LUND UNIVERSITY
School of Economics and Management

Master in Economic History

The Determinants of Industrial Location in Sweden (1913-1950)

Antonio Rodríguez Merello
an4550ro-s@student.lu.se

Abstract: This thesis analyses regional industrial concentration and the determinants of industrial location in Sweden during the first half of the twentieth century. First, it is identified how geographical concentration of industry evolved in this country. Second, by estimating an economic geography model, the forces that were behind industrial location are captured. The inclusion of both procedures provides a good picture of the industrialisation phenomenon in Sweden since both present a certain degree of interrelationship. The results show that in none of the Swedish industries there was a significant degree of concentration at the regional level. Moreover, the evidence provided by the model shows that traditional forces were not operating in the process of industrial location. These findings show that Sweden represents a special case when considering the impact of industrialisation in spatial terms.

Key Words: New economic geography, Heckscher-Ohlin, industrial location

EKHS11
Master's Thesis, First Year (15 Credits ECTS)
June 2017
Supervisor: Erik Bengtsson
Examiner:
Word Count: 10195

Table of Contents

1. Introduction

This thesis explores the determinants of industrial location in Sweden during the first half of the twentieth century. The analysis of this phenomenon allows identifying the key factors that were behind this development. Although there are articles that have analysed industrialisation patterns in Sweden (Enflo, Henning and Andersson, 2011; Berger, Enflo and Henning, 2012), these papers are primarily focused on the twentieth century and on the evolution of both regional convergence as well as divergence. Therefore, this thesis provides new evidence to the discussion related to the evolution of spatial inequalities in Sweden. This is important since research in economic geography is one of the information sources when conducting regional policy. The empirical part of this thesis consists on the measuring of industrial concentration and the estimation of a model based on Midelfart-Knarvik, Overman, Redding and Venables (2000). The implementation of this strategy allows identifying relevant aspects related to this issue by using an approach which differs from what others have contributed to the discussion of Swedish industrialisation. Also, by using this model it is possible to contrast the two main theories that are in the economic literature which are the New Economic Geography (hereafter NEG) and Heckscher-Ohlin theoretical framework. In essence, by considering these two views it is possible to understand how industries located and how the interplay of the forces belonging to these two views changed. Since history is a dynamic process, the aim is to capture how the relative influence of these two approaches evolved during the period observed.

The results provided by the empirical part provide the following conclusions. First, it should be noted that the evolution of industrial concentration in Sweden was not strong, especially when is compared to other Western economies. As a matter of fact, during the first half of the twentieth century none of the industries analysed were concentrated. Second, the findings provided by the model do not show evidence for neither NEG nor Heckscher-Ohlin forces operating in the process of industrial location in Sweden. Therefore, the findings provided in this thesis coincide with others provided in the literature. This means that Sweden represents a special case in economic history when analysing industrialisation and its impact on a regional basis.

2. Theoretical Framework and Background Literature

The analysis of industrialisation processes and their repercussions in terms of spatial distribution provides two major theories. New Economic Geogrpahy (hereafter NEG) theory argues that the phenomenon of industrial location is primarily due to the influence of agglomeration economies which have a pull for centrality (Combes, Mayer and Thisse, 2008, pp. 42). Conversely, the Heckscher-Ohlin framework provides an alternative explanation in which factor endowments are the main determinant behind this phenomenon.

The Heckscher-Ohlin model was proposed by Swedish economists Eli Heckscher and Bertil Ohlin in the 1930s. This model provides an extension to the Ricardian trade theory which is based on the concept of comparative advantage (Krugman, Obstfeld and Melitz, 2012, pp. 80). According to the principles of the Ricardian model, comparative advantage reflects the competitive features that countries have with respect to others in terms of producing certain goods[1]. By following the same reasoning, in the Heckscher-Ohlin model, comparative advantage is determined by the relative factor endowments of countries. Factor endowments indicate the availability of resources that a certain territory has and which are necessary for the production of goods and services. This concept primarily reflects natural resources, human capital and technology. Therefore, by following the reasoning of this model, a country in which there is a relative abundance in iron ore will specialise in those economic activities that use this mineral intensively. Economic research has found evidence for this assertion. For instance, Crafts and Mulatu (2004) found out that coal abundance had a key role behind industrial location during the first industrial revolution in Great Britain.

Proponents of the NEG theory argue that industries based their choices on location according to the economic attractiveness of areas (Combes, Mayer and Thisse, 2008, pp. 302). Therefore, this view emphasises the role of market potential and agglomeration economies as the main factors behind industrial location. The relevance of market potential is based on its capability to attract firms. In general terms, if the size of a market is big, this usually entails a higher availability of suppliers of intermediate goods and also lower costs for these inputs. In the economic literature this concept is known as backward and forward linkages[2]. Moreover, in core areas, there is a higher likelihood of

[1] In the Ricardian model, comparative advantage is determined by labour requirements. This means that if a country needs a lower amount of workers to produce a certain good than other country, it has a comparative advantage, and as a result, it will specialise in this good ceteris paribus.

[2] Backward linkages are identified as the connection between firms and their suppliers while forward linkages reflect the relation between firms and their market (Combes, Mayer and Thisse, 2008, pp. 131). With regards to backward linkages, it should be noted that those firms in which intermediate goods represent an important share of their total costs are the most benefitted from moving to core economic locations.

benefitting from scale economies. Therefore, this is one of the causes behind the phenomenon of circular causality (Myrdal, 1971, pp. 13), which is the base of agglomeration economies. Circular causality occurs when there is an interrelationship between two or more factors and the existence of one reinforces the other and vice versa. In essence, agglomeration economies are based on the pull of centrality of economic activities from core locations.

In the NEG theoretical framework, it is also emphasised the impact of a set of transformations that occurred with the emergence of the industrialisation process. Among them, the decrease in transportation costs is one of the main consequences. According to the findings of Crafts and Mulatu, (2004), technological innovations such as the invention of railroads had a strong impact on the reduction of transportation costs. For instance, in Great Britain the cost of coastal shipping decreased by 47% from 1872 to 1911, while railroad costs per ton decreased by 22% during the same period. This is particularly important since transportation costs determine the incentives to relocate to core economic areas, especially for those industries which benefit more from scale economies. Therefore, the decline in transportation costs usually entails an increase in the concentration of economic activity due to market potential and scale economies. However, when transportation costs fall to a certain threshold, the relative advantage disappears. According to Krugman and Venables (1995), there is an optimum level of transportation costs that triggers divergence in relation to industrial location. However, as the decline in transportation costs continues, the relative gains derived from moving to core economic areas also decreases. This is caused by an increase in congestion that diminishes the benefits derived from agglomeration economies (Tabuchi, 1998). Since incentives to move to central locations become weaker, economic activity starts to move to peripheral areas and relative spatial inequality starts to decline. Moreover, it should be noted that these patterns of transportation costs and its influence on the location of economic activities has occurred in Western economies such as the United States or France (Kim, 1995; Combes Lafourcade, Thisse and Toutain, 2011).

In the analysis of spatial inequalities in a country, attention should be focused on the evolution of both regional convergence and divergence. Regional divergence occurs when there is an increase in the difference between regions in economic terms. These differences are measured by using economic indicators such as gross domestic product (hereafter GDP) per capita or labour productivity. Conversely, regional convergence takes place when the poorest regions grow at a faster rate than the richest, which allows the former converging to the latter. Moreover, it is also relevant the interrelationship between the location of economic activities and the relative impact on regional convergence and/or divergence. In the NEG literature it is argued that the increase in economic integration that took place in the industrialisation phase resulted in the growth of divergence across regions (Rosés, Galarraga and Tirado, 2010). Also as mentioned earlier, the existence of agglomeration economies in certain industries may reinforce the spatial concentration of economic activities which increase the economic differences between regions. By estimating GDP and labour productivity, Geary and Stark (2002) analysed regional convergence process in the British Isles during the late nineteenth century and early twentieth century. The results provided two main conclusions. First, it is possible to conduct reliable GDP estimations by using the information provided by wages across sectors. Second, the Irish economy underwent into a phase of convergence with respect to the English economy during the period analysed.

