LUND UNIVERSITY
School of Economics and Management

Master programme in Economic Growth, Innovation and Spatial Dynamics

What Determines the Location of Industry? Endowments, Market Potential, and Industry Location in Swedish Regions, 1900-1960

Thor Berger
thorberger@gmail.com

Abstract: In this paper I examine the location of manufacturing industries across Swedish regions, between 1900 and 1960, using a novel dataset on manufacturing employment disaggregated into eight industries and 24 (NUTS-III) regions. By means of OLS estimations of a model where the regional share of employment in an industry is determined by interactions between region and industry characteristics I show that factor endowments and market potential jointly determined the regional distribution of industries. More specifically, industries with increasing returns to scale and backward linkages located in regions with high market potential. Conversely, industries with forward linkages located in regions with low market potential. Regions endowed with agricultural land and woodland attracted industries that intensively used inputs from the agricultural and forestry sectors. The regional distribution of iron ore did not have substantial effects on the distribution of industries. In the postwar period the endowment of an educated population became important in attracting skill-intensive industries. In sum, market potential seems to have been a more important determinant of industry location than factor endowments. The main results are confirmed using an instrumental variables approach (two-step GMM) where market potential is instrumented by predetermined and geographical instruments.
Key words: Industry location, Sweden, Industrialization, Heckscher-Ohlin, New Economic Geography

EKHR71
Master thesis (15 credits ECTS)
June 2012
Supervisor: Kerstin Enflo
Examiner: Karl-Johan Lundquist

What Determines the Location of Industry?

Endowments, Market Potential, and Industry Location in Swedish Regions, 1900-1960

Thor Berger*†

EKHR71: First Year Independent Research
Supervisor: Kerstin Enflo
Department of Economic History
Lund University

May 30, 2012

Abstract

In this paper I examine the location of manufacturing industries across Swedish regions, between 1900 and 1960, using a novel dataset on manufacturing employment disaggregated into eight industries and 24 (NUTS-III) regions. By means of OLS estimations of a model where the regional share of employment in an industry is determined by interactions between region and industry characteristics I show that factor endowments and market potential jointly determined the regional distribution of industries. More specifically, industries with increasing returns to scale and backward linkages located in regions with high market potential. Conversely, industries with forward linkages located in regions with low market potential. Regions endowed with agricultural land and woodland attracted industries that intensively used inputs from the agricultural and forestry sectors. The regional distribution of iron ore did not have substantial effects on the distribution of industries. In the postwar period the endowment of an educated population became important in attracting skill-intensive industries. In sum, market potential seems to have been a more important determinant of industry location than factor endowments. The main results are confirmed using an instrumental variables approach (two-step GMM) where market potential is instrumented by predetermined and geographical instruments.

Keywords: Industry location, Sweden, Industrialization, Heckscher-Ohlin, New Economic Geography

*Contact: thorberger@gmail.com

†I would like to thank Kerstin Enflo and participants at the Frontier Research in Economic and Social History (FRESH) seminar at Warwick in Venice 2012 for helpful comments and suggestions.

Contents

List of Tables

List of Figures

1 Introduction

Why are some regions more industrial than others? The ongoing integration efforts within the European Union, where there has been a considerable increase in the regional concentration of industries since the 1980s, have revived interest in the effects of economic integration on the regional distribution of industrial activity (see, e.g., Amiti, 1998; Brülhart, 1998; Brülhart and Torstensson, 2001; Puga, 2002; Midelfart-Knarvik et al., 2004). Similarly, economic historians have recently emphasized that the integration of regional markets in the 19th century was associated with considerable regional variation in industrial activity as central regions attracted dynamic industries whereas peripheral regions stagnated[1].

Industrialization and regional integration in Sweden provides a contrasting experience as early industrialization benefited virtually all regions (Henning et al., 2011). Increasing foreign demand and the expansion of canals and floatways around the mid-19th century promoted an export-led expansion of the forestry industries in the peripheral northern regions (Törnlund and Östlund, 2006). Although exports of timber became less important in the early 20th century the growth in exports of pulp and paper and the abundance of hydroelectric power, used in the pulping processes, ensured the importance of the vast woodlands in northern Sweden well into the 20th century (Fridlizius, 1963; Schön, 2010). Similarly, the construction of the railways, initiated in the late 1850s, lowered transportation costs which in combination with the introduction of the Thomas process in the 1890s promoted an expansion of the extraction of phosphoric iron ore in the northernmost region of Norrbotten (Olsson, 2007)[2]. The distribution of natural resources thus constituted a linchpin of the regional diffusion of economic activity in Sweden around the turn of the century. In addition, institutional changes in the mid 19th century, such as the dismantling of the Guild Ordinance in the 1840s and the abolishment of internal passport requirements in the 1860s, further promoted the diffusion of industrialization and regional migration resulting in strong wage convergence, across regions and occupations, in the late 19th century (Lundh, 2002; Lundh et al., 2005).

Some accounts have stressed that the major reallocation of regional production occurred during the initial period of industrialization, from the 1870s until the outbreak of the First World War, whereafter the regional system remained stable (Söderberg and Lundgren, 1982; Henning et al., 2011). But around the turn of the century the Swedish manufacturing industry became increasingly sophisticated as growth gravitated toward chemical industries, mechanical engineering, and graphical industries promoting the urbanization and concentration of industrial production (Jörberg, 1961; Schön, 2010). The modern industries were to a larger extent dependent on domestic regional markets promoting a pull of centrality where industries were drawn toward the dense urban regions, especially Stockholm, in the early 20th century (Schön, 1997). This gradual shift from a traditional, mainly natural resource-based, industry to a modern industry coincided with the emergence of a distinct north-south divide as the northern regions share in total manufacturing employment decreased from 24

[1]See Tirado et al. (2002) and Rosés (2003) regarding the dominance of Catalonia in early Spanish industrialization, Felice (2011) regarding the divergence between northern and southern Italy, and Klein and Crafts (2011) regarding the concentration of industrial activity to the Manufacturing Belt in the United States.

[2]The annual output of iron ore in the county of Norrbotten increased from 40 tons in the 1880s to roughly 10,000,000 tons in the mid 1930s, or from well below 1 percent to roughly 70 percent of Sweden's total output over the same period. (Statistical Yearbook of Sweden, 1940)

percent in 1900 to around 16 percent in the 1940s (Berger et al., 2012). This '*Northern Problem*', acknowledged by contemporary public officials, was mainly interpreted as structural problems within these regions. It was argued that the growth of the metropolitan regions, and especially the expansion of the capital Stockholm, at the expense of the northern regions was intrinsic to a successful structural shift from agriculture and traditional industries toward modern, predominately urban, manufacturing (see, e.g., Nilsson, 2006). The emergence of this geographical divide occurred to the backdrop of the breakthrough of industrialization in Sweden, as the share of manufacturing in total employment increased from 21 to 36 percent between 1900 and 1960 (Krantz and Schön, 2007). But in addition to this geographical divide industrialization also produced a salient regional distribution of industries as the iron industry was chiefly located in the region of Bergslagen in central Sweden, the foodstuffs industry located in the southernmost region of Scania, the textile industry located in western Sweden, whereas eastern Sweden attracted the mechanical engineering industry in the early 20th century (Söderberg, 1984; Olsson, 2007; Schön, 2010). The gravitation of industries towards mid- and southern Sweden as well as the differences in regional industrial composition during the breakthrough of industrialization raises questions that regard the underlying determinants of industry location.

Scholarly interest in the determinants of industry location have mainly been informed by two theoretical frameworks. In the Heckscher-Ohlin model of international trade, regions differ in their endowments of different factors (see, e.g., Flam and Flanders (1991); Leamer (1995)). When trade costs decrease regions will specialize in production of goods that intensively uses the relatively abundant factors of each region. The distribution of industries is then determined by the underlying pattern of comparative advantage[3]. On the other hand, the New Economic Geography (Krugman, 1991a)[4] has emphasized that that the location of industries is determined by the interaction between transportation costs, increasing returns to scale, and forward and backward linkages. If scale economies are important an industry will locate in the region offering the largest market, incurring transportation costs to supply other regional markets from there (Krugman, 1991b). Similarly, if a large number of firms locate in a region the demand for intermediate products will be higher there, encouraging industries producing these intermediates to relocate to that region (Krugman and Venables, 1995). Although these theories are not mutually exclusive they emphasize two very different sets of determinants of industry location. A central question, ultimately of empirical nature, thus concerns if factor endowments or proximity to markets is the most important determinant of industry location.

Recent empirical work have emphasized that industry location is jointly determined by Heckscher-Ohlin and New Economic Geography forces. Midelfart-Knarvik et al. (2000) derive a general equilibrium model of industry location that nests the Heckscher-Ohlin and New Economic Geography arguments and Midelfart-Knarvik et al. (2004) show that forward and backward linkages and the endowment of skilled labour became increasingly important as determinants of industry location within the European Union from the 1980s and onwards. Building on the pioneering work of Midelfart-Knarvik et al. (2000) economic historians have

[3]This can be extended to a linear relationship between factor endowments and production through the Rybczynski theorem (see, e.g., Kim (1999)). An increase in the supply of a factor will lead to an increase in the production of goods that intensively uses that factor and a reduction in the production of other goods.

[4]The New Economic Geography literature is extensive, in addition see, e.g., Puga, 1999; Neary, 2001; Head and Mayer, 2004; Ottaviano and Thisse, 2004.

recently turned to an examination of the effects of regional integration on industry location in the late 19th and early 20th century. Crafts and Mulatu (2005, 2006) show that the distribution of industries in the United Kingdom was determined both by access to markets and factor endowments between 1870 and 1930. But whereas the importance of factor endowments persisted century the role of market potential faded in the early 20th century. Wolf (2007) examines the relocation of industries in Poland, following the reunification in 1918, and concludes that both comparative advantage and market potential were important in determining the distribution of industries. The most important factors were the endowment of skilled labor and high market potential in order to attract skill-intensive industries and industries with forward linkages. Betran (2011) and Martinez-Galarraga (2012) examine the distribution of industries in Spain and conclude that both factor endowments and market potential was important, with the latter increasing in importance from the mid 19th century until the 1930s. Klein and Crafts (2011) show that, contrary to the conclusions reached by Kim (1995, 1999), the existence of the Manufacturing Belt in the United States relied on the superior market access of these states, whereas the importance of factor endowments, while important in the 19th century, diminished over time.

