	Alicante
Ciudad Real	Alicante
Córdoba	Sevilla
Cuenca	Alicante
Girona	Barcelona
Granada	Málaga
Jaén	Sevilla
Lleida	Barcelona
Teruel	Valencia

Table A11. Inland provinces (Mediterranean) and Atlantic ports, 1867

	Atlantic ports	Mediterranean ports (for France-East)
Badajoz	Santander	Alicante
Cáceres	Santander	Alicante
Guadalajara	Santander	Alicante
Huesca	Bilbao	Barcelona
Madrid	Santander	Alicante
Toledo	Santander	Alicante

It has to be born in mind that in some instances, the closest Mediterranean port does not offer the higher market potential. It depends on the position of each province. As a way of example, it might compensate to reach Barcelona by railway although transport costs are higher, because the maritime distance from this port to Marseille is lower. However, differences are small in volume and are only affecting a share of the French market (East). Therefore, this option is not considered and in order to keep a homogeneous procedure, the closest Mediterranean port is taken. There are some exceptions, nonetheless. There are three provinces (Lugo, Ourense and Palencia) whose location near the Atlantic Ocean and far away of the Mediterranean seaboard suggests a different strategy. Instead of going by railway to the port of Barcelona, and then by ship to Marseille, it is considered that they export directly from the Atlantic ports of Coruña, Vigo and Santander, respectively. Taking into account these considerations, the ports selected for these provinces are displayed in table A12.

Table A12. Inland provinces (Atlantic) and Atlantic ports, 1867

	Atlantic ports	Mediterranean ports (for France-East)
Álava	Guipúzcoa (Pasajes)	Barcelona
Ávila	Santander	Alicante
Burgos	Vizcaya (Bilbao)	Barcelona
León	Santander	Alicante
Logroño	Vizcaya (Bilbao)	Barcelona
Lugo	Coruña	-
Navarra	Guipúzcoa (Pasajes)	Barcelona
Ourense	Pontevedra (Vigo)	-
Palencia	Santander	-
Salamanca	Santander	Alicante
Segovia	Santander	Alicante
Soria	Vizcaya (Bilbao)	Barcelona
Valladolid	Santander	Alicante
Zamora	Santander	Alicante
Zaragoza	Vizcaya (Bilbao)	Barcelona

b) 1900-1930

In this case, a different procedure has been applied again for coastal and inland provinces. The methodology employed is the same in the years 1900, 1914 and 1930.

Coastal provinces. The calculation of the foreign market potential for these provinces is similar to the one described for 1867, following the reduced version of a gravity equation were foreign GDP is discounted on the basis of the distance, the average tariffs in the country of destination and their respective elasticities estimated by Estevadeordal, Frantz and Taylor (2003).

Inland provinces. Overall, the procedure is analogous to that of 1867, although there are some differences derived from the changes operated in the railway network in the second half of the 19th century. The results are affected in the sense that the closest or cheapest Spanish ports differ. Now, sixteen provinces export from the **Mediterranean** ports which means that railway costs to such ports have to be calculated and then, foreign transport costs in the maritime route have to be added. Again, the question is whether or not the Mediterranean ports provide a higher foreign market potential for these provinces. As before, in some cases the lower maritime distances to the foreign markets more than offset the higher internal transport costs by railway and therefore, the Atlantic ports are preferred. Once the cost of both options has been calculated, for eleven provinces exports would leave from the Mediterranean ports (table A13). However, once again, there are some provinces (Cuenca, Madrid, Soria, Toledo and Zaragoza) whose exports should initially leave from the Mediterranean ports, but whose market potential increases if they do it from the Atlantic seaboard. Nevertheless, the market potential that represents France (East) is calculated, as previously, taking the closest Mediterranean port (table A14).

Finally, fourteen inland provinces would trade through the **Atlantic** ports in order to face lower internal transport costs and therefore, a higher foreign market potential (table A15). In the case of the French (East) market, accessibility is calculated assuming that exports leave from the closest Mediterranean port. As before, for the inland provinces of León, Lugo and Ourense, the Mediterranean option does not improve their market potential and hence, it is assumed that exports are sent to Marseille through the Atlantic ports of Gijón, Coruña and Vigo, respectively.

Table A13. Inland provinces (Mediterranean) and Mediterranean ports, 1900-1930

	Mediterranean ports
Albacete	Alicante
Badajoz	Sevilla
Cáceres	Sevilla
Ciudad Real	Alicante
Córdoba	Sevilla
Girona	Barcelona
Granada	Málaga[299]
Huesca	Tarragona
Jaén	Málaga
Lleida	Tarragona
Teruel	Valencia

Table A14. Inland provinces (Mediterranean) and Atlantic ports, 1900-1930

	Atlantic ports	Mediterranean ports (for France-East)
Cuenca	Santander	Alicante
Madrid	Santander	Alicante
Soria	Santander	Tarragona
Toledo	Santander	Alicante
Zaragoza	Guipúzcoa (Pasajes)	Tarragona

[299] In 1914 and 1930, the closest Mediterranean port for Granada is Almería.

Table A15. Inland provinces (Atlantic) and Atlantic ports, 1900-1930

	Atlantic ports	Mediterranean ports (for France-East)
Álava	Guipúzcoa (Pasajes)	Tarragona
Ávila	Santander	Alicante
Burgos	Vizcaya (Bilbao)	Tarragona
Guadalajara	Guipúzcoa (Pasajes)	Alicante
León	Oviedo (Gijón)	-
Logroño	Vizcaya (Bilbao)	Tarragona
Lugo	Coruña	-
Navarra	Guipúzcoa (Pasajes)	Tarragona
Ourense	Pontevedra (Vigo)	-
Palencia	Santander	Tarragona
Salamanca	Santander	Alicante
Segovia	Santander	Alicante
Valladolid	Santander	Tarragona
Zamora	Oviedo (Gijón)	Alicante

Bibliography

Acemoglu, D. (2009). *Introduction to modern economic growth*, Princeton University Press, Princeton (NJ).

Acemoglu, D., Johnson, S. and Robinson, J.A. (2001). The colonial origins of comparative development: an empirical investigation, *American Economic Review* 91(5), 1369-1401.

Acemoglu, D., Johnson, S. and Robinson, J.A. (2002). Reversal of fortune: geography and institutions in the making of the modern world income distribution, *Quarterly Journal of Economics* 117(4), 1231-1294.

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

Aixalá, J. (1999). *La peseta y los precios: un análisis de largo plazo (1868-1995)*, Prensas Universitarias de Zaragoza, Zaragoza.