In Sweden, the evolution of regional inequalities followed a different trend than in other Western economies. Since to a certain extant this was due to the evolution of the industrialisation process, it is important to consider the evolution of this phenomenon as well as its particularities. For instance, Berger, Enflo and Henning (2012), argue that industrialisation was primarily a rural phenomenon until the mid-twentieth century. Moreover, according to Magnusson (2002, pp. 54), in the first industrial revolution, textiles represented an exception since it was the first activity that used mass-production and the most advanced technology. Therefore, since many industries did not experience full mechanisation until the 1870s, this could have influenced the lack of pull in certain sectors, at least during the first phase of industrialisation. In contrast to the previous period, the second industrial revolution arrived in Sweden at roughly the same period as in the advanced industrialised nations. This was particularly reflected by the strong expansion that took place in industry from the 1870s onwards (Magnusson, 2002, pp. 56). At the same time, in the early twentieth century there were forces pulling for both concentration and dispersion (Berger, Enflo and Henning, 2012). For instance, activities sensitive to scale economies became more concentrated at a regional level. Specifically, this took place in metals and machinery and chemical industries. Conversely, those activities which were based on low wages, such as textiles, relocated to the areas which were more attractive in this regard.

When comparing the Swedish case with respect to other countries, it should be noted the particularities of this country. For instance, while in countries such as Spain, the level of inequality was showing a marked increase during the last third of the nineteenth century (Martínez-Galarraga, 2011). The relative level of inequality did not significantly grow in Sweden during this period and it was succeeded by a convergence phase (Henning, Enflo and Andersson, 2011). Moreover, in Berger, Enflo and Henning (2012) it is argued that the initial phase of divergence had a special impact on the northern regions of Sweden. The methodology employed is primarily based on the use of location quotients which measure the relative share of manufacturing employment with respect to the population of an area. The results show the existence of bell-shaped pattern in big cities during the first half of the twentieth century. This means that during the first three decades of the twentieth century, the share of industrial activities located in big cities increased and it was then followed by a decrease. Conversely, in medium-sized cities, this share experienced a decline during the majority of the period observed.

With regards to the regional convergence process in Sweden, Enflo, Henning and Schön (2015) provided an estimation of Swedish regional GDP by using an approach based on Geary and Stark (2002).This approach is primarily based on a GDP estimation which calculates output per worker at all economic sectors. The results show that, since the start of industrialisation in Sweden, the general trend has been directed towards a process of homogenisation across regions, especially in terms of per capita value added. However, the process of convergence has undergone into several phases with different intensities. This has been primarily due to the effects derived from the different migration waves, both internal and external, and the process of economic integration that started during the late eighteenth century.

According to Enflo and Rosés (2015), emigration during the second half of the nineteenth century was particularly strong in those areas in which labour productivity in agriculture was lower. By using the Theil decomposition, they measured the aggregate components of GDP. The results show that migration contributed to improve the relative efficiency of backward areas. Therefore, this contributed to the improvement of the relative efficiency in these areas and, as a result, this allowed them to catch up with the areas with higher productivity levels. Although migration, especially external, reached its peak during this era, it increased again during the second half of the twentieth century. Moreover, the interplay between migration and convergence between industries allowed that the most backward regions could catch-up during this phase. It is important to consider that, at this period, the Swedish economy had started to experience process of structural change in which services became into the main economic activity in terms of population employed. Nevertheless, providing public incentives for internal migration became unpopular and the government stance changed towards the implementation of policies aimed to improve the economic situation of declining areas (Enflo and Rosés, 2015).

Although the general trend in Sweden has been characterised by a process of increasing homogenisation, it seems that this trend changed during the late twentieth century. The main reason behind this change is the emergence of the information and communications technology (hereafter ICT) development block. After the introduction of this new set of technologies, the spillovers derived from its generalisation have a stronger impact in big metropolitan areas than in peripheral areas. One of the reasons that explain this divergence is that core locations are generally better endowed in human capital. Similarly, if the case of France is considered, it is argued that higher education has become into one of the main agglomeration forces during the last decades of the twentieth century (Combes, Lafourcade, Thisse and Toutain, 2011).

As mentioned before, NEG emphasises the implications between transportation costs and the process of industrial location. There are articles that show that this happened in Western economies (Crafts and Mulatu, 2004; Combes, Lafourcade, Thisse and Toutain, 2011). For instance, in the case of Spain, Martínez-Galarraga (2011) argues that, the expansion of the railroad network in the second half of the nineteenth century caused a reduction in transportation costs. As a result, the spatial concentration of industry increased notably in this country. In the literature it is also mentioned that the evolution of spatial inequalities follows a bell-shaped curve in which the concentration of economic activities increases until a certain threshold, and then, it decreases again. In order to test the existence of this pattern, it is common to use the Gini coefficient or the Theil index. For instance, Brülhart and Torstensson (1998) use the Gini index in order to measure the geographical distribution of manufacturing employment in the EU. The analysis for France was done using Theil indices by sector in terms of both employment and value added (Combes Lafourcade, Thisse and Toutain, 2011). As mentioned before, in both cases the increase in spatial inequality was followed by a decrease in concentration. Regarding the first case, this was due to a decline in strength of scale effects. In the case of France, the continuous decline in transportation costs made access to market a secondary factor, and hence, this caused relocation towards peripheral areas. Moreover, in this period was also relevant the set of policies implemented by the French government which aimed to reduce inequality between regions.

With regards to the debate between NEG versus Heckscher-Ohlin forces, one of the most relevant works is Midelfart-Knarvik, Overman, Redding and Venables (2000). This article features an economic geography model (hereafter EGM) in which there are variables which represent both factor endowments and NEG forces. In essence, the main goal is capturing which were the determinants of industrial location in the European Union (hereafter EU). This is particularly interesting since the EU provides a unique example regarding economic integration between several nations. One of the most remarkable consequences is that trade flows between the different members are considerable and this has had an impact on the economic structure of the EU countries. The industries analysed cover a wide range of activities that vary in terms of technological structure (high, medium and low). Therefore, the use of this approach allowed getting a deep view of how this process developed during the late twentieth century. The results show the existence of forces that were pulling for both convergence and divergence during the period observed. For instance, backward linkages were the reason behind the increase in concentration for certain industries. Conversely, a relative weakening of medium and high increasing returns was also identified. Therefore, some industries became more concentrated while others remained unchanged or even decreased in terms of spatial concentration.

The model previously described has been adapted to several economic contexts in order to identify the determinants of industrial location during both the first and the second industrial revolution. For instance, in Martínez-Galarraga (2011) besides differing in terms of historical context, it also considers provincial rather than aggregate national data. The results show that during the first industrial revolution in Spain, factor endowments, and specifically, agricultural and labour abundance, were the main forces explaining industrial location[3]. However, the decline in transportation costs gave rise to NEG forces, and this was reflected by the relationship between market potential and economies of scale. The effects of these factors related to NEG became especially relevant with the emergence of the second industrial revolution, which took place during the first third of the twentieth century. In Spain, the NEG forces that had a stronger impact were scale effects and increasing returns. This is explained by the technological advances that happened during this phase which resulted in an increase in concentration. In the case of Britain, there are certain similarities with the previous article since factor endowments had a key influence during the first industrial revolution (Crafts and Mulatu, 2004). This was particularly true if we consider the role of access to coal and the abundance of educated workforce. With regards to NEG forces, the results show that their influence was weak since transportation costs did not decrease significantly. As a result, the effect of variables such as intermediate inputs is negligible in the model.

[3] It cannot be overlooked that during this period the majority of the manufacturing activities belonged to the foodstuffs category in Spain (Martínez-Galarraga, 2011).

3. Methodology

This section covers the main aspects regarding the empirical analysis which is carried out in this thesis. While industrial concentration is analysed by using the Herfindahl-Hirschman index (hereafter HHI), the determinants of industrial location are measured by using an EGM.

Figure 1: HHI formula

$$HHI = \sum_{i=1}^{N} s_i^2$$

In order to understand how this index works, it is suitable to explain how it is obtained. The first step consists in calculating the shares of a sector in a country. These shares are squared and the sum of these numbers provides a value which ranges from zero to one. This means that the closer the value to zero, the less concentrated an industry is, and conversely, the closer the index is to one, the higher the degree of concentration. In the economic literature it is a common practice the use of indices in order to measure the effects derived from economic integration across countries (Palan, 2010). The HHI was originally created in order to measure the relative degree of concentration of an industry (Herfindahl, 1951; Hirschman, 1964). Moreover, it has also been used in order to measure the sectoral diversity of regions as in the case of France (Combes, Lafourcade, Thisse and Toutain, 2011). In this thesis, the HHI is calculated by considering Swedish county level data for a set of industries. The use of several benchmarks allows capturing the time evolution of concentration and hence, it can be inferred whether it has increased or decreased overtime.