My contribution to the literature is threefold. First, this paper will be an addition to the emerging literature on the determinants of industry location in a historical perspective. Second, by drawing upon a recently constructed database on the Swedish manufacturing industry I will provide a descriptive account of the changes in the regional distribution of manufacturing employment and regional specialization between 1900 and 1960. Third, I will provide an econometric treatment of the relative importance of natural resources, human capital, and market potential in determining the distribution of industries across Swedish regions. More specifically I will answer the following questions:

- How did the regional distribution of the manufacturing industry and regional specialization evolve in Sweden between 1900 and 1960?
- To what extent did Heckscher-Ohlin and New Economic Geography forces shape the regional distribution of manufacturing industries?
- Did the relative influence of these forces change over time?

The results can briefly be summarized as follows. In section 4.1 I show that the Swedish manufacturing industry became increasingly concentrated in the interwar period whereas the postwar period was characterized by strong convergence. Industrial concentration thus traced out a bell-shaped evolution between 1900 and 1960, resulting in a more even regional distribution of manufacturing employment in the 1960s than during any previous decade of the 20th century. But an examination of eight individual industries makes clear that the degree and trends of concentration varied considerably across industries during this period. Regarding regional specialization, it decreased from the early 1900s, flattened out during the interwar period, and decreased further in the postwar period such that Swedish regions became considerably more similar in terms of industrial composition between 1900 and 1960. In order to answer the latter two questions I estimate a version of the Midelfart-Knarvik et al. (2000) model, in section 4.2, where the interaction between region and industry characteristics determine the location of industries. OLS estimations show that the distribution of industries across Swedish regions, between 1900 and 1960, was jointly determined by factor endowments and market potential. More specifically, industries that intensively used inputs from the agricultural and forestry sectors located in regions abundant in agricultural land and woodland, although the importance of the endowment of agricultural land diminished from the 1930s. The distribution of iron ore did not have a substantial effect in attracting industries during this period. From the 1940s and onwards, but not before, the regional endowment of human capital was important in attracting skill-intensive industries. Except for in the interwar period, industries with increasing returns to scale located in regions with high market potential. Industries where a large share of output was sold as inputs to other industries located in regions with high market potential, this effect being especially important in the interwar period. But high market potential seems to have deterred industries where intermediate consumption was high, although the magnitude of this negative relationship declined over time. In addition, to identify the causal effects of market potential I use an instrumental variables approach (two-step GMM) where I use predetermined levels of market potential and the sum of distances between regional capitals as exogenous instruments of market potential, confirming the results obtained by OLS. Standardized coefficients indicate that market potential was a more important determinant of industry location than factor endowments over this period.

The rest of this paper is organized as follows: In section two I discuss the model of Midelfart-Knarvik et al. (2000) and describe the interactions between region and industry characteristics that constitute the core of the model. The construction of the dataset is described in section three, with additional details provided in Appendix A. Section four deals with the descriptive account of the regional distribution of manufacturing employment, the dispersion of industries, and regional specialization as well as an econometric analysis of the determinants of industry location. In section five I provide some concluding remarks.

2 A Model of Industry Location

Location theories, going back to von Thünen (1826), Christaller (1933), and Lösch (1940), emphasize the interaction between the characteristics of different locations and the characteristics of economic activities. Consequently, in order to examine the determinants of industry location I make use of a general equilibrium model, derived by Midelfart-Knarvik et al. (2000), that nests interactions between region and industry characteristics derived from the Heckscher-Ohlin and New Economic Geography theories, such that in equilibrium the regional distribution of industries is simultaneously determined by both sets of factors[5]. Regions differ in their proximity to markets and their endowments of different factors, such as agricultural land and human capital whereas industries value proximity to markets differently and use factors with different intensities. As trade between regions is costly both the supply arguments of the Heckscher-Ohlin theory and the demand argument in New Economic Geography influence the location of industries. The general equilibrium nature of the model implies that although all industries may want to locate in the region most proximate to markets, this cannot be a stable equilibrium. Instead, the relative intensities in the use of different factors determine which industries locate where. For example, an industry with substantial backward linkages will locate in a region with high demand and an industry that intensively uses skilled labour will locate in a region with abundant human capital. The

[5]See Midelfart-Knarvik et al. (2000) or the supplementary material in Klein and Crafts (2011) for a thorough discussion of the model. Note that the original model considers countries, not regions.

model of Midelfart-Knarvik et al. (2000) has recently been applied by a number of economic historians (Crafts and Mulatu, 2005, 2006; Wolf, 2007; Klein and Crafts, 2011; Betran, 2011; Martinez-Galarraga, 2012) that all rely on slight variations of a general reduced-form of the model that can be expressed as:

$$ln(S_{ik}) = Z\alpha + \sum_j \beta^j(y_k^j - \gamma^j)(x_i^j - \chi^j) + \varepsilon_{ik} \quad (1)$$

where the share S of national employment in industry k in region i is jointly determined by region (x) and industry characteristics (y) after controlling for size-differences through the vector of regional size controls Z. β^j, γ^j, and χ^j are the parameters of interest. By expanding equation (1) I obtain an equation that lends itself to estimation:

$$ln(S_{ik}) = Z\alpha + \sum_j \left(\phi_1^j y_k^j x_i^j + \phi_2^j x_i^j + \phi_3^j y_k^j\right) + \varepsilon_{ik} \quad (2)$$

where the terms in summation are the regional characteristics, industry characteristics, and the corresponding interactions. $\phi_1 = \beta^j$ is the coefficient for the interaction between region (x) and industry (y) characteristic j, $\phi_2 = -\beta^j\gamma^j$ and $\phi_3 = -\beta^j\chi^j$ are the coefficient for region and industry characteristic j respectively, and $\varepsilon_{ik} \sim iid(0, \sigma_\varepsilon^2)$. The relative magnitude and significance of the coefficient of the interaction between, for example, the regional endowment of human capital and the use of skilled labor in an industry provides a measure of how important these factors were in determining the distribution of industries across Swedish regions.

In the time dimension there are three sources of variation in this specification. (1) The regional endowments can change, e.g. there is an increase in the endowment of human capital in region i (2) Over time industry characteristics may change, e.g. industry k becomes more intensive in the use of skilled labor (3) The responsiveness of regional industry shares to the interactions change, e.g. the interaction between the endowment of human capital and skilled-labor intensity in shaping the spatial distribution of industries becomes more important. To exploit the variation in the time dimension this suggests that the model should be implemented by using time-variant characteristics and estimated for repeated cross-sections.

As neither theory nor the model itself dictate which interactions are potentially important determinants of industry location I proceed by describing and motivating the seven interaction effects that constitute the core of the model in the next section.

2.1 Region and Industry Characteristics

In this section I describe and motivate four interactions corresponding to the arguments of the Heckscher-Ohlin theory and three interactions corresponding to the arguments in the New Economic Geography. The region and industry characteristics, and corresponding interactions are summarized in Table 1.

2.1.1 Heckscher-Ohlin Interactions

The distribution of natural resources and especially the distribution of coal has been emphasized as an important historical determinant of industry location as proximity to coalfields lowered the costs of steam power (Ross, 1896; Lee, 1971; Pollard, 1981; Stobart, 2000; Balderston, 2010). The dearth of coal in Sweden, with only scant supplies in north-western Scania, and the centrality of water power implies that the impact of coal on industrial location was neglible[6]. Instead, as early industrialization in Sweden mainly was confined to extraction and refinement of iron ore and wood, and exports of oat (see, e.g., Schön, 2010) this nonetheless suggests that the distribution of natural resources constituted a potentially important determinant of industry location. Hence I include three interactions based on central natural resource endowments (arable land, woodland, and iron ore). From the early 1900s and onwards industries became more skill-intensive; between the 1920s and the 1960s the average share of white-collar employment in total employment increased from roughly 9 percent to 26 percent[7]. Thus, the increasing importance of human capital motivates the inclusion of an interaction based on the regional endowment of human capital.[8]

The first interaction regards the regional endowment of agriculture, measured as the share of arable land in total regional area. The corresponding industry characteristic is the share of inputs from the agricultural sector in gross output for each industry. All things equal, an industry that intensively uses inputs from the agricultural sector should want to locate in a region with a comparative advantage in the production of agricultural goods. As Sweden by the early 20th century still was mainly an agrarian country, with roughly two-thirds of the population employed in agriculture (Krantz and Schön, 2007), this suggest that the distribution of agricultural land may have been important in attracting industries using inputs from this sector. Indeed, historical evidence indicates that this factor may have been important as there was considerable variation in arable land across Swedish regions due to underlying variation in soil quality and precipitation. For example, the share of arable land in regional area in the vast northernmost county of Norrbotten was 0.3 percent and in the southernmost county of Malmöhus (*'the Granary of Sweden'*) arable land constituted 72 percent of the regional area. In the early 1900s the county of Norrbotten had no registered employment in the food industry whereas the province of Malmöhus housed more than a third of national employment in that industry[9].

The second interaction considers the relative abundance of wood, measured as the share of woodland in total regional area. The corresponding industry characteristic is inputs from the forestry sector as a share of gross output. Forestry and wood-related industries was of imperious importance to early industrialization in Sweden and was continually important due to the expansion of the paper and pulp industries in the 20th century (Schön, 2010). In addition, in an European comparison the international distribution of woodland is of marked importance to Sweden. On average Western European countries were endowed with roughly 0.4 hectares of wood for every inhabitant whereas Sweden had roughly 4 hectares per inhabitant in the early 1900s. Similarly, roughly half of Sweden's geographical area consisted of woodland whereas the average for the Western European countries was merely

[6]A brief review of the international differences in coal reserves makes this point clear: In 1913 the estimated reserves was 20 tons per head in Sweden, compared with 39,328 for the United States, 11,187 in Germany, and 433 in Spain (Betran, 2005). Similarly, in 1910 the total consumption of coal and coke in Sweden equaled 4,756,000 tons, out of which only 303,000 was produced domestically (Sundbärg and Åmark, 1914).