Alcaide Inchausti, J. (2003). *Evolución económica de las regiones y provincias españolas en el siglo XX*, Fundación BBV, Bilbao.

Alonso, M. (1964). *Location and land use*, Harvard University Press, Cambridge (MA).

Alonso-Villar, O. (2001). Large metropolises in the Third World: an explanation, *Urban Studies* 38(8), 1359-1371.

Álvarez Llano, R. (1986). Evolución de la estructura económica regional de España en la historia: una aproximación, *Situación* 1, 5-61.

Amiti, M. and Cameron, L. (2007). Economic geography and wages, *Review of Economics and Statistics* 89(1), 15-29.

Anderson, J.E. and Van Wincoop, E. (2003). Gravity with gravitas: a solution to the border puzzle, *American Economic Review* 93(1), 170-192.

Anes Álvarez, R. (1978). Relaciones entre el ferrocarril y la economía española (1865-1935). In Artola (ed.), *Los ferrocarriles en España: 1844-1943*, vol.2, Banco de España, Madrid, 355-512.

Anes, R., Tortella, G. and Schwartz, P. (1974). *La banca española en la Restauración*, Servicio de Estudios del Banco de España, Madrid.

Angier, E. (1920). *Fifty years' freigths, 1869-1919*, Fairplay, London.

Anuario Estadístico de España – Spanish Statistical Yearbook (ad annum), Madrid.

Armstrong, H.W. (1995). Convergence among regions of the European Union, 1950-1990, *Papers in Regional Science* 74(2), 143-152.

Aschauer, D.A. (1989). Is public expenditure productive?, *Journal of Monetary Economics* 23(2), 177-200.

Ayuda, M.I., Collantes, F. and Pinilla, V. (2010). From locational fundamentals to increasing returns: the spatial concentration of population in Spain, 1787-2000, *Journal of Geographical Systems* 12(1), 25-50.

Bairoch, P. (1989). European trade policy, 1815-1914. In Mathias and Pollard (eds.), *The Cambridge economic history of Europe*, vol.8, Cambridge University Press, Cambridge (MA), 1-160.

Barquín, R. (1997). Transporte y precio del trigo en el siglo XIX: creación y reordenación de un mercado nacional, *Revista de Historia Económica* 15(1), 17-48.

Barquín, R. (1999). El ahorro social: una perspectiva dinámica. In Muñoz, Sanz and Vidal (eds.), *Siglo y medio del ferrocarril en España 1848-1998. Economía, industria y sociedad*, Fundación de los Ferrocarriles Españoles, Madrid, 337-354.

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

Barro, R.J. (1991). Economic growth in a cross section of countries, *Quarterly Journal of Economics* 106(2), 407-443.

Barro, R.J. and Sala-i-Martin, X. (1991). Convergence across states and regions, *Brookings Papers on Economic Activity* 1, 107-182.

Barro, R.J. and Sala-i-Martin, X. (1992). Convergence, *Journal of Political Economy* 100(2), 223-251.

Barro, R.J. and Sala-i-Martin, X. (1995). *Economic growth*, McGraw-Hill, New York.

Baum, C.F., Schaffer, M.E. and Stillman, S. (2003). Instrumental variables and GMM, *Stata Journal* 3(1), 1-31.

BBV (1999). *Renta nacional de España y su distribución provincial: serie homogénea años 1955 a 1993 y avances 1994 a 1997*, Fundación BBV, Bilbao.

BBV (2000). *Renta nacional de España y su distribución provincial: año 1995 y avances 1996-1999*, Fundación BBV, Bilbao.

Behrens, K. (2003). Assymetric trade and agglomeration, *LEG Document de Travail*, 2003-10.

Behrens, K., Gaigné, C., Ottaviano, G.I.P. and Thisse, J.F. (2006). Is remoteness a locational disadvantage?, *Journal of Economic Geography* 6(3), 347-368.

Behrens, K., Gaigné, C., Ottaviano, G.I.P. and Thisse, J.F. (2007). Countries, regions and trade: on the welfare impacts of economic integration, *European Economic Review* 51(5), 1277-1301.

Behrens, K. and Thisse, J.F. (2007). Regional economics: a new economic geography perspective, *Regional Science and Urban Economics* 37(4), 457-465.

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(3), 663-696.

Boyer, G.R. and Hatton, T.J. (1994). Regional labour market integration in England and Wales, 1850-1913. In Grantham and MacKinnon (eds.), *Labour market evolution. The economic history of market integration, wage flexibility and the employment relation*, Routledge, London, 84-106.

Brakman, S., Garretsen, H. and Schramm, M. (2004). The spatial distribution of wages: estimating the Helpman-Hanson model for Germany, *Journal of Regional Science* 44(3), 437-466.

Bringas, M.Á. (2000). *La productividad de los factores en la agricultura española (1752-1935)*, Banco de España, Madrid.

Brown, J.C. (1995). Imperfect competition and Anglo-German trade rivalry: markets for cotton textiles before 1914, *Journal of Economic History* 55(3), 494-527.

Brülhart, M., Crozet, M. and Koenig, P. (2004). Enlargement and the EU periphery: the impact of changing market potential, *World Economy* 27(6), 853-875.

Brülhart, M. and Torstensson, J. (1996). Regional integration, scale economies and industry location in the European Union, *CEPR Discussion Papers*, 1435.

Button, K.J. and Pentecost, E.J. (1995). Testing for convergence of the EU regional economies, *Economic Inquiry* 33(4), 664-671.

Calvo, Á. (2001). Los inicios de las telecomunicaciones en España: el telégrafo, *Revista de Historia Económica* 19(3), 613-636.

Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics: methods and applications*, Cambridge University Press, New York.

Canova, F. and Marcet, A. (1995). The poor stay poor: non-convergence across countries and regions, *CEPR Discussion Papers*, 1265.

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* 2(1), 127-157.

Carreras, A. (1990a). Fuentes y datos para el análisis regional de la industrialización española. In Nadal and Carreras (eds.), *Pautas regionales de la industrialización española (siglos XIX y XX)*, Ariel, Barcelona, 3-20.

Carreras, A. (1990b). *Industrialización española: estudios de historia cuantitativa*, Espasa-Calpe, Madrid.

Carreras, A. and Tafunell, X. (2004). *Historia económica de la España contemporánea*, Crítica, Barcelona.

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

Caselli, F., Esquivel, G. and Lefort, F. (1996). Reopening the convergence debate: a new look at cross-country growth empirics, *Journal of Economic Growth* 1(3), 363-389.

Castañeda, L. and 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(2), 367-383.

Censos de Población – Spanish Population Censuses (ad annum), Madrid.