Regarding the usefulness of the HHI it should be noted that it presents some advantages as well as limitations. For instance, this index complies with the axiom of anonymity (Palan, 2010). This means that if the production of a small firm is absorbed by a bigger firm, overall concentration will increase. However, one of the shortcomings of this index is the fact that is not decomposable. However, according to Acar and Sankaran (1999), if this index is considered as a measure of diversity, it is possible to measure both intersectoral and inter-industry HHI diversification. Nevertheless, it should be noted that this type of issues do not affect the results of the thesis since the goal is to assess concentration per type of industry.

As mentioned in the previous section, this thesis features a model which is based on Midelfart-Knarvik, Overman, Redding and Venables (2000). Since the original model focuses on the last third of the twentieth century, some of the variables will not be included and others are adapted to the second industrial revolution context. The structure of the model is the following, the dependent variable is $ln(s^k_i)$ which reflects the share of industry *k* in province *i*. Among the independent variables there are $ln(pop_i)$ and $ln(man_i)$ and a set of interaction variables. The variable $ln(pop_i)$ reflects the share of population in Sweden that is living in province *i* and $ln(man_i)$ is the share of manufacturing that province *i* represents from the total.

Equation 1: Variables of the Model

$$\begin{aligned}\ln(s_i^k) = \beta_0 &+ \beta_1 \ln(pop_i) + \beta_2 \ln(man_i) + \beta_3 Market\ Potential + \beta_4 Agriculture(\%GDP) \\ &+ \beta_5 Intermediates(\%\ costs) + \beta_6 Economies\ of\ Scale \\ &+ \beta_7 Agric.input\ (\%\ costs) + \beta_8 Sales\ to\ Industry \\ &+ \beta_9 MP * (Intermediate\ Goods) + \beta_{10} MP * (Economies\ of\ Scale) \\ &+ \beta_{11} MP * (Sales\ to\ Industry) + \beta_{12} Agric.\%GDP * (Agric.\%\ costs)\end{aligned}$$

Table 1 shows the interaction variables that are used in the model, while the left hand side is formed by variables that reflect provincial characteristics, the variables from the right hand side are industry features. The use of interaction variables is motivated by the existence of heterogeneity in the characteristics of both country and industry variables. For instance, not all regions have the same market potential or the use of intermediate inputs differs across industries. Therefore, in a context in which economic integration was growing, the process of industrial location is going to be determined by the interaction of the different regional and industrial features (Crafts and Mulatu, 2004).

Table 1: Interaction Variables

County Variables	**Industry Variables**
Market Potential	Intermediate goods as % of total costs
Market Potential	Economies of Scale
Market Potential	Sales to Industry
Agricultural Production as % of GDP	Agricultural Inputs as % of total costs

Source: Midelfart-Knarvik, Overman, Redding and Venables (2000)

Considering the theoretical concepts previously mentioned country characteristics belong to the Heckscher-Ohlin framework while industry features are related to the NEG field. Regarding the first category, market potential is the most relevant. This variable reflects the availability that a county has in terms of accessing to other counties markets, and it also reflects its own economic potential. Therefore, this variable is expressed as the GDP of other areas divided by the distance of county *i* plus the GDP of county *i* over its within-distance. This internal distance is expressed with a coefficient that is obtained by applying the formula in figure 2 which is frequently used in economic geography (Crafts and Mulatu, 2004; Martínez-Galarraga, 2011). In essence, market potential express the economic possibilities that a county offers both in terms of its own wealth, and also, on its geographical location.

Figure 2: Within-Distance Formula

$$d_{rr} = 0{,}333\sqrt{\frac{area\ of\ county_r}{\pi}}$$

The first three interactions are related to NEG forces since they measure the relative pull for centrality that some industries may have with each other. The interaction *Market Potential*Intermediate goods as percentage of total costs*, reflects the importance of intermediate goods in the production process. Moreover, *Market Potential*Sales to Industry* considers the relevance of sales to other industries in terms of total gross output. In essence, it can be stated that these two interactions are based on backward and forward linkages, which are key when determining industrial location. The use of economies of scale as a variable, *Market Potential*Economies of Scale,* is based on the following conjecture. If industries have increasing returns, they have a tendency to locate to core areas in which markets are bigger, and as a result, the pull for centrality is going to be stronger (Krugman, 1991). Finally, the last interaction variable, *Agricultural Production as % of GDP*Agricultural Inputs as Percentage of Total Costs*, belongs to the Heckscher-Ohlin theoretical framework. This is because it considers the relative endowment of agriculture across the Swedish counties. Therefore, in principle, the industries that use agricultural goods intensively will locate in those counties in which agriculture is abundant.

Regarding the interaction variables, it should be noted that only one captures Heckscher-Ohlin forces. Although some papers include other interactions such as educated population*white-collar workers (Crafts and Mulatu, 2004; Martínez-Galarraga, 2011), the availability of data has not allowed to construct more variables related to this theoretical approach.

Since this thesis focus on the second industrial revolution, the process of data collection is very similar as in Crafts and Mulatu, (2004) or Martínez-Galarraga (2011). For instance, information regarding intermediate inputs, agricultural inputs and sales to industry is obtained from input-output tables. The variable economies of scale reflects employment per unit of production. Therefore, this information is retrieved from industrial surveys that provide data regarding number of both workers and factories. Market potential is the result of dividing county GDP per capita over the distance between counties capital cities in kilometres. Within-distance is obtained by applying the formula in figure 2. County area is expressed in squared kilometres and this information is provided by the statistical yearbook of 1914 from Statistics Sweden.

4. Data

This section provides a discussion of the main issues related to the collection and use of data for the empirical exercises that are carried out in this thesis. As mentioned above, the analysis consists on the measurement of industrial concentration as well as the estimation of an econometric model. It is important to consider that, since there was not available data for industries at a county level until 1913, this has limited the period of analysis which extends from 1913 until 1950.

Table 2: Industrial Groups

Group 1	Mining and Metal Industries
Group 2	Stone, Clay and Glass
Group 3	Wood Industries
Group 4	Pulp, Paper and Graphical Industries
Group 5	Food Product Industries
Group 6	Textile and Clothing Industries
Group 7	Leather, Hair and Rubber Industries
Group 8	Chemical Industries
Group 9	Power, Gas and Waterworks

Source: *Industri och bergshantering*

The HHI has been calculated by considering the number of firms at a county level of nine industrial groups. Although, in the calculation of this index it is common to use production levels, in quantities or monetary units, the lack of this information has motivated to use the number of firms. The source used for this purpose is the industrial survey known as *Industri och bergshantering* which is provided by Statistics Sweden. By using the number of firms it is assumed that a higher number of firms in a particular county entails a higher market share of an industrial activity in that geographical unit. Conversely, if there are not many differences across counties in terms of number of firms, the concentration level will be low. Table 2 shows the classification of the nine industrial groups that have been analysed. The structure that has been chosen is the same as the one provided in the industrial survey[4]. The advantage of using this classification is that it comprises the vast majority of industrial activities that belong to the period analysed.

As mentioned above, in this thesis it is used an EGM which has been adapted to the first half of the twentieth century context. Therefore, it should be noted that some modifications have been made, and also, some techniques based on both Crafts and Mulatu (2004) and Martínez-Galarraga, (2011) have been applied. This is due to the fact that these two articles have covered periods which are more similar to the chosen for this thesis. Moreover, they have provided some methodological innovations that solve the main issues related to the estimation of the model. Among them, the most significant is the use of a fixed effects estimation which captures both industry and county characteristics.

The dependent variable is $ln(s^k_i)$ and it has been obtained by calculating the natural logarithm of the share of the nine industrial groups in every of the Swedish counties[5]. The source employed for this variable is the same as in the previous case. Therefore, it is assumed that a higher number of firms in a county implies that it has a higher share of production with respect to the total. The data used for constructing the independent variables is provided by Enflo, Henning and Schön (2015). The procedure of obtaining the first two independent variables, $ln(pop_i)$ and $ln(man_i)$, is similar as in the previous case. The first reflects the share of population in every county and the second is the share of active population that works in industry in every county. The use of employment data in industry rather than its contribution to national GDP is due to the unavailability of data at a county level. With regards to the variable *agriculture as % of GDP*, there is a similar problem. Therefore, the share of agricultural employment at a county level has been used in order to express the contribution of this sector to GDP. The use of this approach rests on the assumption that a higher share of agricultural employment entails a higher contribution of agriculture in national GDP.