[7]Calculated based on data obtained from *BiSOS-D 1920* and *SOS: Industri 1960*, see section three.

[8]Although one of the central factors in the original Heckscher-Ohlin model, capital is excluded based on the assumption that regional capital markets were integrated by the early 20th century.

[9]Calculated based on data obtained from SMID, see Berger et al. (2012).

Heckscher-Ohlin

Region Characteristic	Industry Characteristic
Agricultural Land	Agricultural Inputs
(Share of Regional Area)	(Share of Gross Output)
Human Capital	White-Collar
(Population Enrolled in Sec. Schooling)	(Share of Total Employment)
Woodland	Forestry Inputs
(Share of Regional Area)	(Share of Gross Output)
Iron Ore	Mining Inputs
(Gross Output of Iron Ore)	(Share of Gross Output)

New Economic Geography

Region Characteristic	Industry Characteristic
Market Potential	Forward Linkages
(Distance-Deflated Sum of Regional GDPs)	(Intermediates in Gross Output)
Market Potential	Backward Linkages
(Distance-Deflated Sum of Regional GDPs)	(Sales to Industry in Gross Output)
Market Potential	Increasing Returns to Scale
(Distance-Deflated Sum of Regional GDPs)	(Mean firm size)

Table 1: Interactions Between Region and Industry Characteristics

a quarter (Örtenblad, 1914). This suggest that the regional distribution of woodland may have been and important determinant of industry location in Sweden.

The third interaction concerns the distribution of iron ore. With the introduction of the Thomas process in the 1890s the deposits of phosphoric iron ore in the Mining District of Central Sweden (*Bergslagen*) and in the northernmost region of Norrbotten served as the foundation for a renewed expansion of the iron industry (Söderberg, 1984; Olsson, 2007). To account for the endowment of iron ore I include the regional production of iron ore for each benchmark year which is interacted with the share of inputs from the mining industry in gross output for each industry.

The fourth interaction considers the regional endowment of human capital and the skill intensity, proxied by the share of white-collar employees in total employment, in each industry. Owing to the early introduction of compulsory public schooling, in 1842, the literacy rate in the late 19th century exceeded 90 percent (see, Ljungberg and Nilsson, 2009 and Schön, 2010). This seemingly invalidates the use of literacy rates as a measure of regional differences in human capital[10]. The regional human capital stock is therefore, similar to Betran (2011), proxied by the population enrolled in secondary schooling.

[10] Literacy rates are commonly used as a proxy for human capital in the literature, see, e.g., Martinez-Galarraga (2012). Also, note that recent attempts to estimate regional levels of human capital using age-heaping methods are unfeasible for Sweden as individuals did not themselves state their age in the censuses (Hippe and Baten, 2011).

2.1.2 New Economic Geography Interactions

The regional characteristic considered in all New Economic Geography interactions is market potential, i.e. the economic centrality of each region, measured as the distance-deflated sum of regional GDPs (see section 3.1). As Swedish regions are separated by vast distances the peripheral regions were potentially disadvantaged due to lower market potentials. In addition, recent findings suggest that smaller distances to large urban markets encouraged industrialization in rural areas during the late 19th century (Jonsson et al., 2009). This suggest that differences in market potential thus constitutes a potentially important determinant of industry location across Swedish regions.

The first New Economic Geography interaction regards market potential and increasing returns to scale, proxied by the mean firm size in each industry. The mean firm size should in this context be interpreted as the minimum efficient scale (Crafts and Mulatu, 2005). An industry with increasing returns to scale faces a trade-off between exhausting scale economies by locating in proximity to a large market and operating at an inefficient scale in order to supply smaller markets locally (Krugman, 1991a). Jörberg (1961) argues that poor communications and high distribution costs promoted the growth of small firms serving local markets in late 19th century Sweden. The consequent expansion of infrastructure led to an increased growth for larger firms in the first decade of the 20th century. The importance of scale economies in conjunction with market potential thus potentially became increasingly important during the breakthrough of industrialization in the early 20th century.

The other two New Economic Geography interactions considers market potential in conjunction with forward and backward linkages (Krugman and Venables, 1995). The interaction between market potential and forward linkages, measured as the share of intermediates in gross output, are based on the notion that firms will locate as to minimize inter-industry transportation costs. An industry where intermediates comprise a large share of gross output should tend to locate in a region where market potential, and the supply of intermediate goods, is high. Similarly, the importance of backwards linkages depend on the extent that an industry supplies other industries with intermediate goods. Backward linkages is measured as the sales to industry as a share of gross output. An industry where sales to industry constitute a large share of output should localize as to minimize transport costs with respect to its industrial customers. The growth of the Swedish manufacturing industry in the early 20th century produced an industrial structure with higher degree of linkages as exemplified by the growth of the mechanical workshops, mainly producing machines for other industries (Jörberg (1961)). This suggest that the importance of linkages may have been increasing from the early 20th century and onwards.

3 Data

In order to examine the determinants of industry location across Swedish regions I have constructed a unique dataset for the 24 Swedish NUTS-III regions[11], eight industries[12], and for the seven benchmark years 1900, 1910, 1920, 1930, 1940, 1950, and 1960. Additional

[11] Corresponding to counties or '*län*'.

[12] The industries are: (1) Metals & Machinery (2) Stone & Earthenware (3) Wood (4) Paper, Pulp, and Graphical (5) Food (6) Textiles & Clothing (7) Leatherware & Rubber (8) Chemicals.

information on sources and considerations taken when constructing the dataset is provided in Appendix A.

3.1 Region Characteristics

Data on regional employment was obtained from the recently constructed *Swedish Manufacturing Industry Database* (SMID), which in turn is based on the official industrial statistics published by Statistics Sweden[13] (see, Berger et al. 2012). The dependent variable S is the share of employment in industry k located in region i which is simply calculated as: $S_{ik} = e_{ik} / \sum_i e_{ik}$, where e denotes employment.

I have digitalized the regional characteristics by hand from various publications from Statistics Sweden. Data on regional population for each benchmark year has been collected from the population censuses, obtained from Statistics Sweden. Enrollment rates in secondary schooling has been obtained from *BiSOS-P: Undervisningsväsendet.* Arable land per region was obtained from the agricultural census, in *BiSOS-N: Jordbruk och Boskapsskötsel* and the *Statistical Yearbook of Sweden,* and has been divided by total regional area for each benchmark year. Similarly, the share of woodland is calculated analogously based on data obtained from the forestry statistics, in *BiSOS-Q: Skogsväsendet* and the *Statistical Yearbook of Sweden.* Data on the output of iron ore in each region was obtained from the *Statistical Yearbook of Sweden, 1940.* Each regional characteristic, with some minor exceptions (see Appendix A), has been obtained for each of the seven benchmark years.

Regional market potential can primarily be estimated in two ways. Redding and Venables (2004) have shown that market potential can be estimated using inter-regional trade data and a general gravity model. As data on inter-regional trade flows is not available for Sweden for the period at hand the estimation of market potential follows the classical approach in Harris (1954) where market potential is estimated as the inverse distance-weighted sum of regional GDPs[14]. Market potential M for region i is estimated as:

$$M_i = D_{ii}^{-\delta} GDP_i + \sum_{j=1}^{n} D_{ij}^{-\delta} GDP_j \tag{3}$$

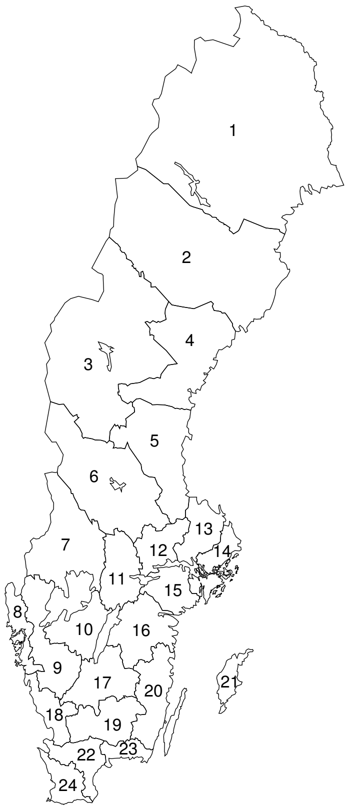
where D is the bilateral geodesic distance, i.e. as the crow flies, in kilometers between provincial capitals of region i and j and δ is a distance-decay parameter. Following Klein and Crafts (2011) own-region distance is approximated by:

$$D_{ii}^{-\delta} = \left[2/3\sqrt{A_i/\pi} \right]^{-\delta} \tag{4}$$

where A is the area of region i. Geodesic distances between the regional capitals of Sweden were obtained from an electronic atlas (see Appendix A). In addition, in order to calculate market potentials for the Swedish regions I need an estimate of δ. Here I follow the traditional assumption in the literature that $\delta = 1$ which is motivated by gravity estimates obtained from contemporary and modern trade flows (see, e.g., Head and Mayer, 2004). Recent

[13]For the years 1900 and 1910 the industrial statistics was published in *Bidrag till Sveriges Offentliga Statistik D: Fabriker och Manufakturer*, henceforth referred to as *BiSOS*-D, and for the years 1920-1960 in *Sveriges Officiella Statistik: Industri* henceforth referred to as *SOS: Industri.*

[14]For recent applications of this methodology see, e.g., Crafts (2005) and Schulze (2007). Regarding market potential estimated directly from regional trade flows see Wolf (2007).