Chamberlin, E. (1933). *The theory of monopolistic competition*, Harvard University Press, Cambridge (MA).

Chastagnaret, G. (2000). *L'Espagne, puissance minière dans l'Europe du XIXe siècle*, Casa de Velázquez, Madrid.

Christaller, W. (1933). *Die zentralen orte in süddeutschland*, Gustav Fischer, Jena (English translation, 1966: *Central places in Southern Germany*, Prentice Hall, Londres).

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

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

Clark, E.C., Herring, W.E., Knapp, T.A. and White, N.E. (2003). Migration and implicit amenity markets: does incomplete compensation matter?, *Journal of Economic Geography* 3(3), 289-307.

Clark, C., Wilson, F. and Bradley, J. (1969). Industrial location and economic potential in Western Europe, *Regional Studies* 3(2), 197-212.

Clemens, M.A. and Williamson, J.G. (2001). A tariff-growth paradox? Protection's impact the world around 1875-1997, *NBER Working Papers*, 8459.

Coll, S. (1985). El sector minero, *Información Comercial Española* 623, 83-96.

Combes, P.Ph., Duranton, G. and Gobillon, L. (2008). Spatial wage disparities: sorting matters!, *Journal of Urban Economics* 63(2), 723-742.

Combes, P.Ph. and Lafourcade, M. (2005). Transport costs: measures, determinants, and regional policy implications for France, *Journal of Economic Geography* 5(3), 319-349.

Combes, P.Ph., Lafourcade, M., Thisse, J.F. and Toutain, J. (2008). The rise and fall of spatial inequalities in France: a long-run perspective, *CEPR Discussion Papers*, 7017.

Combes, P.Ph., Mayer, T. and Thisse, J.F. (2008). *Economic geography*, Princeton University Press, Princeton (NJ).

Combes, P.Ph. and Overman, H. (2004). The spatial distribution of economic activities in the European Union. In Henderson and Thisse (eds.), *Handbook of Regional and Urban Economics*, vol. 4, Elsevier, Amsterdam, 2845-2909.

Comín, F., Martín Aceña, P., Muñoz, M. and Vidal, J. (1998). *150 años de historia de los ferrocarriles españoles*, Fundación de los Ferrocarriles Españoles, Madrid.

Cordero, R. and Menéndez, F. (1978). El sistema ferroviario español. In Artola (ed.), *Los ferrocarriles en España: 1844-1943*, vol.1, Banco de España, Madrid, 161-338.

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

Crafts, N. (2005b). Market potential in British regions, 1871-1931, *Regional Studies* 39(9), 1159-1166.

Crafts, N. and Mulatu, A. (2005). What explains the location of industry in Britain, 1871–1931?, *Journal of Economic Geography* 5(4), 499-518.

Crafts, N. and Mulatu, A. (2006). How did the location of industry respond to falling transport costs in Britain before World War I?, *Journal of Economic History* 66(3), 575-607.

Crozet, M. (2004). Do migrants follow market potentials? An estimation of a new economic geography model, *Journal of Economic Geography* 4(4), 439-458.

Crozet, M. and Koenig, P. (2004a). Trade liberalization and the internal geography of countries. In Mayer and Mucchielli (eds.), *Multinational firms' location and economic geography*, Edward Elgar, Cheltenham, 91-109.

Crozet, M. and Koenig, P. (2004b). EU enlargement and the internal geography of countries, *Journal of Comparative Economics* 32(2), 265-279.

Crozet, M., Mayer, T. and Mucchielli, J. (2004). How do firms agglomerate? A study of FDI in France, *Regional Science and Urban Economics* 34(1), 27-54.

Cuadrado Roura, J.R. (2010). *Regional policy, economic growth and convergence: lessons from the Spanish case*, Springer, Heidelberg.

Davis, D.R. and Weinstein, D.E. (1999). Economic geography and regional production structure: an empirical investigation, *European Economic Review* 43(2), 379-407.

Davis, D.R. and Weinstein, D.E. (2003). Market access, economic geography and comparative advantage: an empirical test, *Journal of International Economics* 59(1), 1-23.

De Bruyne, K. (2003). The location of economic activity. Is there a spatial employment structure in Belgium?, *Katholieke Universiteit Leuven*.

de la Fuente, A. (2002). On the sources of convergence: a close look at the Spanish regions, *European Economic Review* 46(3), 569-599.

Devereux, M.P. and Griffith, R. (1998). Taxes and the location of production: evidence from a panel of US multinationals, *Journal of Public Economics* 68(3), 335-367.

Dirección General de Obras Públicas (1861). *Memoria sobre el progreso de las obras públicas en España, en los años 1859 y 1860*, Madrid.

Disdier, A. and Head, K. (2008). The puzzling persistence of the distance effect on bilateral trade, *Review of Economics and Statistics* 90(1), 37-48.

Dixit, A.K. and Stiglitz, J.E. (1977). Monopolistic competition and optimum product diversity, *American Economic Review* 67(3), 297-308.

Dobado, R. (2004). Un legado peculiar: la geografía. In Llopis (ed.), *El legado económico del Antiguo Régimen en España*, Crítica, Barcelona, 97-119.

Dobado, R. (2006). Geografía y desigualdad económica y demográfica de las provincias españolas (siglos XIX-XX), *Investigaciones de Historia Económica* 5, 133-170.

Dolado, J.J., González-Páramo, J.M. and Roldán, J.M. (1994). Convergencia económica entre las provincias españolas: evidencia empírica (1955-1989), *Moneda y Crédito* 198, 81-131.

Domenech, J. (2008). Mineral resource abundance and regional growth in Spain, 1860-2000, *Journal of International Development* 20(8), 1122-1135.

Domínguez, R. (2002). *La riqueza de las regiones. Las desigualdades económicas regionales en España, 1700-2000*, Alianza, Madrid.

Eichengreen, B. and Irwin, D.A. (1995). Trade blocs, currency blocs and the reorientation of world trade in the 1930s, *Journal of International Economics* 38(1-2), 1-24.

Ellison, G. and Glaeser, E.L. (1999). The geographic concentration of industry: does natural advantage explain agglomeration?, *American Economic Review* 89(9), 311-316.

Engerman, S.L. and Sokoloff, K.L. (2002). Factor endowments, inequality, and paths of development among new world economies, *NBER Working Papers*, 9259.

Erdozáin, P. and 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.

Estadística Administrativa de la Contribución Industrial y de Comercio – EACI (ad annum), Madrid.

Estadística de la Contribución de Utilidades – ECU (ad annum), Madrid.