[4] Table 2 of this survey shows the information regarding firms, workers and horsepower by industry at a county level. This information has been retrieved for the benchmark years which are 1913, 1920, 1930, 1940 and 1950.

[5] For reasons of simplicity and coherence, the number of counties is 24, which means that both *Stockholm Stad* and *Stockholm Län* form one county. Also, the old regional classification has not been modified. Therefore, considering that nine industrial activities are covered, the number of observations in every benchmark is 216.

The process of calculating market potential is carried out in two steps. First, the distance between capitals cities of counties and also the within-distance is calculated. While, in the first case this is calculated by considering the straight line distance between capital cities, within-distance is obtained by applying the formula provided in figure 2. The second step consists on dividing county GDP per capita of every county over the distance between every county and the rest of the counties[6]. The sum provides the overall market potential of every county. The reasoning behind the use of this procedure is that, if a county has a high GDP per capita and it is close to other counties with a similar situation, its level of market potential is going to be high.

With regards of how industry characteristics have been obtained, the procedure has consisted on deriving information from the Swedish Historical National Accounts (Schön and Krantz, 2015). The variable *Sales to industry* has been calculated by considering both input in and investment in industry as a share of gross total output which appears in table 7 Schön and Krantz (2015, pp. 67). The process of calculating both *agricultural goods* and the *share of intermediates* consisted of calculating the share of these two components as part of gross total output. It should be noted that, with the information provided in the tables, gross total output is the result of summing inputs in and investment in, consumption and total. Since in Schön and Krantz (2015, pp. 63) only the overall contribution of agriculture to industry is shown, it has been assumed that agricultural goods have a higher contribution in those industries which transform these goods. Therefore, this has been taken into account by applying weights that are multiplied by the total contribution of agriculture to industry[7]. Then, this is divided by the gross total output and it gives the share of agriculture as total output. Intermediates goods have been calculated by considering the contribution to industry of construction, transportation and forestry (Schön and Krantz, 2015, pp. 63, 74 and 81). In this case, weights are also employed but it is assumed that they are the same for all industrial categories except for forestry[8]. Finally, the contribution to each sector is divided by the gross total output of every industry. The last variable that reflects industry features is *economies of scale*. This variable reflects the number of workers per firm; therefore, it has been created by dividing workers over firms at every county in the nine industrial groups. As it happens with the variable $ln(s^k_i)$, the source employed for creating this variable is the industrial survey known as *Industri och bergshantering*.

[6] County level GDP per capita employed for calculating market potential is obtained by dividing GDP in constant prices over the county population. The data is provided by Enflo, Henning and Schön (2015), which estimated GDP as well as population at a county level from 1860 to 2010.

[7] Since there are nine industries, food industries account for 5/9 of the contribution of agriculture to industry. At the same time, each of the remaining sectors are assumed to have a proportional share of the remaining agricultural production that is used in industry.

[8] As it happens with the previous case, there is a sector which is wood industries which accounts for the lion share of forestry. Therefore, it is assumed that it consumes the same percentage as a part of gross total output.

5. Main Results and Discussion

The following section covers the two empirical exercises which are the analysis of the HHI and the estimation of an EGM. Although, the first part measures the evolution of industrial concentration and the second captures the determinants of industrial location, they present some similarities. First, the benchmarks are the same for both cases (1913, 1920, 1930, 1940 and 1950). This allows having a view of how these two processes evolved overtime. Also, the industrial groups are the same in the first and second part of this section, the description of every group appears in table 2. The industrial classification follows the same structure as the one provided by the Swedish industrial survey.

Before getting into the discussion of the results of the HHI, it is convenient to discuss certain issues related to the spatial location of the nine industrial activities that are analysed. Table 3 shows the number of firms of the nine industries at a county level. The analysis of the three main metropolitan areas provides the following conclusions. Stockholm County has an important share of production in all sectors and is the leading county in mining and metal industries, pulp, paper and graphical industries and chemical industries. With regards to Gothenburg and Bohus, this county accounts for the highest share in the production of Stone, clay and glasses. Moreover, textiles and clothing industries are also relevant in this country, although in Älvsborgs the number of firms is the highest in this sector. This can be explained by the lower wages which made firms to relocate to areas which had these features as argued by Berger, Enflo and Henning, (2012). With regards to Malmöhus County, it has the highest number of firms in leather hair and rubber industries. At the same time, the production of chemicals and food products is also relevant in this county. With regards to the other industries, food product industries has the highest number of firms in the county of Skaraborgs while in the case of wood industries, the county with more firms in this activity is Gävleborg. It can be stated that in these two activities the relative abundance of factor endowments might have played a role. This is due to the fact that in these counties agriculture and forestry had an important role.

Table 3: Number of firms per Industrial Group at County Level in 1913

	Mining and Metal Industries	Stone, Clay and Glass	Wood Industries	Pulp, Paper and Graphical Industries
Stockholm	255	75	138	139
Uppsala	43	24	35	17
Södermanland	121	27	54	15
Östergötland	77	42	110	32
Jönköping	88	26	142	30
Kronoberg	18	53	81	12
Kalmar	43	52	95	28
Gotland	3	17	8	8
Blekinge	21	79	18	9
Kristianstad	36	66	54	12
Malmöhus	108	101	68	68
Hallands	20	26	24	12
Gothenburg and Bohus	79	129	47	56
Älvsborgs	43	32	77	34
Skaraborgs	28	62	62	28
Värmlands	56	22	91	56
Örebro	109	59	61	26
Västmandland	76	36	66	14
Kopparberg	91	22	104	21
Gävleborg	63	21	188	32
Västernorrlands	21	5	110	35
Jämtlands	11	6	40	7
Västerbotten	6	2	82	11
Norrbotten	11	8	88	7