1	Norrbotten
2	Västerbotten
3	Jämtland
4	Västernorrland
5	Gävleborg
6	Kopparberg
7	Värmland
8	Gothenburg & Bohus
9	Älvsborg
10	Skaraborg
11	Örebro
12	Västmanland
13	Uppsala
14	Stockholm
15	Södermanland
16	Östergötland
17	Jönköping
18	Halland
19	Kronoberg
20	Kalmar
21	Gotland
22	Kristianstad
23	Blekinge
24	Malmöhus

Notes: The regions correspond to NUTS-III regions or counties (*län*) as defined by Statistics Sweden prior to 1997.

Figure 1: The Regions of Sweden

estimates of GDP for the 24 (NUTS-III) regions of Sweden was obtained from Enflo et al. (2010), estimated according to the Geary and Stark (2002) methodology. I only include domestic regional GDPs in the empirical analysis due to the difficulties associated with proper weighting of foreign GDPs (see Schulze, 2007). In Appendix B I provide estimates based on market potential that include foreign markets and show that the main econometric results are similar although not identical. Foreign GDPs enter equation (3) in the same way as domestic regional GDPs and are deflated by the geodesic distances between each regional capital and the node (in parentheses) in Germany (Berlin), Poland (Warsaw), Norway (Oslo), Finland (Helsinki), France (Paris), Denmark (Copenhagen), the Netherlands (Amsterdam), the Czech Republic (Prague), Austria (Vienna), United Kingdom (London), and Belgium (Brussels). Geodesic distances were again obtained from an electronic atlas and GDP for foreign countries, in 1990 $GK, were obtained from Maddison (2003).

Figure 1 provides the names and locations of the 24 Swedish (NUTS-III) regions and Table 2 presents the estimated market potentials for each region normalized to the market potential of Stockholm, for each benchmark year. The county of Stockholm had the highest market potential in all benchmark years. In addition, the market potential of Stockholm

Region	1900	1910	1920	1930	1940	1950	1960
Blekinge	0.66	0.67	0.59	0.56	0.48	0.49	0.46
Gävleborg	0.65	0.61	0.58	0.55	0.49	0.51	0.49
Gothenburg & Bohus	0.86	0.81	0.84	0.84	0.73	0.77	0.70
Gotland	0.52	0.51	0.47	0.45	0.41	0.43	0.40
Halland	0.65	0.65	0.60	0.58	0.50	0.54	0.49
Jämtland	0.32	0.30	0.28	0.27	0.24	0.25	0.24
Jönköping	0.72	0.72	0.66	0.63	0.56	0.61	0.56
Kalmar	0.60	0.59	0.53	0.51	0.44	0.46	0.43
Kopparberg	0.61	0.58	0.54	0.52	0.46	0.49	0.47
Kristianstad	0.68	0.70	0.62	0.59	0.51	0.54	0.49
Kronoberg	0.65	0.66	0.59	0.56	0.49	0.52	0.49
Malmöhus	0.83	0.87	0.77	0.74	0.64	0.66	0.60
Norrbotten	0.21	0.20	0.19	0.18	0.16	0.17	0.16
Skaraborg	0.78	0.77	0.70	0.67	0.60	0.65	0.60
Södermanland	0.81	0.81	0.76	0.73	0.69	0.71	0.69
Stockholm	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Uppsala	0.85	0.82	0.80	0.78	0.74	0.76	0.74
Värmland	0.66	0.64	0.60	0.58	0.51	0.55	0.51
Västerbotten	0.29	0.27	0.26	0.25	0.21	0.23	0.22
Västernorrland	0.42	0.38	0.36	0.35	0.30	0.31	0.30
Västmanland	0.84	0.81	0.77	0.74	0.69	0.72	0.71
Älvsborg	0.70	0.68	0.66	0.65	0.57	0.62	0.57
Örebro	0.79	0.77	0.73	0.69	0.63	0.66	0.63
Östergötland	0.79	0.79	0.72	0.69	0.62	0.66	0.62
Coefficient of Variation	0.29	0.31	0.32	0.32	0.35	0.34	0.35

Sources and *notes:* Regional nominal GDPs were obtained from Enflo et al. (2010) and geodesic distances between regional capitals were obtained from an electronic atlas, see text and Appendix A for more information. Regional market potentials are normalized to the market potential of Stockholm for each benchmark year.

Table 2: Market Potential of Swedish Regions, 1900-1960

was increasing relative to other regions which corresponds to the increasing importance of Stockholm in the national economy during this period (see, Enflo et al., 2010). Looking at each cross-section it is apparent that the northern regions (Norrbotten, Västernorrland, Jämtland, and Västerbotten) had low market potentials, due to low regional GDPs and to the vast distances to the central markets of Gothenburg and Stockholm. Conversely, the regions located in proximity to Stockholm and Gothenburg (Uppsala, Södermanland, Västmanland, and Älvsborg) generally had high market potentials. Turning to the coefficient of variation, presented in the bottom row of Table 2, there is an indication that there was an increasing dispersion of market potentials between 1900 and 1960 suggesting that this factor may be important in explaining the emerging divide between northern and southern Sweden.

Industry	Agricultural Inputs	Wood Inputs	Iron Inputs	White Collar	Firm Size	Intermed. Inputs	Sales to industry
Metals & Machinery	0.00	0.00	0.42	0.17	51	0.58	0.43
Stone & Earthenware	0.00	0.01	0.00	0.10	40	0.27	0.11
Wood Industries	0.00	0.39	0.00	0.09	28	0.63	0.16
Paper, Pulp, and Graphical	0.00	0.25	0.00	0.16	62	0.60	0.20
Food Industries	0.49	0.00	0.00	0.17	39	0.78	0.20
Textiles & Clothing	0.04	0.00	0.00	0.12	76	0.64	0.38
Leatherware & Rubber	0.00	0.00	0.00	0.13	51	0.75	0.43
Chemical Industries	0.00	0.06	0.00	0.25	49	0.49	0.34

Sources and *notes:* Agricultural inputs, wood inputs, and iron inputs are all calculated as shares of gross output in each industry. White-collar is the average share of white-collar employees in total employment, 1900-1960. Firm size is the average mean firm size between 1900 and 1960. Intermediate inputs and sales to industry are calculated as shares in gross output. All characteristics, except firm size and white-collar, are based on the technical coefficients in 1913 from Bohlin (2007). Mean firm size is calculated based on data in SMID (see Appendix A) and white-collar is calculated based on data obtained from *BiSOS-D* and *SOS: Industri.*

Table 3: Industry Characteristics

3.2 Industry Characteristics

Industry characteristics are primarily based on Bohlin (2007) who derives an input-output table for the Swedish manufacturing industry disaggregated into nine industries[15] based on the technical coefficients in 1913. The intensity in use of agricultural inputs is calculated as domestically produced agricultural goods as a share of gross output in each industry. The intensity in use of inputs from the forestry sector is similarly calculated as the inputs from this sector as a share of gross output in each industry. Since only one industry (Metals & Machinery) used iron ore as an input in production, and data on consumption of ore is not readily available, the use of iron ore is approximated by the share of own-industry inputs for this industry. As time-variant data is not available regarding the input-output relationships these industry characteristics are solely based on the technical coefficients in 1913. To proxy for the skill-intensity of each industry I have obtained data on the number of white-collar employees and total employment in each industry from the industrial statistics (*BiSOS-D* and *SOS: Industri*). The skill-intensity is proxied by the share of white-collar employment in total employment for each industry. Time-varying data is available from 1920 and onwards (see Appendix A). Regarding the industry characteristics derived from the New Economic Geography, forward linkages is calculated as the gross output net value added in each industry based on the technical coefficients in Bohlin (2007). Backward linkages are similarly calculated as the share of output used as inputs in other industries. To proxy for increasing returns to scale at the industry level I have calculated the mean firm size, corresponding to the minimum efficient scale, for each industry and benchmark year based on data obtained from SMID (see, Berger et al., 2012).

Table 3 summarizes the industry characteristics for the eight industries. In the first column it is evident that only the food, textiles, and clothing industries used inputs from the agricultural sector. But the former dwarfs the latter in terms of quantities as almost

[15]In addition to the eight industries included in this paper Bohlin (2007) also include the power industry.

half of the gross output of the food industry consisted of agricultural inputs. Wood inputs was primarily used by the wood, paper, and pulp industries, whereas these inputs were of marginal importance in the stone, earthenware, and chemical industries. As previously mentioned metals and machinery was the only industry that used iron ore as an input in production. The most skill-intensive industry was the chemical industry where white-collar employment on average constituted 25 percent of total employment, whereas the least skill-intensive were the stone, earthenware, and wood industries where white-collar employment constituted roughly 10 percent of total employment. The textiles and clothing industries had the largest average firm size of 76 employees, whereas the average firm size was considerably smaller in the wood industries with an average size of 28 employees per firm. The food industry had the highest share of intermediate inputs where the bulk of inputs was sourced from the agricultural sector. The metals, machinery, textiles, clothing, leatherware, rubber, and chemical industries all sold a large share of gross output as inputs, as the sales to industry as a share of gross output ranged between 34 and 43 percent for these industries. Given the manifest variation in these industry characteristics it is likely that industries valued regional characteristics differently thus encouraging different localization patterns.