Estadística Industrial (EI) de 1958 – Industrial Statistic 1958, Madrid.

Estevadeordal, A. (1997). Measuring protection in the early twentieth century, *European Review of Economic History* 1(1), 89-125.

Estevadeordal, A., Frantz, B. and Taylor, A.M. (2003). The rise and fall of world trade, 1870-1939, *Quarterly Journal of Economics* 118(2), 359-407.

Fagerberg, J. and Verspagen, B. (1996). Heading for divergence? Regional growth in Europe reconsidered, *Journal of Common Market Studies* 34(3), 431-448.

Faíña, J.A. and López-Rodríguez, J. (2006). Renta per cápita, potencial de mercado y proximidad: el caso de España, *Papeles de Economía Española* 107, 268-276.

Fally, T., Paillacar, R. and Terra, C. (2010). Economic geography and wages in Brazil: evidence from micro-data, *Journal of Development Economics* 91(1), 155-168.

Flandreau, M. and Maurel, M. (2005). Monetary union, trade integration, and business cycles in 19th century Europe, *Open Economies Review* 16(2), 135-152.

Foro Hispánico de Cultura (1957). *La población activa española de 1950 a 1957*, Madrid.

Frax, E. (1981). *Puertos y comercio de cabotaje en España, 1857-1934*, Banco de España, Madrid.

Frax, E. and Madrazo, S. (2001). El transporte por carretera, siglos XVIII-XX, *Transportes, Servicios y Telecomunicaciones* 1, 31-53.

Friedman, J., Gerlowski, D.A. and Silberman, J. (1992). What attracts foreign multinational corporations? Evidence from branch plant location in the United States, *Journal of Regional Science* 32(4), 403-418.

Frost, M. and Spence, N. (1995). The rediscovery of accessibility and economic potential: the critical issue of self-potential, *Environment and Planning A* 27, 1833-1848.

Fujita, M., Krugman, P. and Venables, A.J. (1999). *The spatial economy: cities, regions, and international trade*, The MIT press, Cambridge (MA).

Funcas (2006). *Balance económico regional (autonomías y provincias). Años 2000 a 2005*, Gabinete de Estadística Regional, Madrid.

Gallup, J.L. and Sachs, J.D. (2001). The economic burden of malaria, *American Journal of Tropical Medicine and Hygiene* 64, 85-96.

Gallup, J.L., Sachs, J.D. and Mellinger, A.D. (1999). Geography and economic development, *International Regional Science Review* 22(2), 179-232.

Galor, O. (1996). Convergence? Inferences from theoretical models, *Economic Journal* 106, 1056-1069.

Garcia Pires, A. (2006). Estimating Krugman's economic geography model for the Spanish regions, *Spanish Economic Review* 8(2), 83-112.

García Sanz, Á. (1980). Jornales agrícolas y presupuesto familiar campesino en España a mediados del siglo XIX, *Anales de CUNEF*, curso 1979-80, 51-71.

García-Greciano, B., Raymond, J.L. and Villaverde, J. (1995). Convergencia de las provincias españolas, *Papeles de Economía Española* 64, 38-53.

Garrabou, R. and Sanz, J. (1985). La agricultura española durante el siglo XIX: ¿inmovilismo o cambio? In Garrabou and Sanz (eds.), *Historia agraria de la España contemporánea. 2. Expansión y crisis (1850-1900)*, Crítica, Barcelona, 7-191.

Geary, F. and Stark, T. (2002). Examining Ireland's post-famine economic growth performance, *Economic Journal* 112, 919-935.

Germán, L., Llopis, E., Maluquer, J. and Zapata, S. (2001). *Historia económica regional de España, siglos XIX y XX*, Crítica, Barcelona.

Goerlich, F.J. and Mas, M. (2008). Drivers of agglomeration: geography vs. history, *MPRA Paper*, 15802.

Goldin, C. and Katz, L.F. (1998). The origins of technology-skill complementarity, *Quarterly Journal of Economics* 113(3), 693-732.

Gómez Mendoza, A. (1981). *Railways and Spanish economic growth in the late 19th century*. PhD dissertation (unpublished), University of Oxford.

Gómez Mendoza, A. (1982). *Ferrocarriles y cambio económico en España 1855-1913*, Alianza, Madrid.

Gómez Mendoza, A. (1999). Caballo de sangre, caballo de vapor en el transporte de fines del siglo XIX. In Carreras, Pascual, Reher and Sudrià (eds.), *Doctor Jordi Nadal. La industrialització i el desenvolupament econòmic d'Espanya*, vol.I. Universitat de Barcelona, Barcelona, 720-735.

Gómez Mendoza, A. and San Román, E. (2005). Transportes y comunicaciones. In Carreras and Tafunell (eds.), *Estadísticas históricas de España siglos XIX-XX*, vol.II, Fundación BBVA, Bilbao, 509-572.

Grubel, H.G. and Lloyd, P.J. (1975). *Intra-industry trade: the theory and measurement of internatinal trade in differetiated products*, Basingstoke, London.

Gutiérrez, J. and Urbano, P. (1996). Accessibility in the European Union: the impact of the trans-European road network, *Journal of Transport Geography* 4(1), 15-25.

Haaland, J.I., Kind, H.J., Midelfart-Knarvik, K.H. and Torstensson, J. (1999). What determines the economic geography of Europe?, *CEPR Discussion Papers*, 2072.

Hanson, G.H. (1996). Economic integration, intraindustry trade, and frontier regions, *European Economic Review* 40(3-5), 941-949.

Hanson, G.H. (1997). Increasing returns, trade and the regional structure of wages, *Economic Journal* 107, 113-133.

Hanson, G.H. (1998). Market potential, increasing returns, and geographic concentration, *NBER Working Papers*, 6429.

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

Harley, C.K. (1988). Ocean freight rates and productivity, 1740-1913: the primacy of mechanical invention reaffirmed, *Journal of Economic History* 48(4), 851-876.

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(4), 315-348.

Hatton, T.J. and Williamson, J.G. (1998). *The age of mass migration: causes and economic impact*, Oxford University Press, NY.

Head, C.K., Ries, J.C. and Swenson, D.L. (1999). Attracting foreign manufacturing: investment promotion and agglomeration, *Regional Science and Urban Economics* 29(2), 197-218.

Head, K. and Mayer, T. (2004a). Market potential and the location of Japanese investment in the European Union, *Review of Economics and Statistics* 86(4), 959-972.

Head, K. and Mayer, T. (2004b). The empirics of agglomeration and trade. In Henderson and Thisse (eds.), *Handbook of Regional and Urban Economics*, vol.4, Elsevier, Amsterdam, 2609-2669.