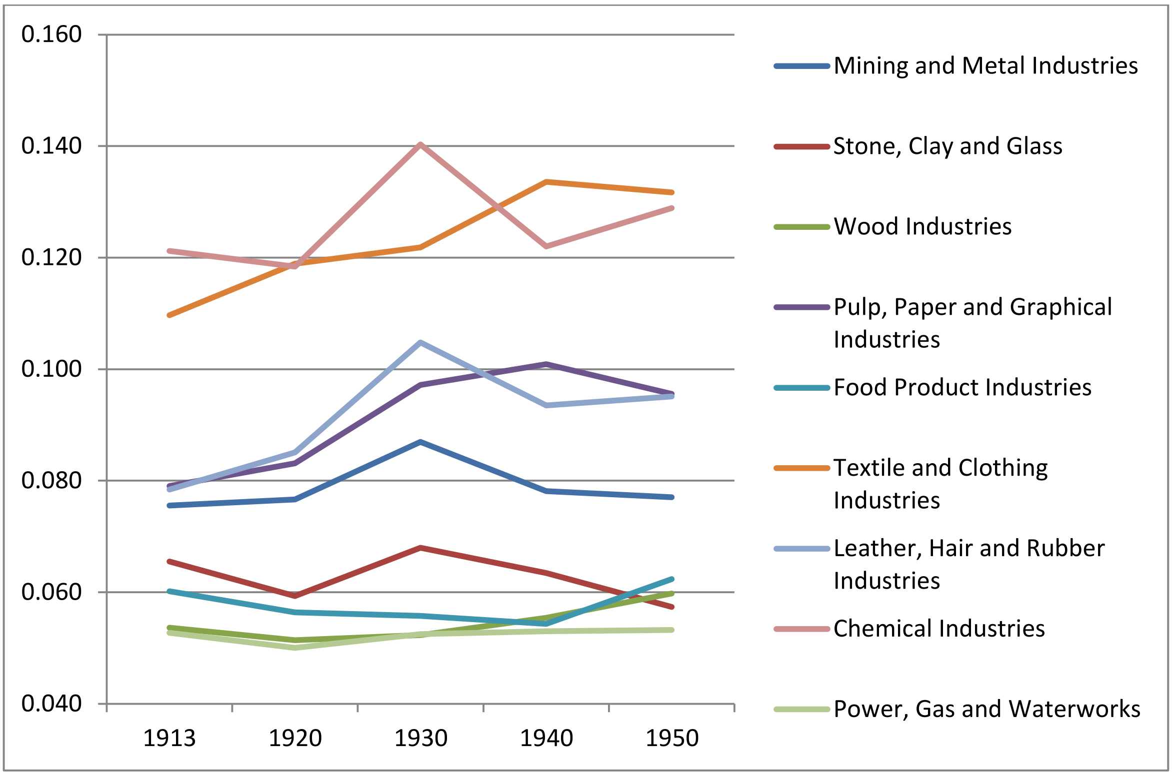
Source: *Industri och bergshantering*

Table 3 (cont.): Number of firms per Industrial Group at County Level in 1913

	Food Product Industries	Textile and Clothing Industries	Leather, Hair and Rubber Industries	Chemical Industries	Power, Gas and Waterworks
Stockholm	238	60	39	61	27
Uppsala	80	5	7	2	5
Södermanland	125	8	4	4	17
Östergötland	213	65	12	8	25
Jönköping	71	17	21	10	16
Kronoberg	29	2	5	5	6
Kalmar	125	6	7	12	14
Gotland	31	1	1	0	1
Blekinge	124	7	6	6	11
Kristianstad	262	12	9	8	12
Malmöhus	325	66	53	61	28
Hallands	106	9	9	2	13
Gothenburg and Bohus	149	69	20	31	12
Älvsborgs	111	90	13	11	18
Skaraborgs	352	15	12	5	19
Värmlands	98	9	12	4	17
Örebro	97	11	30	14	27
Västmandlands	131	3	6	4	13
Kopparberg	105	8	4	10	21
Gävleborg	87	19	5	22	16
Västernorrlands	56	7	5	5	16
Jämtlands	63	3	1	0	6
Västerbotten	60	1	14	0	3
Norrbotten	50	0	15	0	6

Source: *Industri och bergshantering*

Graph 1: HHI for 1913-1950



Graph 1 shows the evolution of the HHI index for the nine industrial groups. The groups which present a higher degree of concentration are textile and clothing industries and chemical industries. During the period observed, although the trends are different, the concentration level of both groups increased. Other groups that also experienced an increase in concentration were pulp, paper and graphical industries, food product industries and wood industries. The rest of the groups either suffered a decline in concentration, such as stone clay and glasses, or experienced a very weak change in terms of industrial concentration. Moreover, it is important to highlight the fact that even in the case of the industrial groups which present a higher degree of concentration, the value is lower than 0.15. According to Horizontal Merger Guidelines of the United States Department of Justice, when the HHI index of an industry is below 0.15 there is not industrial concentration.

In order to have a better understanding of the results form graph 1, it is suitable putting them into an international context. For instance, if the case of France is considered, the HHI for this country shows that the concentration levels were higher, as it appears in graph 2 of the appendix section. Although the available categories for France are not the same for Sweden, it is possible to make some comparisons. For instance, both cast iron and steel show concentration levels which are considerably higher than the levels of mining and metal industries in Sweden. The production of liquors can be compared to food industries in the Swedish industrial classification. Although, in 1905 the concentration levels were very high for this industry, in 1920 and 1930 shows concentration levels which are similar to food industries in Sweden. These results confirm the idea that Sweden represents a special case with respect to other Western economies in terms of industrial concentration.

In essence, the results show that despite the increase in spatial concentration, it cannot be stated that those industries were concentrated at a national level. Therefore, although there are different trends for the industrial groups, it can be stated that the Swedish industry did not present a high degree of concentration during this period. These findings seem to go in line with the results provided in the literature, which emphasise the particularities of Sweden with regards to the evolution of spatial inequalities and industrial location. As mentioned above, in Berger, Enflo and Henning (2012) it is argued that the Swedish industrialisation was primarily a rural phenomenon until the mid-twentieth century. Therefore, this is probably the main reason that explains why concentration levels are low during the period observed. Also, according to NEG theoretical framework, if there is not a strong pull for centrality from urban centres then it is unlikely that the concentration of industrial activities will increase. Nevertheless, these issues are going to be analysed in the next part which deals with the determinants of industrial location.

With regards to the model, the first step consists on carrying out a pooled ordinary least squares (hereafter OLS) estimation which includes the nine industrial categories and the data from the five benchmarks (1913, 1920, 1930, 1940 and 1950). In order to test the validity of the pooled estimation, a Chow test for structural change has been carried out. Since the results have invalidated the pooled estimation, an estimation that corresponds to every benchmark has been done. In order to confirm the reliability of the results, some endogeneity tests have been carried out. Finally, the last step of the model consists on carrying out a fixed effects estimation that controls both industry and county characteristics. For reasons of simplicity, the four interaction variables have been classified according to the following structure. *Market*Intermediates* reflects the interaction between market potential and the share of intermediates. The second interaction is *Market*economies of scale* and it captures the effect of market potential and economies of scale. The last interaction related to NEG is *Market*Sales to industry*. Finally, the interaction that capture Heckscher-Ohlin forces is the share of agriculture in county GDP times the percentage of agricultural goods in gross total output which is renamed as *Agriculture*inputs*.

When analysing the results, the main focus has to be directed to the interaction variables since they reflect the existent relationship between both country and industry characteristics. Also, this allows inferring which forces, NEG or Hecksher-Ohlin, had a stronger influence during the period observed. As mentioned in Martínez-Galarraga (2011), the data used in this type of models present certain issues that have to be addressed. Among them, the most relevant are endogeneity and the presence of a clustered sample. The main reason behind these problems is the fact that observations are not spatially independent since there are nine industrial groups for each of the counties. As a result, this may distort standard errors. In order to solve these issues, the first estimation includes both robust standard errors and cluster robust standard errors. While the first type it is generally used for correcting heteroscedasticity, the second is used when there is intra-cluster dependence, in this case at the county level.

Table 4 shows the estimation output of the pooled regression, this means that all the observations belonging to the five benchmarks have been estimated at the same time. The results show that three out of four of the interaction variables are significant after applying cluster robust standard errors. Also, it should be noted that the sign of all the interaction coefficients are negative, in which the first interaction has the strongest impact. This means that intermediate goods has a negative influence on the process of industrial location. The same applies to both economies of scale and sales to industry. The fourth interaction which is related to the Heckscher Ohlin theoretical framework is not significant. Therefore, according to the results, there is no evidence for either NEG or Heckscher-Ohlin forces operating in industrial location during the period observed. However, it is also true that the use of a pooled estimation implies that parameters are constant over time (Martínez-Galarraga, 2011). Therefore, in order to test this assumption, a Chow test for structural change is carried out. The results show that the F-statistic is equal to 2.57 and this gives evidence to reject the null hypothesis at the 1% level[9]. Taking this information into account, in order to have a more accurate estimation of the model it is necessary to carry out a separate estimation for every benchmark year.

[9] In this type of test, the null hypothesis states that there are no breaks at a number of specified breakpoints. In this case, it was used the years 1913, 1920, 1930 and 1940. The year 1950 was not taken into account since its inclusion in the test would lead to a singular matrix error.

Table 4: Pooled OLS Regression with Robust and Cluster Robust Standard Errors

VARIABLES	Robust	Cluster
Log(Population)	-0.614	-0.614
	(0.396)	(0.527)
Log(Manufacturing)	1.625***	1.625***
	(0.320)	(0.413)
Market potential	0.018***	0.018***
	(0.005)	(0.006)
Agriculture/GDP	16.84***	16.84**
	(5.580)	(6.716)
Share of intermediates	4.914***	4.914**
	(1.610)	(2)
Economies of scale	0.017**	0.017***
	(0.007)	(0.006)
Sales to industry	1.093	1.093
	(1.554)	(1.038)
Share of agricultural goods	0.852	0.852
	(4.628)	(2.843)
Market*Intermediate goods	-0.026**	-0.026**
	(0.001)	(0.013)
Market*Economies of scale	-9.98e-05**	-9.98e-05**
	(4.67e-05)	(4.17e-05)
Market*Sales to Industry	-0.012	-0.012**
	(0.008)	(0.005)
Agriculture*Inputs	-24.98	-24.98
	(89.49)	(53.97)
Constant	-3.698***	-3.698**
	(1.107)	(1.556)
Observations	1,080	1,080
R-squared	0.264	0.264

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 5: Robust and Cluster Robust Estimation for 1913-1930