4 Empirical Analysis

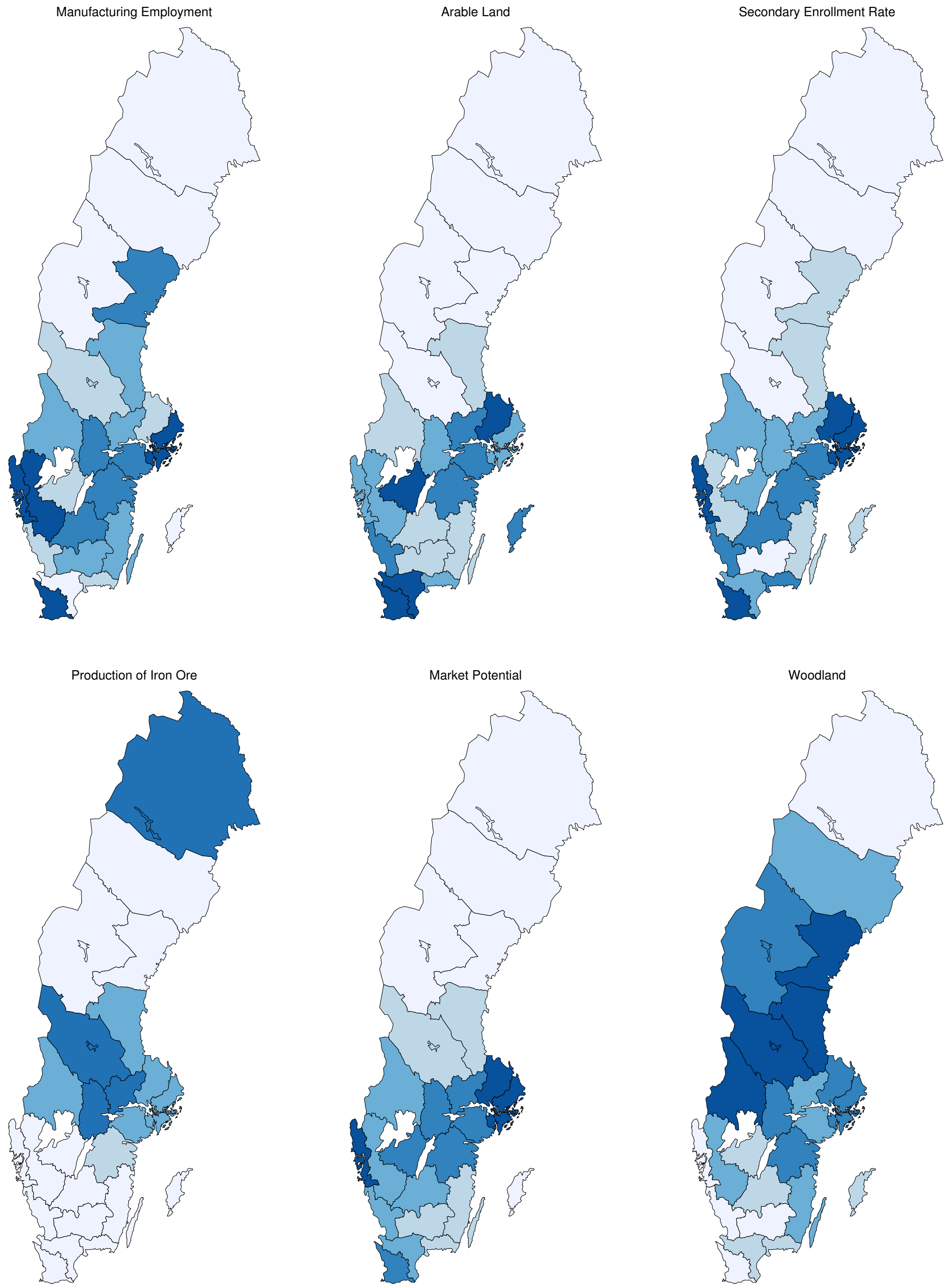
I take an unabashedly descriptive approach in the first part (4.1) of the empirical analysis, motivated by the fact that I am working with a novel dataset, to provide an account of the changes in regional specialization and industry location across Swedish regions between 1900 and 1960. In the second part (4.2) I proceed with an econometric treatment of the determinants of industry location for each of the seven benchmark years by estimation of the model put forth in section two.

4.1 The Regional Distribution of the Manufacturing Industry

In this section I begin by documenting the correspondence between the five regional characteristics (arable land, woodland, iron ore, human capital, and market potential) and the distribution of regional manufacturing employment for the benchmark year 1930. I then proceed with an examination of the trends in the regional distribution of manufacturing employment, industrial concentration, and regional specialization respectively.

4.1.1 Endowments, Markets, and Manufacturing in 1930

To provide a broad-brush impression of the extent to which regional manufacturing employment coincided with the regional characteristics Figure 2 projects these variables, divided into quintiles, for the benchmark year 1930. The northern regions were evidently disadvantaged as measured by all regional characteristics except for those capturing natural resource endowments (wood and iron). Human capital, market potential, agricultural land, and manufacturing employment were all at considerably lower levels than in the regions of mid- and southern Sweden. This constitutes a potential source of the north-south divide, in terms of manufacturing employment, that emerged during this period. Regions with an apparent comparative advantage in agriculture, e.g. the counties of Uppsala, Kristianstad, and Skaraborg were among the least industrial regions although they were located in proximity



Sources: Data on manufacturing employment was obtained from the *Swedish Manufacturing Industry Database* (SMID), see text for further information, and is normalized by regional population. Market potential is based on regional GDPs obtained from Enflo et al. (2010) and regional characteristics are based on data obtained from various publications from Statistics Sweden, see text for further information.
Notes: All variables are presented for the benchmark year 1930. Each variable is divided into quintiles and is increasing from light to dark.

Figure 2: Quintiles of Regional Characteristics & Manufacturing Employment, 1930

to the large urban markets of Stockholm, Malmö, and Gothenburg respectively. The major exception to this apparent negative relationship between the share of arable land and the level of industrial activity was the region of Malmöhus, richly endowed with both industry and agriculture. Both market potential and human capital, as proxied by the enrollment rate in secondary schooling, coincided with high levels of manufacturing employment. The most well-endowed regions in terms of human capital are the metropolitan regions (Stockholm, Gothenburg & Bohus, and Malmöhus), these regions also had the highest levels of industrial activity. The one major exception to this relationship seems to have been the region of Älsvborg where human capital levels were low although the level of industrial activity was on par with that in the metropolitan regions. The dominance of the textiles and clothing industry with modest usage of human capital (Table 3) in this region may partly explain this anomaly. The endowment of iron ore and woodland does not seem to have been associated with high levels of manufacturing employment. Instead, the regions that were well-endowed with these natural resources were less industrial than other regions.

But these broad-brush impressions are limited in that they only consider one cross-section. Consequently, I proceed by examining the distribution of regional manufacturing employment for the seven benchmark years in the next section.

4.1.2 How Concentrated Was Manufacturing Employment?

To examine the regional distribution of manufacturing employment between 1900 and 1960 I make use of two traditional measures of inequality, namely the Gini coefficient and the Theil index. The Gini (1912) coefficient is a traditional measure of inequality where a coefficient of zero corresponds to an even regional distribution whereas a coefficient of one means that all manufacturing employment is concentrated in one region. Following Deaton (1994) the Gini coefficient (G) can be calculated as:

$$G = \frac{N+1}{N-1} - \frac{2}{N(N-1)\mu}\sum_{i=1}^{N} r_i e_i \tag{5}$$

where N is the number of regions, μ is the average employment, and r and e are the rank and employment of region i respectively. The Theil (1967) index is a measure of inequality, derived from a general entropy measure, where an index of zero corresponds to an even distribution and the upper bound, corresponding to perfect inequality, is the natural logarithm of the number of regions (i.e. $ln(24) = 3.18$). The Theil index (T), using the same notation as in (5), is calculated as:

$$T = 1/N \sum_{i=1}^{N} \frac{e_i}{\mu} ln\left(\frac{e_i}{\mu}\right) \tag{6}$$

These two measures and their corresponding bootstrapped standard errors, using 100 replications, are presented in Table 3. The regional distribution of manufacturing employment became slightly more equal between 1900 and 1920 as both the Gini and Theil decreased marginally. But during the interwar period the manufacturing industry became increasingly concentrated as the Gini and Theil increased from 0.41 and 0.28 to 0.44 and 0.33 respectively. Conversely, in the postwar period there was rapid convergence in employment levels as the Gini and Theil decreased from 0.33 and 0.44 to 0.39 and 0.26 respectively

Index	1900	1910	1920	1930	1940	1950	1960
Gini Coefficient	0.42	0.41	0.41	0.43	0.44	0.42	0.39
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
Theil Index	0.29	0.28	0.28	0.31	0.33	0.29	0.26
	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)

Sources: Data on regional employment was obtained from the *Swedish Manufacturing Industry Database* (SMID), based on the official industrial statistics published by Statistics Sweden, see section three and Appendix A for details.
Notes: The Gini coefficient and Theil index are calculated based on the 24 (NUTS-III) regions of Sweden. Bootstrapped standard errors, using 100 replications, are in parentheses.

Table 4: Regional Distribution of Manufacturing Employment, 1900-1960

between 1940 and 1960. This gives rise to a bell-shaped pattern of industry concentration as previously found, for example, in the United States (Kim, 1995) and France (Combes et al., 2011). It should be noted that these indices are associated with relatively high standard errors such that the indices in 1900 and 1960 are within the limits of the degree of concentration in 1940 $\pm$ one standard error that suggests that the degree of concentration was of a rather small magnitude.

What explains the evolution of industry concentration? One peculiarity of industrialization in Sweden was its rural character that promoted a dispersion of industries, as evident in column two of Table 4 (Söderberg and Lundgren, 1982). The electrification of the manufacturing industry, initiated during the First World War, undermined the relative advantage of rural localization and promoted a relocation to urban areas that could bear the large fixed costs invoked by electrification (Schön, 2010). By the early 1920s electricity constituted roughly three-quarters of the aggregate motive power in manufacturing and the hedonic price of electrical motors decreased by roughly 13 percent over this decade, promoting a deeper penetration of electrification (Schön, 2000; Edquist, 2010). Industries that were chiefly located in rural areas, such as the sawing mills in the northern regions, lagged in electrification, which was achieved only by the 1950s (Schön, 2000)[16]. This promoted a drift in the location of industries toward the metropolitan regions that increased their share of manufacturing employment by roughly 7 percentage units between 1900 and 1940 (Berger et al., 2012), corresponding to the increased inequality in the interwar period in Table 4. The diffusion of manufacturing from the 1940s and onwards occurred as employment was shifted from the metropolitan regions of Gothenburg and Stockholm to adjacent regions in mid-central Sweden. The dispersion of employment in the postwar period resulted in a regional distribution of the manufacturing industry that was more equal, in terms of the Gini coefficient and Theil index, in the 1960s than during any previous decade of the 20th century. These aggregate trends in the distribution of manufacturing employment also closely corresponds to the evolution of regional income per capita where there was a slowdown in the long-run trend of convergence in the first half of the 20th century, with slightly increasing dispersion in the interwar period, and strong convergence following the Second World War (Henning et al., 2011).