Head, K. and Mayer, T. (2006). Regional wage and employment responses to market potential in the EU, *Regional Science and Urban Economics* 36(5), 573-594.

Helpman, E. (1998). The size of regions. In Pines, Sadka and Zilcha (eds.), *Topics in public economics*, Cambridge University Press, New York, 33-54.

Helpman, E. and Krugman, P. (1985). *Market structure and foreign trade*, The MIT Press, Cambridge (MA).

Henderson, J.V. and Kuncoro, A. (1996). Industrial centralization in Indonesia, *World Bank Economic Review* 10(3), 513-540.

Hering, L. and Paillacar, R. (2008). On the relation between market access, migration and wages: an empirical analysis, *European University Institute*, Florence.

Hering, L. and Paillacar, R. (2009). How does market access shape internal migration, *Université de Paris 1*.

Hering, L. and Poncet, S. (2006). Market access impact on individual wage: evidence from China, *CEPII Working Papers*, 2006-23.

Herranz, A. (2001). Una aproximación al stock de infraestructuras español (1844-1935), *Revista de Historia Económica* 19, 129-152.

Herranz, A. (2002). *Infrastructure and economic growth in Spain, 1845-1935*. PhD dissertation (unpublished), London School of Economics.

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 del ICE* 70, 183-203.

Herranz, A. (2007a). Infrastructure investment and Spanish economic growth, 1850–1935, *Explorations in Economic History* 44(3), 452-468.

Herranz, A. (2007b). The spatial distribution of Spanish transport infrastructure between 1860 and 1930, *Annals of Regional Science* 41(1), 189-208.

Herranz, A. (2008). *Infraestructuras y crecimiento económico en España (1850-1935)*, Fundación de los ferrocarriles españoles, Madrid.

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

Hotelling, H. (1929). Stability in competition, *Economic Journal* 39, 41-57.

Hummels, D. (1999). Toward a geography of trade costs, *GTAP Working Papers*, 1162.

Hummels, D. (2001). Time as a trade barrier, *GTAP Working Papers*, 1152.

Instituto Geográfico y Estadístico (1903). *Estadística de la emigración e inmigración de España, 1896-1900*. Ministerio de Instrucción Pública y Bellas Artes, Madrid.

Instituto Nacional de Estadística (INE) – National Statistics Institute, www.ine.es.

Institut National de la Statistique et des Études Économiques (INSEE) – National Institute of Statistics and Economic Studies, www.insee.fr.

Isserlis, L. (1938). Tramp shipping cargoes and freights, *Journal of the Royal Statistical Society* 101(1), 53-146.

Jacks, D.S., Meissner, C.M. and Novy, D. (2009). Trade booms, trade busts, and trade costs, *NBER Working Papers*, 15267.

Jacks, D.S. and Pendakur, K. (2008). Global trade and the maritime transport revolution, *NBER Working Papers*, 14139.

Jiménez Blanco, J.I. (1986). Introducción. In Garrabou, Barciela and Jiménez Blanco (eds.), *Historia agraria de la España contemporánea*, vol.3, Crítica, Barcelona, 9-141.

Kaldor, N. (1935). Market imperfection and excess capacity, *Economica* 2(5), 33-50.

Kancs, D. (2005). Can we use NEG models to predict migration flows? An example of CEE accession countries, *Migration Letters* 2(1), 32-63.

Kancs, D. (2010). The economic geography of labour migration: competition, competitiveness and development, *Applied Geography* (forthcoming).

Kaukiainen, Y. (2003). How the price of distance declined: ocean freights for grain and coal from the 1870s to 2000, *University of Helsinki*.

Keeble, D., Offord, J. and Walker, S. (1988). *Peripheral regions in a Community of twelve*, Commission of the European Communities, Brussels/Luxembourg.

Keeble, D., Owens, P. and Thompson, C. (1982). Regional accessibility and economic potential in the European community, *Regional Studies* 16(6), 419-432.

Kim, S. (1995). Expansion of markets and the geographic distribution of economic activities: the trends in U.S. regional manufacturing structure, 1860-1987, *Quarterly Journal of Economics* 110(4), 881-908.

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

Kim, S. (1999). Regions, resources, and economic geography: sources of U.S. regional comparative advantage, 1880–1987, *Regional Science and Urban Economics* 29(1), 1-32.

Kim, S. and Margo, R.A. (2004), Historical perspectives on U.S. economic geography. In Henderson and Thisse (eds.), *Handbook of Regional and Urban Economics*, vol.4, Elsevier, Amsterdam, 2981-3019.

Kiso, T. (2005). Does New Economic Geography explain the spatial distribution of wages in Japan?, *University of Tokyo*.

Klein, A. and Crafts, N. (2009). Making sense of the manufacturing belt: determinants of U.S. industrial location, 1880-1920, *University of Warwick*.

Knaap, T. (2006). Trade, location, and wages in the United States, *Regional Science and Urban Economics* 36(5), 595-612.

Kolko, J. (2007). Agglomeration and co-agglomeration of services industries, *MPRA Paper*, 3362.

Krugman, P. (1979). Increasing returns, monopolistic competition, and international trade, *Journal of International Economics* 9(4), 469-479.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade, *American Economic Review* 70(5), 950-959.

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

Krugman, P. (1992a). *Geography and trade*, The MIT Press, Cambridge (MA).

Krugman, P. (1992b). A dynamic spatial model, *NBER Working Papers*, 4219.

Krugman, P. (1993). First nature, second nature, and metropolitan location, *Journal of Regional Science* 33(2), 129-144.

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

Krugman, P. and Venables, A.J. (1995). Globalization and the inequality of nations, *Quarterly Journal of Economics* 110(4), 857-880.

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

Kuznets, S. (1966). *Modern economic growth: rate, structure, and spread*, Yale University Press, New Haven (CT).

Kuznets, S. (1971). *Economic growth of nations: total output and production structure*, Harvard University Press, Cambridge (MA).

Lafourcade, M. and Thisse, J.F. (2008). New economic geography: a guide to transport analysis, *PSE Working Papers*, 2008-02.

Lanaspa, L., Pueyo, F. and Sanz, F. (2003). The evolution of Spanish urban structure during the twentieth century, *Urban Studies* 40(3), 567-580.

League of Nations (1927). *Tariff level indices*, International Economic Conference, Geneva.

Liepmann, H. (1938). *Tariff levels and the economic unity of Europe: an examination of tariff policy, export movements and the economic integration of Europe, 1913-1931*, George Allen and Unwin, London.