VARIABLES	1913 Robust	1913 Cluster	1920 Robust	1920 Cluster	1930 Robust
Log(Population)	0.508	0.508	-0.122	-0.122	1.095
	(1.381)	(0.956)	(1.043)	(1.519)	(0.728)
Log(Manufacturing)	0.669	0.669	1.184	1.184	0.142
	(1.016)	(0.705)	(0.882)	(1.124)	(0.634)
Market Potential	0.001	0.001	0.073**	0.073***	0.069**
	(0.043)	(0.019)	(0.037)	(0.0164)	(0.028)
Agriculture/GDP	24.41	24.41	16.75	16.75	5.750
	(19.65)	(16.51)	(12.76)	(15.17)	(15.50)
Share of Intermediates	0.615	0.615	15.94**	15.94***	19.66**
	(6.417)	(4.501)	(7.388)	(3.867)	(9.102)
Economies of Scale	0.03*	0.03***	0.052**	0.0522***	0.044
	(0.015)	(0.00770)	(0.026)	(0.019)	(0.027)
Sales to Industry	-22.63	-22.63***	4.243	4.243**	9.488**
	(13.86)	(8.628)	(5.067)	(1.662)	(4.021)
Share of Agricultural Goods	-12.90	-12.90	5.915	5.915	-6.846
	(15.95)	(10.43)	(10.90)	(4.537)	(14.61)
Market*Intermediate Goods	-0.019	-0.019	-0.142**	-0.142***	-0.132**
	(0.066)	(0.047)	(0.07)	(0.037)	(0.062)
Market*Economies of Scale	-0.000*	-0.000***	-0.000*	-0.000**	-0.000
	(0.000)	(8.96e-05)	(0.000)	(0.000)	(0.000)
Market*Sales to Industry	0.207	0.207**	-0.024	-0.024	-0.06**
	(0.144)	(0.096)	(0.044)	(0.017)	(0.025)
Agriculture*Inputs	89.12	89.12	10.15	10.15	253.3
	(304.8)	(193.3)	(233.4)	(85.86)	(301.3)
Constant	0.862	0.862	-9.58**	-9.576***	-10.37**
	(4.224)	(2.900)	(4.782)	(3.304)	(4.781)
Observations	216	216	216	216	216
R-squared	0.361	0.361	0.291	0.291	0.361

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5 shows the estimation output with both robust and cluster robust standard errors for the five benchmarks. The first interaction which is *Market*Intermediate Goods* is significant from the year 1920 onwards. Regarding the sign of its coefficient, it is negative during the whole period of analysis. With regards to the second interaction, which identifies the impact of economies of scale, its significance varies across the different benchmarks. In the first three benchmarks (1913, 1920 and 1930), the coefficient is significant at the 5% level. However it changes and it loses significance in 1940 and 1950. Nevertheless, it also should be noted that the value of the coefficient is negative as in the previous interaction variable. The variable *Market*Sales to industry*, shows a different trend than the other two variables. In 1913, the coefficient of this variable is significant as well as positive. Moreover, by considering that the value of the coefficient is higher than the other interaction that is significant but negative, it can be stated that NEG forces were operating in this year. However, this variable losses significance in 1920 and it also becomes negative. Although it recovers significance in 1930 and 1940 it still has a negative value. Finally in the year 1950 the coefficient is positive but not significant. With regards to the variable that captures Heckscher-Ohlin forces (*Agriculture*Inputs*) there is a phase in which it has a positive sign (1913, 1920 and 1930) and other in which is negative (1940 and 1950). Regarding the significance of this variable, it is only significant at the 10% level in 1950, which means that it does not have enough influence in the model. In essence, the results provided by the estimation show that there is no evidence for NEG or Heckscher-Ohlin forces behind the process of industrial location in Sweden.

Table 5 (cont.): Robust and Cluster Robust Estimation for 1930-1950

VARIABLES	1930 Cluster	1940 Robust	1940 Cluster	1950 Robust	1950 Cluster
Log(Population)	1.095	-0.0775	-0.078	-0.750	-0.750
	(1.249)	(0.544)	(0.616)	(0.679)	(0.490)
Log(Manufacturing)	0.142	1.114	1.114	1.594**	1.594***
	(0.938)	(0.682)	(0.679)	(0.739)	(0.475)
Market Potential	0.069***	0.049*	0.049**	0.01*	0.01**
	(0.016)	(0.028)	(0.021)	(0.005)	(0.004)
Agriculture/GDP	5.750	31.09**	31.09**	29.02***	29.02***
	(14.42)	(14.42)	(13.55)	(10.52)	(8.943)
Share of Intermediates	19.66***	15.50**	15.50**	8.950***	8.950***
	(6.405)	(7.453)	(6.314)	(3.216)	(2.394)
Economies of Scale	0.044**	0.053	0.053	0.025	0.025**
	(0.019)	(0.033)	(0.033)	(0.022)	(0.012)
Sales to Industry	9.488***	12.73**	12.73***	-5.198	-5.198
	(2.62)	(5.732)	(2.896)	(6.606)	(5.424)
Share of Agricultural Goods	-6.846	13.78	13.78**	5.974	5.974**
	(7.703)	(11.33)	(6.571)	(4.373)	(2.393)
Market*Intermediate Goods	-0.132***	-0.0834*	-0.083**	-0.035**	-0.035***
	(0.044)	(0.044)	(0.038)	(0.014)	(0.011)
Market*Economies of Scale	-0.000**	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(6.19e-05)
Market*Sales to Industry	-0.059***	-0.053*	-0.053***	0.014	0.014
	(0.019)	(0.029)	(0.014)	(0.026)	(0.02)
Agriculture*Inputs	253.3	-146.9	-146.9	-214.8*	-214.8*
	(172.5)	(190.2)	(118.1)	(124.1)	(110.4)
Constant	-10.37***	-11.90**	-11.90***	-3.727**	-3.727***
	(3.401)	(5.806)	(4.331)	(1.754)	(1.257)
Observations	216	216	216	216	216
R-squared	0.361	0.383	0.383	0.377	0.377

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In order to confirm the results shown in table 5, attention should be paid to dealing with the problem of endogeneity because it may distort the true value of both the coefficients and standard errors. The variable that may cause an endogeneity problem is market potential as well as its interactions. This is due to the fact that this variable is influenced by the location decisions of both firms and workers (Combes, Lafourcade, Thisse and Toutain, 2011). In order to test for endogeneity, a Two-Stage Least Squares (hereafter TSLS) estimation it is carried out in which the instrumental variable is a lagged market potential which corresponds to the previous decade of every benchmark. The results of the Durbin and Wu-Hausman test in do not provide enough evidence to reject the null hypothesis of exogeneity. Therefore, in none of the five benchmarks there is an endogeneity problem. The information related to this test is available at table 7 in the appendix section.

As mentioned earlier, the type of data used in this type of estimations may cause measurement error due to the use of both county and industry characteristics (Martínez-Galarraga 2011). In order to solve this potential problem, it is necessary to carry out a fixed effects estimation which may provide a more accurate estimation that overcomes the potential measurement error problem.

Table 6 shows the estimation output of a fixed effects estimation with both robust and cluster robust standard errors. Regarding the first interaction (Market*Intermediate Goods) it has a negative sign throughout the whole period of analysis. Also, it should be noted that this variable is significant from the year 1920 onwards. The interaction which reflects the effect of economies of scale it is negative from 1913 to 1950 but its coefficient is only significant during the first three benchmarks. In essence, it can be stated that neither intermediate goods nor economies of scale were influencing the process of industrial location. The interaction between market potential and sales to industry presents certain differences with respect to the previous interactions. First, the coefficient is significant but positive in 1913. Also by looking at the value of the other interaction which is significant, it can be stated that its impact is stronger. Therefore, in 1913 there is evidence that forward linkages were influencing the process of industrial location in Sweden. However, it also should be noted that this variable losses significance in the next period, and also, its sign changes. Therefore, although this variable becomes significant again in 1930, it no longer has a positive effect. Finally, the last interaction captures Heckscher-Ohlin forces in the model estimated. The coefficient changes its sign during the period observed since it is positive in 1913 and 1930 while negative in 1920, 1940 and 1950. However, in terms of significance, this variable is only significant at the 10% level in 1950. Therefore, this shows that there is no evidence that Heckscher-Ohlin forces were operating during the period observed.