[16]It is significant that manufacturing employment was roughly five times as urban in southern Sweden as in the north in 1940, with ratios of urban to rural employment of 0.5 and 2.5 respectively (calculated based on data obtained from SMID, see Appendix A for more information).

But the aggregate trends may obscure different concentration patterns across industries. As the concentration trends of individual industries likely varied over this period I proceed by an examination of the dispersion of the eight included industries in the next section.

4.1.3 How Concentrated Were Industries?

In order to examine the dispersion of individual industries I have calculated the coefficient of variation of regional manufacturing employment for each industry[17], where an increase in the coefficient of variation corresponds to an increase in the dispersion of regional employment in an industry. Figure 3 presents the coefficient of variation for the eight industries and for each benchmark year. Here both the level, i.e. the degree of concentration, and its evolution over time are of interest. Apparently, comparing the industries in Figure 3, there was considerable variation in both concentration levels and trends, although four broadly different trajectories seem to stand out.

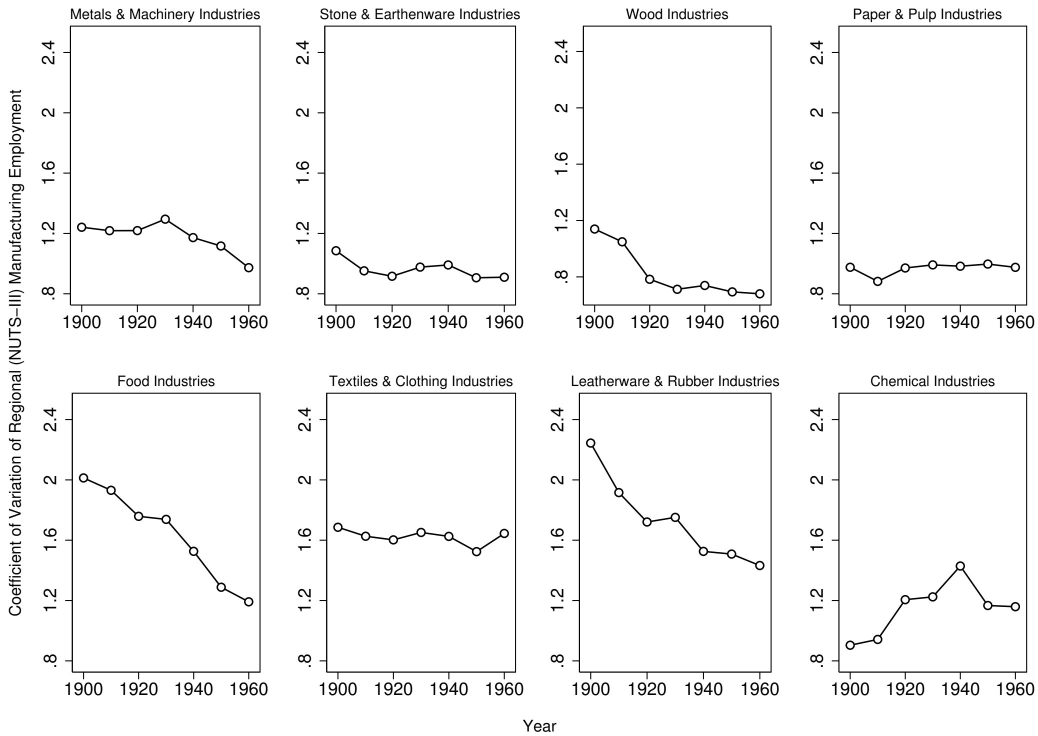
The regional distribution of the paper, pulp, textiles, and clothing industries remained remarkably stable between 1900 and 1960. Thus, the increasing demand for paper in the modern industrial society that led to an expansion and consolidation of the paper and pulp mills in the 1920s in the northern regions seems to have had neglible effects on the concentration of these industries (Schön, 2010). The textiles and clothing industry was considerably concentrated, although the apparent stability cloaks the rise of the county of Älsvborg as the most prominent producer of textiles.

Metals, machinery, stone, earthenware, and the wood industries became gradually more dispersed over the period. But whereas the metals and machinery industries became more dispersed mainly from the 1940s and onwards the stone, earthenware, and wood industries appear to primarily have diffused in the early 1900s. The diffusion of the metals and machinery industries was mainly due to an even distribution of the mechanical workshops, a central industry in the Second Industrial Revolution in Sweden (Berger et al., 2012). The diffusion could have been driven by the need to locate in proximity to industrial customers as this industry primarily produced inputs, such as machines and tools, for other industries (Jörberg, 1961). The stone, earthenware, and wood industries were to a large extent dependent on natural resources suggesting that the diffusion may have been related more to the decreases in transportation costs and expansion of infrastructure over this period.

The most radical changes occurred in the degree of concentration of the food, leatherware, and rubber industries that were the most concentrated in the early 1900s but became considerably dispersed during the first half of the 20th century. These industries depended on substantial intermediate consumption (Table 3) suggesting that this dependence on intermediate products became less of a constraint on localization over time as the food industry became increasingly focused on domestic regional markets.

The chemical industry followed a somewhat idiosyncratic trend, mirroring the evolution of aggregate manufacturing employment documented in the previous section, as the degree of concentration traced out a bell-shaped evolution between 1900 and 1960. This industry was the most skill-intensive industry, as proxied by the share of white-collar employment (Table 3), and quite dispersed in the early 1900s but became increasingly concentrated to the region of Stockholm, well endowed with human capital and with an abundant supply of

[17] The coefficient of variation is calculated as the unweighted standard deviation of employment for the 24 (NUTS-III) regions of Sweden divided by the average regional employment in each year



Sources: Data on regional employment was obtained from the *Swedish Manufacturing Industry Database* (SMID), based on the official industrial statistics published by Statistics Sweden, see section three and Appendix A for details.
Notes: The coefficient of variation is calculated as the unweighted standard deviation of regional manufacturing employment divided by average manufacturing employment for each benchmark year.

Figure 3: Concentration of Industries (CV), 1900-1960

producer services, during the interwar period. From the 1940s and onwards the dispersion coincided with the relative decline of manufacturing in the largest cities and metropolitan regions (Berger et al., 2012). The trajectory of this industry is also broadly consistent with recent evidence regarding the different needs over the industrial life-cycle where early growth occurs in diverse large urban regions and then diffuse to specialized low-cost areas (Neffke et al., 2011).

As the degree of concentration differed across industries this suggests that regions were specialized. In addition, the different trends in industry concentration further suggest that regional specialization changed between 1900 and 1960, which I turn to in the next section.

4.1.4 How Specialized Were Regions?

In this section I examine the degree of regional specialization by means of calculating Krugman (1991a) specialization indices. Let E denote the level of employment in industry i in region j and k, then the Krugman specialization index (KSI) is calculated as:

$$KSI_{jk} = \sum_{i=1}^{n} \mid \frac{E_{ij}}{E_j} - \frac{E_{ik}}{E_k} \mid, \; KSI \in [0, 2] \tag{7}$$

If the index is equal to zero the regions j and k are completely de-specialized, and conversely, if the index equals two, the regions are completely specialized. The indices are calculated based on 24 (NUTS-III) regions and eight manufacturing industries. The unweighted and weighted average of the Krugman indices are presented in Table 5.

Evidently Swedish regions became considerably less specialized between 1900 and 1960, as both the unweighted and weighted Krugman indices decrease from 0.80 and 0.73 to 0.50 and 0.48, or by roughly 35 and 40 percent respectively. But there are divergent intra-period dynamics. From the early 1900s, regional specialization decreased until the interwar period, implied both by the unweighted and weighted Krugman index. In the interwar period regional specialization flattened out, and even increased slightly when focusing on the weighted index. In the postwar period there is again convergence in industrial composition such that the level of regional specialization is lower in 1960 than during any previous decade of the 20th century. Seen over the entire period Swedish regions thus became increasingly more alike in terms of industrial composition.

In terms of geographic specialization there was an evident north-south divide where the northern regions were the most specialized throughout the period[18]. The average Krugman specialization index of the northern regions of 1.28 in 1900 imply that 64 percent of total employment in these regions would have to be shifted to other industries to converge with the the average industry composition. This relates to the uneven distribution of natural resources such as wood and iron ore (see Figure 2) where the industrial composition of the northern regions was biased toward traditional industries less prevalent in the southern regions. This also potentially constitutes a source of regional lock-in into industrial trajectories with little scope for branching out into more dynamic industries, promoting the relative stagnation in these regions (Neffke et al., 2011).

The evolution of regional specialization can readily be compared with the experience in other countries. In the United States, Kim (1995) shows that regional specialization increased from the mid 19th century until the outbreak of the First World War, flattened out in the interwar period, and continued to decrease in the postwar period. In Spain, regional specialization increased until the outbreak of the First World War, increased during during the interwar period, and decreased after the Second World War (Betran, 2011)[19]. Thus, although the trends as seen over the entire period are similar (convergence) there seems to have been somewhat different subtleties to the national trajectories of regional specialization.

Summing up the descriptive section on the regional distribution of the manufacturing industry across Swedish regions it is evident that the distribution of employment followed a bell-shaped pattern with rising inequality in the interwar period, that industries differed both in their levels of concentration as well as in their trends over the period, and that

[18] The unweighted average KSI, of the northern regions of Norrbotten, Jämtland, Västerbotten, and Västernorrland was 1.28, 1.09, and 0.66 in 1900, 1930, and 1960 respectively.

[19] The Krugman specialization index in the United States, calculated based on two-digit SIC industries was 0.66 in 1958 (Kim, 1995), 0.67 for Spain in 1955 (Betran (2011)), whereas it was 0.50 in Sweden in 1960. Corresponding figures for 1900 are 0.75 for the US, 0.60 (1893) for Spain (Martinez-Galarraga (2012)), and 0.80 for Sweden.