Llopis, E. (2001). El legado económico del antiguo régimen desde la óptica regional. In Germán, Llopis, Maluquer and Zapata (eds.), *Historia económica regional de España, siglos XIX y XX*, Crítica, Barcelona, 507-524.

Llopis, E. and Sotoca, S. (2005). Antes, bastante antes: la primera fase de la integración del mercado español de trigo, 1725-1808, *Historia Agraria* 36, 225-262.

López-Córdova, J.E. and Meissner, C. (2003). Exchange-rate regimes and international trade: evidence from the classical gold standard era, *American Economic Review* 93(1), 344-353.

Lösch, A. (1940). *Die räumliche ordnung der wirtschaft*, Gustav Fischer, Jena (English translation, 1954: *The economics of location*, Yale University Press, New Haven).

Lozano, C. (2007). *Ideología, política y realidad económica en la formación profesional industrial española (1857-1936)*, Milenio, Lleida.

Lucas, R.E. (1988). On the mechanics of economic development, *Journal of Monetary Economics* 22(1), 3-42.

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

Luna, G. (1988). La población urbana en España, 1860-1930, *Boletín de la Asociación de Demografía Histórica* 6(1), 25-68.

Madrazo, S. (1984). *El sistema de transportes en España, 1750-1850*, Turner, Madrid.

Magrini, S. (2004). Regional (di)convergence. In Henderson and Thisse (eds.), *Handbook of Regional and Urban Economics*, vol.4, Elsevier, Amsterdam, 2741-2796.

Maixé-Altés, J.C. and 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(3), 496-521.

Maluquer, J. (1994). El índice de la producción industrial de Cataluña. Una nueva estimación (1817-1935), *Revista de Historia Industrial* 5, 45-71.

Mankiw, G., Romer, D. and Weil, D. (1992). A contribution to the empirics of economic growth, *Quarterly Journal of Economics* 107(2), 407-437.

Marshall, A. (1890). *Principles of economics*, Macmillan, London.

Martín Aceña, P. and Pons, M.Á. (2005). Sistema monetario y financiero. In Carreras and Tafunell (eds.), *Estadísticas históricas de España, siglos XIX-XX*, vol.II, Fundación BBVA, Bilbao, 645-706.

Martín, M. (1992). Pautas y tendencias de desarrollo económico regional en España: una visión retrospectiva. In García Delgado and Pedreño (eds.), *Ejes territoriales de desarrollo: España en la Europa de los noventa*, Madrid, 133-155.

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. (2009). *La producción industrial en el País Valenciano (1861-1920). Tirando del 'hilo industrial'*, Milenio, Lleida.

Martínez-Galarraga, J., Paluzie, E., Pons, J. and Tirado, D.A. (2008). Agglomeration and labour productivity in Spain over the long term, *Cliometrica* 2(3), 195-212.

Martorell, M. (2001). *Historia de la peseta: la España contemporánea a través de su moneda*, Planeta, Barcelona.

Mas, M., Maudos, J., Pérez, F. and Uriel, E. (1994). Disparidades regionales y convergencia en las comunidades autónomas, *Revista de Economía Aplicada* 4, 129-148.

Matilla, M., Pérez, P. and Sanz, B. (2008). La integración del mercado español a finales del siglo XIX: los precios del trigo entre 1891 y 1905, *Documentos de Trabajo de Funcas*, 391/2008.

Mayer, T. (2008). Market potential and economic development, *CEPR Discussion Papers*, 6798.

Midelfart-Knarvik, K.H., Overman, H., Redding, S. and Venables, A.J. (2002). The location of European industry. In European Commission (ed.), *European economy. European integration and the functioning of product markets*, 2/2002. 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 Papers*, 2618.

Ministerio de Fomento (1856). *Memoria sobre el estado de las obras públicas en España en 1856*, Dirección General de Obras Públicas, Madrid.

Ministerio de Hacienda (1867). *Información sobre el derecho diferencial de bandera y sobre los de aduanas exigibles á los hierros, el carbón de piedra y los algodones, presentada al gobierno de su majestad por la comisión nombrada al efecto en real decreto de 10 de noviembre de 1865*, Madrid.

Ministerio de Obras Públicas (1902). *Estadísticas de obras públicas, años 1899 y 1900*, Dirección General de Obras Públicas, Madrid.

Ministerio de Trabajo, Comercio e Industria (1927). *Estadística de los salarios y jornadas de trabajo referida al período 1914-25*, Dirección General de Trabajo y Acción Social, Madrid.

Mion, G. (2004). Spatial externalities and empirical analysis: the case of Italy, *Journal of Urban Economics* 56(1), 97-118.

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.

Mitchener, K.J. and Weidenmier, M. (2008). Trade and empire, *NBER Working Papers*, 13765.

Mohammed, S.I.S. and Williamson, J.G. (2004). Freight rates and productivity gains in British tramp shipping 1869–1950, *Explorations in Economic History* 41(2), 172-203.

Monfort, P. and Nicolini, R. (2000). Regional convergence and international integration, *Journal of Urban Economics* 48(2), 286-306.

Moral Ruiz, J. (1979). *La agricultura española a mediados del siglo XIX, 1850-1870*, Ministerio de Agricultura, Madrid.

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

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, Carreras and Sudrià (eds.), *La economía española en el siglo XX. Una perspectiva histórica*, Ariel, Barcelona, 23-61.

Nadal, J. and Carreras, A. (1990). *Pautas regionales de la industrialización española (siglos XIX y XX)*, Ariel, Barcelona.

Nadal, J. and Tafunell, X. (1992). *Sant martí de Provençals, pulmó industrial de Barcelona (1847-1992),* Columna, Barcelona.

Neven, D. and Gouyette, C. (1995). Regional convergence in the European Community, *Journal of Common Market Studies* 33(1), 47-65.

Nicolau, R. (2005). Población, salud y actividad. In Carreras and Tafunell (eds.), *Estadísticas históricas de España: siglos XIX-XX*, Fundación BBVA, Bilbao, 77-154.

Niebuhr, A. (2006). Market access and regional disparities, *Annals of Regional Science* 40(2), 313-334.

North, D. (1958). Ocean freight rates and economic development 1750-1913, *Journal of Economic History* 18(4), 537-555.

Núñez, C.E. (1992). *La fuente de la riqueza: educación y desarrollo económico en la España contemporánea*, Alianza, Madrid.

Nunn, N. and Puga, D. (2009). Ruggedness: the blessing of bad geography in Africa, *NBER Working Papers*, 14918.

O'Rourke, K.H. (2000). Tariffs and growth in the late 19th century, *Economic Journal* 110, 456-483.