Table 6: Fixed Effects Estimation for 1913-1930

VARIABLES	(1913) Robust	(1913) Cluster	(1920) Robust	(1920) Cluster	(1930) Robust
Log(Population)	-8.419	-8.419***	-12.83	-12.83***	-7.175
	(11.40)	(0.236)	(8.277)	(1.19)	(5.73)
Log(Manufacturing)	2.36	2.360***	6.19	6.19***	2.928
	(4.217)	(0.178)	(4.063)	(0.575)	(2.793)
Market Potential	0.123	0.123***	0.148	0.148***	0.117*
	(0.136)	(0.022)	(0.09)	(0.006)	(0.063)
Agriculture/GDP	167.5	167.5***	-55.13	-55.13***	-9.098
	(126.3)	(11.90)	(113.8)	(12.18)	(105.5)
Share of Intermediates	-4.529	-4.529	14.14**	14.14***	19.83**
	(6.473)	(5.428)	(6.627)	(3.573)	(8.158)
Economies of Scale	0.03**	0.0296***	0.045**	0.045***	0.038*
	(0.014)	(0.005)	(0.022)	(0.017)	(0.022)
Sales to Industry	-30.66**	-30.66***	-3.053	-3.053	3.369
	(13.63)	(11.74)	(8.22)	(7.99)	(4.462)
Share of Agricultural Goods	-37.40*	-37.40*	-50.89	-50.89	-39.71
	(19.30)	(19.62)	(34.32)	(41.97)	(24.94)
Market*Intermediate Goods	-0.014	-0.014	-0.129**	-0.129***	-0.125**
	(0.062)	(0.049)	(0.063)	(0.034)	(0.055)
Market*Economies of Scale	-0.000*	-0.000***	-0.000	-0.000**	-0.000
	(0.000)	(6.70e-05)	(0.000)	(0.000)	(0.000)
Market*Sales to Industry	0.206	0.206**	-0.03	-0.03*	-0.061**
	(0.126)	(0.097)	(0.043)	(0.016)	(0.025)
Agriculture*Inputs	43.88	43.88	-10.13	-10.13	223.0
	(275.0)	(219.7)	(218.5)	(90.93)	(294.2)
County Dummies	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes
Constant	-29.52	-29.52***	-30.24	-30.24***	-27.45
	(35.61)	(4.439)	(23.29)	(5.188)	(19.84)
Observations	216	216	216	216	216
R-squared	0.483	0.483	0.419	0.419	0.482

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6 (cont.): Fixed Effects Estimation for 1930-1950

VARIABLES	1930 Cluster	1940 Robust	1940 Cluster	1950 Robust	1950 Cluster
Log(Population)	-7.175***	1.063	1.063	0.395	0.395
	(0.821)	(2.413)	(1.395)	(1.681)	(0.657)
Log(Manufacturing)	2.928***	-1.82	-1.82*	1.390	1.39***
	(0.533)	(2.423)	(0.946)	(1.054)	(0.505)
Market Potential	0.117***	0.082*	0.082***	0.003	0.003
	(0.011)	(0.049)	(0.018)	(0.014)	(0.005)
Agriculture/GDP	-9.098	96.58	96.58***	53.98	53.98***
	(7.594)	(72.4)	(11.73)	(37.54)	(5.156)
Share of Intermediates	19.83***	13.38**	13.38**	5.955*	5.955**
	(6.504)	(6.734)	(6.697)	(3.23)	(2.701)
Economies of Scale	0.038**	0.049	0.049	0.024	0.024*
	(0.015)	(0.031)	(0.032)	(0.021)	(0.013)
Sales to Industry	3.369	3.55	3.55	-14.46	-14.46
	(3.768)	(5.224)	(5.399)	(10.11)	(11.04)
Share of Agricultural Goods	-39.71*	-17.52	-17.52	-29.51*	-29.51
	(23.59)	(16.38)	(13.02)	(17.13)	(20.03)
Market*Intermediate Goods	-0.125***	-0.079*	-0.079**	-0.033**	-0.033**
	(0.042)	(0.04)	(0.04)	(0.014)	(0.013)
Market*Economies of Scale	-0.000**	-0.000	-0.000	-7.84e-05	-7.84e-05
	(0.000)	(0.000)	(0.000)	(9.28e-05)	(6.83e-05)
Market*Sales to Industry	-0.061***	-0.053**	-0.053***	0.013	0.013
	(0.019)	(0.026)	(0.015)	(0.025)	(0.021)
Agriculture*Inputs	223.0	-152.4	-152.4	-218.8*	-218.8*
	(183.4)	(210.6)	(124.0)	(127.0)	(115.0)
County Dummies	Yes	Yes	Yes	Yes	Yes
Industry Dummies	Yes	Yes	Yes	Yes	Yes
Constant	-27.45***	-19.83	-19.83***	4.253	4.253
	(1.787)	(14.32)	(2.94)	(7.787)	(4.76)
Observations	216	216	216	216	216
R-squared	0.482	0.520	0.520	0.472	0.472

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

When comparing the results of both the OLS and fixed effects estimations, it should be noted that the results are very similar. This is particularly true when assessing the value and the significance of the interaction coefficients. Moreover, another remarkable finding is that when a fixed effect estimation is conducted, the value of the r-squared increases in all the periods included. This is important since it makes the results more similar as the ones provided in the literature (Crafts and Mulatu, 2004; Martínez-Galarraga, 2011). Therefore, the analysis is derived from the results provided by the fixed effects estimation.

Regarding the economic interpretation of the estimation results, it should be noted the special features of the country that has been analysed. As mentioned before, the evolution of the industrialisation phenomenon in Sweden presents certain particularities which makes this country a special case. For instance, as argued by Enflo, Henning and Schön (2015) since the eve of industrialisation, Swedish regions underwent into a process of increasing homogenisation. This is the contrary case of countries such as the United States or France in which there was an increase of spatial inequalities during the early twentieth century (Kim, 1995; Combes, Lafourcade, Thisse and Toutain, 2011). Therefore, by taking this information into account it is possible to have a deeper understanding of the findings provided in the empirical section of this thesis. The results of the estimation only show evidence for NEG forces influencing industrial location in 1913, and specifically, forward linkages. The rest of the interaction variables related to NEG are always negative during the period analysed. Moreover, the interaction which belongs to the Heckscher-Ohlin theoretical framework is not significant in none of the benchmarks. Therefore, with the exception of 1913, there is no evidence that these two type of forces influenced industrial location in the first half of the twentieth century in Sweden.

In essence, the results of the model confirm the findings provided in the literature specialised in Swedish economic history. The weak influence of NEG forces during the period observed is probably due to the lack of pull for centrality of urban agglomerations. As mentioned before, industry was primarily a rural phenomenon until the mid-twentieth century. Therefore, the non-significance of the coefficients related to NEG gives evidence that agglomeration economies did not play a role behind industrial location in Sweden during the period observed. Moreover, Heckscher-Ohlin forces, identified as agricultural abundance, did not have an influence on industrial location. When these results are compared in an international context, some similarities as well as some differences can be found. For instance, with regards to NEG forces, while their impact in Britain was almost inexistent, in Spain, they started to operate with the arrival of the second industrial revolution (Crafts and Mulatu, 2004; Martínez-Galarraga, 2011). Although this has not been tested in this thesis, in Britain the weakness of NEG can be explained by the fact that transport costs were not sufficiently low, and as a result, this type of forces could not arise. With regards to agricultural abundance, it had a positive influence in the process of industrial location, although the evolution is different in both Britain and Spain. Finally, it should be noted that due to data limitations it has not been possible to include in the analysis periods before the year 1913. This represents a limitation of this study since it is possible that the relative impact of both NEG and Heckscher-Ohlin forces may evolved from the first to the second industrial revolution.

6. Conclusion

In this thesis both industrial concentration and industrial location have been analysed by using the HHI and by estimating an EGM. The results derived from using the index provide the following conclusions. First, although most industrial groups increased in terms of concentration, it is also true that in most cases the increase was not remarkable. Second, even in those industries in which spatial concentration increased in a strong way, (textile and clothing industries and chemicals) this did not mean that they transformed into concentrated industries during the period analysed. In essence, the information provided by this index seems to confirm that industrialisation was primarily a rural phenomenon in Sweden (Berger, Enflo and Henning, 2012). This is due to the fact that industrial activities did not show a high level of concentration. Moreover, when compared to other countries the concentration of levels of Swedish industry can be considered as low.