	1900	1910	1920	1930	1940	1950	1960
Unweighted Average	0.80	0.78	0.72	0.72	0.65	0.54	0.50
Weighted Average	0.73	0.70	0.62	0.64	0.58	0.51	0.48

Sources: Data on regional employment was obtained from the *Swedish Manufacturing Industry Database* (SMID), based on the official industrial statistics published by Statistics Sweden, see section three and Appendix A for details.
Notes: The Krugman indices are calculated for the 24 Swedish (NUTS-III) regions and eight industries. The weighted average index is weighted by regional employment.

Table 5: Krugman's Index of Regional Specialization, 1900-1960

regions were specialized, although there was strong convergence between 1900 and 1960. What accounts for these changes? In the next section I proceed with estimations of the model of Midelfart-Knarvik et al. (2000), discussed in section two, that explains industry location and regional specialization based on the interactions between region and industry characteristics presented in Table 3.

4.2 Econometric Results

In this section I provide a brief discussion of the appropriate estimation techniques when estimating the model of Midelfart-Knarvik et al. (2000), as described in section two. In section 4.2.2 and 4.2.3 I present the results of OLS and two-step GMM estimations respectively. These sections also provide a discussion of the *statistical* results. Since the estimated parameters cannot be compared directly, as the variables are measured in different units, I standardize the coefficients from the two-step GMM estimations to evaluate the relative importance of each interaction in section 4.3. In this section I also provide a discussion of the economic results in their historical context.

4.2.1 Estimation Techniques

Estimation of equation (2) can either be performed by including the region and industry characteristics (Crafts and Mulatu, 2005; Martinez-Galarraga, 2012), or by using region and industry dummies as controls (Wolf, 2007; Klein and Crafts, 2011). All results in the rest of this paper was estimated using both characteristics and dummies, with similar results (see Appendix B). Thus, I focus on the following version of equation (2) where dummies replace the size controls and the region and industry characteristics:

$$ln(S_{ik}) = \alpha + \delta_i + \theta_k + \sum_j \beta^j x_i^j y_k^j + \varepsilon_{ik} \tag{8}$$

where S is the share of industry k located in region i, δ_i and θ_k are a set of region and industry dummies that replace the size controls and characteristics in equation (2), $x_i^j y_k^j$ is the interaction between region and industry characteristic j and β_t^j is the set of parameters to estimate (see Table 1).

Klein and Crafts (2011) emphasize that there are three potential problems when estimating equation (8) where the latter two has not been taken into account in the previous literature. First, there may be a problem of heteroscedasticity across regions and industries. Second, since the sample is constructed for regions, there may be an unobserved cluster-effect *within* regions. Third, market potential and the associated interactions are likely endogenous which implies that the estimates of these interactions are inconsistent. The solutions to the first two problems are straightforward following White (1980) and by estimating the equation using cluster-robust errors. The problem of endogeneity calls for an instrumental variables approach, described in detail in section 4.2.3.[20]

4.2.2 Baseline Estimations (OLS)

Equation (8) is estimated by OLS using the White (1980) correction to account for heteroscedastic errors[21]. The results are presented in Table 6. The estimated models explains roughly 60 percent of the variation in regional shares of manufacturing employment, with slightly increasing R-squared over the period which is very similar to that obtained when using the same methodology based on data for other European countries (see, e.g., Crafts and Mulatu (2005); Martinez-Galarraga (2012)).

In the top panel the three interactions corresponding to the New Economic Geography are presented. The mean firm size-interaction is only significant in 1900 and 1910. Forward linkages are similarly most significant in the early 20th century but still significant in 1940 and 1950. The estimated coefficient is negative throughout the whole period which suggests that industries with high intermediate consumption were drawn toward regions with low market potential. The interaction capturing backward linkages is significant throughout the entire period and becomes increasingly significant in the interwar period. This suggests that industries that sold a large share of output as inputs to other industries tended to locate in regions with high market potential.

In the bottom panel of Table 6 the interactions corresponding to the Heckscher-Ohlin theory are presented. The regional endowment of wood seems to have been important in determining the location of industries intensive in the usage of inputs from the forestry sector as the interaction is strongly significant throughout the period. Whereas wood seems to have remained an important determinant of localization the regional endowment of agricultural land seems to have declined in importance over time. The interaction between arable land and agricultural inputs is significant in the early 20th century but from the 1930s it is insignificant. To the contrary the endowment of human capital seems to have become important only in the latter part of the period. From the 1940s the endowment of human capital seems to have been important in attracting industries intensive in white-collar labor as the estimated coefficient is significant. The iron ore-interaction is significant in the interwar and postwar periods, suggesting that industries were drawn toward regions rich in iron ore during these decades.

But since these estimates may be biased due to the presence of endogeneity I proceed in the next section with an instrumental variables approach.

[20] One option would be to use panel data techniques, but pooling data is not desirable since this would amount to assuming constant parameters over time. Something that is unlikely given the changes in industry location over this period (see section 4.1).

[21] This is motivated by a Breusch-Pagan test that leads to rejection of the null of homoscedastic errors at a 95 percent significance level for all years. OLS estimates using cluster-robust errors are provided in Appendix B.

VARIABLES	(1) 1900	(2) 1910	(3) 1920	(4) 1930	(5) 1940	(6) 1950	(7) 1960
	ln(Share of manufacturing industry i in region j)						
New Economic Geography							
Increasing Returns to Scale	0.086*	0.045**	0.001	-0.000	0.002	0.001	0.000
*(Market Potential*Mean Firm Size)*	(0.046)	(0.022)	(0.003)	(0.002)	(0.002)	(0.001)	(0.000)
Forward Linkages	-5.006**	-3.535**	-0.601*	-0.642	-0.345*	-0.167*	-0.059
*(Market Potential*Share of Intermediates)*	(2.310)	(1.384)	(0.323)	(0.427)	(0.204)	(0.098)	(0.039)
Backward Linkages	5.639*	3.164*	1.114***	1.166***	0.445*	0.293**	0.149**
*(Market Potential*Sales to Industry)*	(2.978)	(1.907)	(0.377)	(0.397)	(0.266)	(0.126)	(0.057)
Heckscher-Ohlin							
Endowment of Iron Ore	1.079	-0.223	1.193***	0.325*	0.105	0.283**	0.290**
*(Iron Ore Output*Share of Iron in GO)*	(1.974)	(0.699)	(0.372)	(0.189)	(0.170)	(0.136)	(0.143)
Human capital Endowment	3.512	2.801	4.158	0.350	0.935*	0.620**	0.220**
*(Pop. in Secondary Schooling*Share White-Collar)*	(7.705)	(5.405)	(3.797)	(0.556)	(0.515)	(0.270)	(0.110)
Endowment of Agriculture	8.485**	12.157***	8.052**	8.041**	3.536	2.355	2.703
*(Arable Land*Agricultural Inputs)*	(3.271)	(3.423)	(3.371)	(3.284)	(2.344)	(2.235)	(1.997)
Endowment of Wood	7.444***	7.884***	8.295***	13.142***	13.935***	9.595***	9.331***
*(Woodland*Forestry Inputs)*	(2.476)	(2.632)	(3.099)	(2.929)	(2.939)	(3.009)	(2.632)
Constant	-7.557***	-5.787***	-5.097***	-3.452**	-3.724***	-4.487***	-4.918***
	(2.108)	(1.449)	(1.518)	(1.444)	(1.023)	(1.283)	(1.212)
Industry Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	174	178	183	184	182	187	189
R-squared	0.571	0.582	0.583	0.582	0.590	0.640	0.653

Notes: The estimated equation is $ln(S_{ik}) = \alpha + \delta_i + \theta_k + \sum_j \beta^j x_i^j y_k^j + \varepsilon_{ik}$. Robust White (1980) standard errors in parentheses, where significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table 6: OLS with Heteroscedasticity-Robust Standard Errors

4.2.3 Instrumenting for Market Potential (Two-step GMM)

If market potential is endogenous, such that the OLS estimations have failed to identify the causal effect of market potential on regional shares of industry employment, the estimates of the market potential-interactions are biased since the regressors are correlated with the error term. Market potentials are likely endogenous as the relocation of firms to a region with high market potential in turn increases the market potential of that region, thus giving rise to a process of circular causation (see, e.g., Head and Mayer (2004)). To account for the potential endogeneity problems I adopt an instrumental variables approach, where I consider two sets of instruments. First, following Wolf (2007) and Martinez-Galarraga (2012) I use the predetermined market potentials, i.e. lagged one decade, as an instrument for current market potential. This effectively eliminates the simultaneity problem as lagged market potentials are not affected by contemporaneous location decisions. Second, following Klein and Crafts (2011) I use the sum of distances between all regional capitals as an additional instrument for market potential. Geographical distances between regional capitals can be considered exogenous as the cities were founded prior to industrialization. All instruments are created for the three New Economic Geography interactions respectively yielding a total of six instruments and three endogenous variables for each benchmark year. Since it is likely that the errors are heteroscedastic a GMM estimator is preferable to IV/2SLS, since the latter is, although unbiased, inefficient in the presence of arbitrary heteroscedasticity (Baum et al., 2003).

Consequently, equation (8) is estimated using two-step GMM simultaneously using the two sets of instrument. There are two problems to deal with. First, if the instruments are weak in the sense that they are not strongly correlated with the endogenous regressors the results may be even more biased than when using OLS. Second, the instruments may be endogenous. In the first stage regression I reject that instruments are weak based on the heteroscedasticity-robust F-statistic, at the 99 percent level, using the critical values in Stock and Yogo (2005), for all years. In addition, the instruments perform well in the first stage with Shea's (1997) partial R-squared generally above 98 percent (Table 7) suggesting that the instruments are strong. Since I have more instruments than the required moment conditions to estimate the three parameters in the model this provides the overidentification restrictions that allows me to test for the exogeneity of the instruments using Hansen's J-test. For all decades, aside from 1950 and 1960, I do not reject the null of valid instruments (Table 7). For 1950 and 1960 I reject the null of valid instruments at the 95 percent level which indicates that the instruments are not exogenous. I proceeded by backward elimination of instruments. When excluding the sum of distances for the increasing returns to scale-interaction I can no longer reject the null of valid instruments, while all coefficients remain essentially unchanged which suggests that the problem of weak instruments does not have important effects on inference for these years.