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

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

Ottaviano, G.I.P., Tabuchi, T. and Thisse, J.F. (2002). Agglomeration and trade revisited, *International Economic Review* 43(2), 409-436.

Paluzie, E. (2001). Trade policy and regional inequalities, *Papers in Regional Science* 80(1), 67-85.

Paluzie, E., Pons, J., Silvestre, J. and Tirado, D.A. (2009). Migrants and market potential in Spain over the twentieth century: a test of the new economic geography, *Spanish Economic Review* 11(4), 243-265.

Paluzie, E., Pons, J. and Tirado, D.A. (2004). The geographical concentration of industry across Spanish regions, 1856-1995, *Jahrbuch Für Regionalwissenschaft* 24(2), 143-160.

Paluzie, E., Pons, J. and Tirado, D.A. (2005). Potencial de mercado y estructura espacial de los salarios en las regiones españolas, 1955-1995, *Cuadernos Aragoneses de Economía* 15, 17-37.

Paluzie, E., Pons, J. and Tirado, D.A. (2007). Aglomeración y productividad del trabajo en las regiones españolas. In Reig (ed.), *Competitividad, crecimiento y capitalización de las regiones españolas*, Fundación BBVA, Bilbao, 249-276.

Paluzie, E., Pons, J. and Tirado, D.A. (2009). A test of the market potential equation in Spain, *Applied Economics* 41(12), 1487-1493.

Parejo, A. (1997). *La producción industrial de Andalucía, 1830-1935*, Instituto de Desarrollo Regional, Sevilla.

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.

Parejo, A. (2004). La industrialización de las regiones españolas durante la primera y la segunda revolución tecnológica. Andalucía, Cataluña, País Vasco (1830-1975), *Revista de Historia Económica* 22(3), 669-706.

Pascual, P. (1990). *Agricultura i industrialització a la Catalunya del segle XIX. Formació i desestructuració d'un sistema econòmic*, Crítica, Barcelona.

Peña, D. and Sánchez-Albornoz, N. (1983). *Dependencia dinámica entre precios agrícolas. El trigo en España, 1857-1890. Un estudio empírico*, Servicio de Estudios de Historia Económica, Banco de España, Madrid.

Peña, D. and 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(3), 341-345.

Picard, P. and Zeng, D.Z. (2005). Agricultural sector and industrial agglomeration, *Journal of Development Economics* 77(1), 75-106.

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

Pons, J., Paluzie, E., Silvestre, J. and Tirado, D.A. (2007). Testing the new economic geography: migrations and industrial agglomerations in Spain, *Journal of Regional Science* 47(2), 289-313.

Pons, J. and Tirado, D.A. (2006). The determinants of regional economic inequality in Spain in the 20th century. Who is right: Sachs or Krugman?, *Universitat de Barcelona*.

Pons, J. and Tirado, D.A. (2008). Los determinantes de la desigualdad económica regional en España, *ICE. Tribuna de Economía* 842, 195-216.

Prados de la Escosura, L. (1982). *Comercio exterior y crecimiento económico en España, 1826-1913: tendencias a largo plazo*, Banco de España, Madrid.

Prados de la Escosura, L. (1988). *De imperio a nación: crecimiento y atraso económico en España, 1780-1930*, Alianza, Madrid.

Prados de la Escosura, L. (2000). International comparisons of real product, 1820–1990: an alternative data set, *Explorations in Economic History* 37(1), 1-41.

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

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

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

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

Quah, D. (1996). Regional convergence clusters across Europe, *European Economic Review* 40(3-5), 951-958.

Quah, D. (1997). Empirics for growth and distribution: stratification, polarization, and convergence clubs, *Journal of Economic Growth* 2(1), 27-59.

Rappaport, J. and Sachs, J.D. (2003). The United States as a coastal nation, *Journal of Economic Growth* 8(1), 5-46.

Raymond, J.L. and García-Greciano, B. (1994). Las disparidades en el PIB per cápita entre comunidades autónomas y la hipótesis de convergencia, *Papeles de Economía Española* 59, 37-58.

Redding, S. (2010). The empirics of new economic geography, *Journal of Regional Science* 50(1), 297-311.

Redding, S. and Venables, A.J. (2004). Economic geography and international inequality, *Journal of International Economics* 62(1), 53-82.

Reher, D. (1994). Ciudades, procesos de urbanización y sistemas urbanos en la Península Ibérica, 1550-1991. In Guàrdia, Monclús and Oyón (eds.), *Atlas histórico de las ciudades europeas. Península ibérica*, Centre de Cultura Contemporània de Barcelona-Salvat, Barcelona, 1-30.

Reher, D. (2001). Producción, precios e integración de los mercados regionales de grano en la España preindustrial, *Revista de Historia Económica* 19(3), 539-572.

Rich, D. (1980). Potential models in human geography, *Geoabstracts* 26.

Ringrose, D.R. (1970). *Transportation and economic stagnation in Spain: 1750-1850*, Duke University Press, Durham (NC).

Rodríguez-Pose, A. (1999). Convergence or divergence? Types of regional responses to socio-economic change in Western Europe, *Tijdschrift voor Economische en Sociale Geografie* 90(4), 365-378.

Romer, P.M. (1986). Increasing returns and long-run growth, *Journal of Political Economy* 94(5), 1002-1037.

Roos, M. (2001). Wages and market potential in Germany, *Jahrbuch für Regionalwissenschaft* 21(2), 171-195.

Roos, M. (2005). How important is geography for agglomeration?, *Journal of Economic Geography* 5(5), 605-620.

Rosés, J.R. (2003). Why isn't the whole of Spain industrialized? New economic geography and early industrialization, 1797-1910, *Journal of Economic History* 63(4), 995-1022.

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

Rosés, J.R. and Sánchez-Alonso, B. (2004). Regional wage convergence in Spain 1850–1930, *Explorations in Economic History* 41(4), 404-425.

Sabaté, M., Gadea, M.D. and Serrano, J.M. (2001). Cambio, precios, balanza y expectativas. In Sudrià and Tirado (eds.), *Peseta y protección. Comercio exterior, moneda y crecimiento económico en la España de la Restauración*, Edicions de la Universitat de Barcelona, Barcelona, 143-176.

Sachs, J.D. (2000). Tropical underdevelopment, *CID Working Paper*, 57.

Sachs, J.D. and Warner, A.M. (2001). The curse of natural resources, *European Economic Review* 45(4-6), 827-838.