The results of the model are somewhat related to the evolution of the HHI. This is due to the fact that there is only evidence for NEG forces on industrial location in the first year of the analysis. Also, it should be noted that the Heckscher-Ohlin forces did not have an influence on the location of industries from 1913 to 1950. Therefore, the lack of influence of these two types of forces may be related to the evolution of spatial inequalities in Sweden. The weakness of NEG forces means that agglomeration economies did not play a significant role. As a result, the big urban centres did not become into areas that pulled for centrality, and hence, industrial concentration did not grow. Finally, although there is the possibility that some NEG forces were operating before 1913, the lack of available data before this year makes that this hypothesis cannot be tested.

7. Appendix

The following list contains the data sources which have been used for this thesis:

The Swedish industrial survey, also known as *Industri och bergshantering,* is retrieved from the Statistics Sweden website. It contains the information regarding the number of firms and workers of the nine industrial groups. This source has been used for both the calculation of the HHI as well as the dependent variable of the model and the variable *economies of scale*.

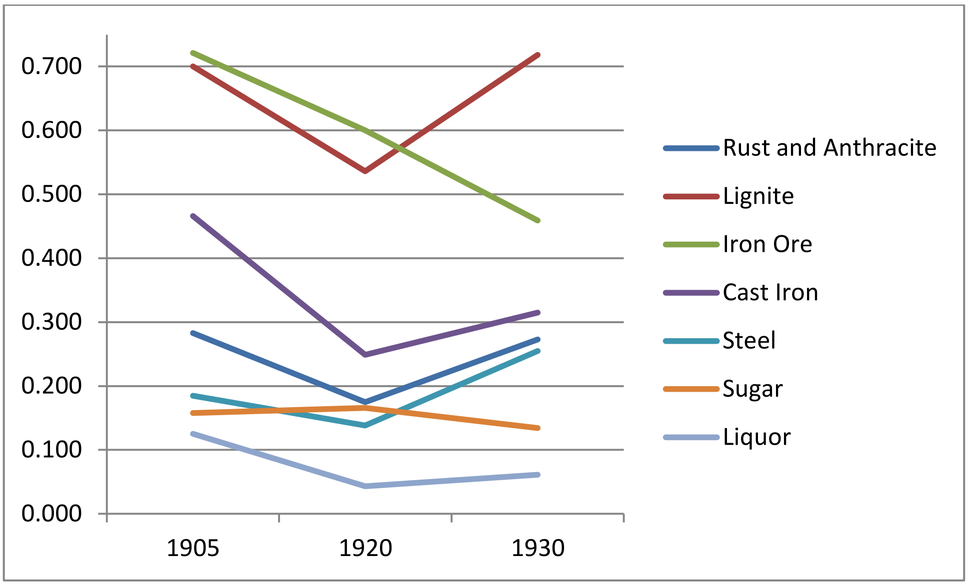
The information regarding the surface expressed in squared kilometres has been retrieved from the Statistical Yearbook of Sweden of 1914 which is available at the Statistics Sweden website.

The Swedish Historical National Accounts from Schön and Krantz (2015) contains the data related to input-output which is necessary to calculate *intermediate inputs/gross total output,* agricultural *inputs/gross total output* and *sales to industry/gross total output*.

The Swedish Historical Regional Accounts (1860-2010) from Enflo, Henning and Schön (2015) provides information regarding employment in agriculture, manufacturing, population and GDP at a county level. The variables obtained from these regional accounts are *agriculture as % of GDP*, the *share of manufacturing/national* ($ln(man_i)$) and the *share of population/national* ($ln(pop_i)$) and *market potential*.

The data for calculating the HHI in France is provided by the Statistical Yearbooks of the years 1905, 1920 and 1930, which are available at the website of the National Library of France.

Graph 2: HHI Index in France



Source: Statistical Yearbook of France of 1905, 1920 and 1930

Table 7: Endogeneity tests for the OLS estimation

	1913		1920		1930	
	Test	*P-Value*	*Test*	*P-Value*	*Test*	*P-Value*
Durbin	1.216	0.270	0.001	0.970	0.862	0.353
Wu-Hausman	1.144	0.286	0.001	0.971	0.809	0.369

	1940		1950	
	Test	*P-Value*	*Test*	*P-Value*
Durbin	0.063	0.801	1.332	0.248
Wu-Hausman	0.059	0.808	1.253	0.264

8. Bibliography

Acar, W. and Sankaran, K., (1999). The Myth of the Unique Decomposability: Specializing the Herfindahl and Entropy Measures? Strategic Management Journal, Vol. 20, 969-975.

Berger, T., Enflo, K. and Henning, M., (2012). Geographical location and urbanisation of the Swedish manufacturing industry, 1900–1960: evidence from a new database. Scandinavian Economic History Review, Vol. 60, No. 3, pp. 290-308.

Brülhart, M. and Torstensson, J., (1998). Regional Integration, Scale Economies and Industry Location in the European Union. Discussion paper no. 1435. Centre for Economic Policy Research, London.

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

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

Crafts, N. and Mulatu, A., (2004). What Explains the Location of Industry in Britain, 1871-1931? CEPR Discussion Paper No. 4356, ISSN 0265-8003, April.

Enflo, K., Henning, M. and Schön, L., (2015). Swedish Regional GDP 1855-2000. Estimations and General Trends in the Swedish Regional System. Research in Economic History, Vol. 30, pp. 47-89.

Enflo, K. and Rosés, J.R., (2015). Coping with regional inequality in Sweden: structural change, migrations, and policy, 1860–2000. Economic History Review, Vol. 68, Iss. 1, pp. 191-217.

Geary, F. and Stark, T., (2002). Examining Ireland's Post-Famine Economic Growth Performance. The Economic Journal, Vol. 112, No. 482, pp. 919-935.

Henning, M., Englo, K. and Andersson, F.N.G., (2011). Trends and cycles in regional economic growth How spatial differences shaped the Swedish growth experience from 1860–2009. Explorations in Economic History, Vol. 48, Iss. 4, pp. 538-555.

Herfindahl, O. C., (1951). Concentration in the steel industry. Ann Arbor: University Microfilms.

Hirschman, A.O., (1964). The Paternity of an Index. The American Economic Review, Vol. 54, No. 5, pp. 761.

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, Vol. 110, No. 4, pp. 881-908.

Krantz, O. and Schön, L., (2015). Swedish historical national accounts 1800-2000. Lund Studies in Economic History 41. Almqvist & Wiksell International, Stockholm, Sweden.

Krugman, P., (1991). Increasing Returns and Economic Geography. Journal of Political Economy, Vol. 99, pp. 483-499.

Krugman, P. and Venables, A.J., (1995). Globalization and the Inequality of Nations. The Quarterly Journal of Economics, Vol. 110, No. 4, pp. 857-880.

Krugman, P., Obstfeld, M. and Melitz, M.J., (2012). International Economics: Theory & Policy, Ninth Edition. Boston, United States, Pearson.

Magnusson, L., (2002). An Economic History of Sweden. London: Taylor and Francis. Retrieved from http://ebookcentral.proquest.com/lib/lund/detail.action?docID=170465

Martínez-Galarraga, J., (2011). The determinants of industrial location in Spain, 1856-1929. Explorations in Economic History, Vol. 49, Iss. 2, pp. 255-275.

Midelfart-Knarvik, K.H., Overman, H.G., Redding, S.J. and Venables A.J., (2000). The Location of European Industry. Economic Papers No. 142, April. ECFIN/318/00-EN. [EU Commission - Working Document]

Myrdal, G., (1971). Economic Theory and Underdeveloped Regions. New York, United States, Harper & Row.

Palan, N., (2010). Measurement of Specialization – The Choice of Indices. Research Centre of International Economics (FIW), FIW Working Paper No. 62.

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

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, Vol. 47, Iss. 2, pp. 404-425.

Solow, R.M., (1956). A Contribution to the Theory of Economic Growth. The Quarterly Journal of Economics, Vol. 70, No. 1, pp. 65-94.

Tabuchi, T., (1998). Urban Agglomeration and Dispersion: A Synthesis of Alonso and Krugman. Journal of Urban Economics, Vol. 44, pp. 333-351.

U.S. Department of Justice and Federal Trade Commission (2010). Horizontal Merger Guidelines, August 2010.