Sáiz, J.P. (2005). Investigación y desarrollo: patentes. In Carreras and Tafunell (eds.), *Estadísticas históricas de España, siglos XIX-XX*, vol.II, Fundación BBVA, Bilbao, 835-872.

Sala-i-Martin, X. (1996). Regional cohesion: evidence and theories of regional growth and convergence, *European Economic Review* 40(6), 1325-1352.

Samuelson, P.A. (1954). The transfer problem and transport costs. II. Analysis of effects of trade impediments, *Economic Journal* 64, 264-289.

Sánchez-Albornoz, N. (1968). *España hace un siglo: una economía dual*, Península, Madrid.

Sánchez-Albornoz, N. (1975). *Los precios agrícolas durante la segunda mitad del siglo XIX*, Banco de España, Madrid.

Sánchez-Albornoz, N. (1987). *The economic modernization of Spain, 1830-1930*, New York University Press, New York.

Sánchez-Alonso, B. (1995). *Las causas de la emigración española, 1880-1913*, Alianza, 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(3), 730-755.

Sardà Dexeus, J. (1948). *La política monetaria y las fluctuaciones de la economía española en el siglo XIX*, CSIC, Madrid.

Schulze, M.S. (2007). Regional income dispersion and market potential in the late nineteenth century Hapsburg Empire, *LSE Working Papers*, 106/07.

Serrano, J.M., Gadea, M.D. and Sabaté, M. (1998). Tipo de cambio y protección. La peseta al margen del patrón oro, 1883-1931, *Revista de Historia Industrial* 13, 83-112.

Silvestre, J. (2003). *Migraciones interiores y mercado de trabajo en España, 1877-1936*. PhD dissertation (unpublished), Universidad de Zaragoza.

Silvestre, J. (2005). Internal migrations in Spain, 1877-1930, *European Review of Economic History* 9(2), 233-265.

Silvestre, J. (2007). Temporary internal migrations in Spain, 1860-1930, *Social Science History* 31(4), 539-574.

Simpson, J. (1995a). *Spanish agriculture: the long siesta, 1765-1965*, Cambridge University Press, Cambridge (MA).

Solow, R.M. (1956). A contribution to the theory of economic growth, *Quarterly Journal of Economics* 70(1), 65-94.

Starret, D. (1978). Market allocations of location choice in a model with free mobility, *Journal of Economic Theory* 17(1), 21-37.

Stock, J.H. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In Andrews and Stock (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 and Marichal (eds.), *La formación de los bancos centrales en España y América Latina (siglos XIX y XX)*, Banco de España, Madrid, 81-107.

Tabuchi, T. and Thisse, J.F. (2002). Taste heterogeneity, labor mobility and economic geography, *Journal of Development Economics* 69(1), 155-177.

Tedde de Lorca, P. (1994). Cambio institucional y cambio económico en la España del siglo XIX, *Revista de Historia Económica* 12(3), 525-538.

Temple, J. (1999). The new growth evidence, *Journal of Economic Literature* 37(1), 112-156.

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

Tirado, D.A., Paluzie, E. and Pons, J. (2002). Economic integration and industrial location: the case of Spain before World War I, *Journal of Economic Geography* 2(3), 343-363.

Tirado, D.A., Pons, J. and Paluzie, E. (2003). Industrial agglomerations and wage gradients: the Spanish economy in the interwar period, *Universitat de Barcelona Working Papers in Economics*, 01/03.

Tirado, D.A., Pons, J. and 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.

Tirado, D.A., Pons, J. and Paluzie, E. (2009). Trade policy and wage gradients: evidence from a protectionist turn, *CSGR Working Paper*, 258/09.

Tortella, G. (1970). La evolución del sistema financiero español de 1855 a 1868. In Schwartz (ed.), *Ensayos sobre la economía española a mediados del siglo XIX*, Banco de España, Madrid, 17-174.

Tortella, G. (1973). *Los orígenes del capitalismo en España. Banca, industria y ferrocarriles en el siglo XIX*, Tecnos, Madrid.

Tortella, G. (1981). La economía española, 1830-1900. In Tuñón de Lara (ed.), *Historia de España*, vol.8, Labor, Barcelona, 9-167.

Tortella, G. (1994). *El desarrollo de la España contemporánea*, Alianza, Madrid.

Tortella, G. (1999). Introducción. La paradoja del ferrocarril español. In Muñoz, Sanz Fernández and Vidal (eds.), *Siglo y medio del ferrocarril en España, 1848-1898. Economía, industria y sociedad*, Fundación de los Ferrocarriles Españoles, Madrid, 249-253.

Venables, A.J. (1996). Equilibrium locations of vertically linked industries, *International Economic Review* 37(2), 341-359.

Venables, A.J. (2005). Spatial disparities in developing countries: cities, regions, and international trade, *Journal of Economic Geography* 5(1), 3-21.

Vicens Vives, J. (1959). *Manual de historia económica de España*, Teide, Barcelona.

Visecretaría Nacional de Ordenación Económica (1962), *Tablas input-output de la economía española: año de 1958*, Organización Sindical Española, Madrid.

Vickerman, R., Spiekermann, K. and Wegener, M. (1999). Accessibility and economic development in Europe, *Regional Studies* 33(1), 1-15.

Von Thünen, J.H. (1826). *Der isolierte staat in beziehung auf landtschaft und nationalökonomie*, Perthes, Hamburgo (English translation, 1966: *Von Thünen's isolated state*, Pergamon Press, Oxford).

Wais, F. (1987). *Historia de los ferrocarriles españoles*, Fundación de los Ferrocarriles Españoles, Madrid.

Weber, A. (1909). *Über den standort der industrien*, Mohr, Tübingen (English translation, 1929: *The theory of the location of industries*, University of Chicago Press).

Williamson, J.G. (1965). Regional inequality and the process of national development: a description of the patterns, *Economic Development and Cultural Change* 13:4, Part II, July, 3-84.

Williamson, J.G. (2006). Explaining world tariffs 1870-1938: Stolper-Samuelson, strategic tariffs and state revenues. In Findlay, Henriksson, Lindgren and Lundahl (eds.), *Eli Heckscher, international trade, and economic history*, The MIT Press, Cambridge (MA), 199-228.

Wolf, N. (2004). Endowments, market potential, and industrial location: evidence from interwar Poland (1918-1939), *CEP Working Paper*, dp0609.

Wolf, N. (2007). Endowments vs. market potential: what explains the relocation of industry after the Polish reunification in 1918?, *Explorations in Economic History* 44(1), 22-42.

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

Yamamoto, D. (2008). Scales of regional income disparities in the United States, 1955-2003, *Journal of Economic Geography* 8(1), 79-103.