In the top panel of Table 7 the New Economic Geography interactions are presented. Except for the interwar period the interaction between increasing returns to scale and market potential is significant. The interactions capturing forward and backward linkages are significant throughout the period, where the former is more significant in the first two decades of the 20th century and the latter increases in significance in the interwar period. As in the OLS estimations the estimated parameter of the forward linkages is negative suggesting that industries where intermediates constituted a large share of gross output located in regions with low market potential.

	(1) 1900	(2) 1910	(3) 1920	(4) 1930	(5) 1940	(6) 1950	(7) 1960
VARIABLES	*ln(Share of manufacturing industry i in region j)*						
New Economic Geography							
Increasing Returns to Scale	0.080**	0.045**	0.001	0.001	0.003**	0.001*	0.001*
*(Market Potential*Mean Firm Size)*	(0.040)	(0.020)	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)
Forward Linkages	-5.187***	-3.414***	-0.726**	-0.785**	-0.433**	-0.204**	-0.076**
*(Market Potential*Share of Intermediates)*	(2.013)	(1.231)	(0.290)	(0.377)	(0.184)	(0.081)	(0.031)
Backward Linkages	5.376**	3.131*	1.185***	1.284***	0.485**	0.251**	0.133***
*(Market Potential*Sales to Industry)*	(2.647)	(1.682)	(0.341)	(0.353)	(0.232)	(0.109)	(0.049)
Heckscher-Ohlin							
Endowment of Iron Ore	0.002	-0.000	0.001***	0.343**	0.104	0.217*	0.224**
*(Iron Ore Output*Share of Iron in GO)*	(0.002)	(0.001)	(0.000)	(0.149)	(0.135)	(0.117)	(0.108)
Human capital Endowment	1.464	2.940	4.006	0.136	0.697	0.499**	0.178*
*(Pop. in Sec. School*Share White-Collar)*	(6.712)	(4.583)	(3.336)	(0.477)	(0.442)	(0.231)	(0.094)
Endowment of Agriculture	9.740***	12.907***	9.192***	9.568***	4.763**	2.618	2.984*
*(Arable Land*Agricultural Inputs)*	(2.768)	(2.862)	(2.984)	(2.886)	(2.057)	(1.948)	(1.750)
Endowment of Wood	7.702***	6.944***	7.553***	11.694***	13.071***	10.222***	10.588***
*(Woodland*Forestry Inputs)*	(2.116)	(2.235)	(2.629)	(2.541)	(2.589)	(2.655)	(2.277)
Constant	-6.855***	-5.861***	-4.394***	-3.397***	-3.800***	-4.013***	-4.597***
	(1.832)	(1.316)	(1.319)	(1.293)	(0.933)	(1.000)	(0.965)
Shea's Partial R-squared (First Stage)							
Increasing Returns to Scale	0.979	0.984	0.974	0.998	0.983	0.996	0.991
Forward Linkages	0.986	0.990	0.977	0.998	0.985	0.996	0.992
Backward Linkages	0.987	0.989	0.975	0.998	0.984	0.996	0.992
Hansen J-statistic	2.871	2.143	4.726	5.565	5.548	10.042**	8.727**
p-value	(0.412)	(0.543)	(0.193)	(0.135)	(0.136)	(0.018)	(0.033)
Industry Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region Dummies?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	174	178	183	184	182	187	189
R-squared	0.566	0.579	0.579	0.578	0.584	0.631	0.644

Notes: The estimated equation is $ln(S_{ik}) = \alpha + \delta_i + \theta_k + \sum_j \beta^j x_i^j y_k^j + \varepsilon_{ik}$. The New Economic Geography interactions are instrumented with the sum of distances between regional capitals and predetermined levels of market potential. Robust standard errors in parentheses, where significance is denoted by *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Two-Step GMM with Predetermined and Geographical Instruments

Turning to the Heckscher-Ohlin interactions in Table 7 the interaction capturing the endowment of iron ore is significant in the interwar period and in the postwar period. Human capital is not significant except for 1950 and 1960. Except for arable land in 1950 both interactions capturing agricultural and forestry inputs are highly significant throughout the period, suggesting that industries that used inputs from these sectors were drawn to regions with abundant arable land and woodland.

But these results may be sensitive to different estimation techniques. Therefore, I proceed in the next section with a brief discussion of various robustness checks that I have performed. Since the coefficients are not comparable across interactions and across time, since they are measured in different units, I then proceed with standardizing the coefficients to evaluate the relative importance of the Heckscher-Ohlin and New Economic Geography determinants of industry location.

4.2.4 Robustness Checks

I have performed a number of robustness checks, presented and discussed in Appendix B and briefly reviewed here. First, I have estimated market potentials including foreign GDPs (see section 3.1) and reestimated the model using OLS with White (1980) errors. The results remain similar although not identical to the results obtained when using market potentials only based on domestic regional GDPs. Second, I have estimated the model by OLS using clustered errors to take into account the potential unobserved cluster-effect arising from the fact that the sample is constructed for regions. These results are similar to the results obtained using OLS and the method of White (1980). Third, I have estimated the model by OLS including the full set of size controls, regional characteristics, and industry characteristics. The interactions considering increasing returns to scale and intermediate consumption becomes considerably less prevalent although the overall results are similar to the ones obtained using the dummy-variable approach. Fourth, I have estimated the model by two-step GMM using only the sum of distances between regional capitals as instruments. The increasing returns to scale and human capital interactions loses their significance but the overall results are again similar to the ones obtained when including all available instruments.

4.3 Discussion: Markets or Endowments?

Since the estimated parameters does not lend themselves to comparison across interactions and years, as the variables are measured in different units and vary over time, I proceed by standardizing the coefficients in order to discuss the relative importance of factor endowments and market potential as determinants of industry location across Swedish regions. The coefficients are standardized such that $\beta_{jt} = b(x_{jt}) \left[\sigma^x_{jt}/\sigma^y_t\right]$ where b is the estimated coefficient for interaction x_{jt} and σ^x_{jt} and σ^y_t are the standard deviations of the interaction and dependent variable respectively. As the standardized coefficients are measured in the same units (standard deviations) they can readily be compared across variables and time. The standardized coefficients are presented for the seven interactions for all years in Table 8.

The effect of increasing returns to scale seems to have been the individually most important determinant in the first two decades of the 20th century implying that industries with increasing returns to scale were drawn toward regions with high market potential. This squares well with the the location of the industries with the highest degree of increasing

Variable	1900	1910	1920	1930	1940	1950	1960
Increasing Returns to Scale	*1.08*	*1.07*	0.11	0.10	*0.41*	*0.29*	*0.70*
Forward Linkages	*-0.93*	*-0.99*	*-0.82*	*-0.71*	*-0.64*	*-0.69*	*-0.58*
Backward Linkages	*0.62*	*0.59*	*0.85*	*0.74*	*0.45*	*0.53*	*0.63*
Iron Ore	0.00	0.00	*0.00*	*0.05*	0.02	*0.06*	*0.06*
Human Capital	0.04	0.13	0.22	0.03	0.20	*0.26*	*0.17*
Arable Land	*0.29*	*0.41*	*0.28*	*0.28*	*0.15*	0.09	*0.09*
Woodland	*0.41*	*0.41*	*0.48*	*0.66*	*0.81*	*0.67*	*0.70*
Total New Economic Geography	2.63	2.65	1.67	1.45	1.50	1.51	1.91
Total Heckscher-Ohlin	0.70	0.82	0.76	0.99	0.96	0.99	1.02

Sources and notes: The standardized coefficients are based on the two-step GMM estimates provided in Table 7 and the standard deviation of each interaction and the dependent variable for each benchmark year. See text for calculations. Significance at the 90 percent level is denoted by italics. The total effect of the New Economic Geography and Heckscher-Ohlin interactions are calculated as the sum of the absolute standardized coefficients of all significant interactions.

Table 8: Standardized β-Coefficients, 1900-1960

returns , i.e. the textiles and clothing industries (Table 3), that were mainly located in the metropolitan regions and the cities of Borås and Norrköping, proximate to the metropolitan markets of Stockholm and Gothenburg (Schön, 2010). These findings are also consistent with the more general pull of centrality observed for the modern manufacturing industries in the early 20th century (Schön, 1997). In the interwar period this effect seems to have become less prevalent at the same time as the importance of backward linkages rose in importance. This squares well with recent findings regarding industry location in the United Kingdom (Crafts and Mulatu, 2005) where the role of increasing returns diminished considerably from interwar period and onwards[22]. Thus, in the interwar period the pull of centrality seems to mainly have affected industries that sold a large share of output to other industries. From the outbreak of the Second World War the role of increasing returns increased again at the same time as the role of backward linkages decreased. Over the entire period industries where intermediate consumption was high located in regions with low market potential. This could be related to the fact that industries such as the food industry that relied on primary products as inputs in production (Table 3) were more sensitive to the congestion costs in the regions with high market potential. The negative effect seems to have decreased throughout the period, suggesting that the centrifugal forces became less prevalent.

Turning to the Heckscher-Ohlin determinants of industry location it is evident that most important factors in the early 1900s were the endowments of arable land and woodland. This is not surprising given the weight attached to the exports of agricultural goods and timber during early Swedish industrialization (Schön, 2010). Although the role of arable land diminished significantly throughout the period the distribution of woodland seems to have become more important in the interwar period. This relates to the expansion of the paper and pulp industries during this period where employment seems to have been drawn to the

[22]In Spain the effects of increasing returns persisted throughout the interwar period although with a slightly decreasing magnitude (Martinez-Galarraga, 2012).

regions most well endowed with wood. The results indicate that the distribution of iron did not affect the distribution of manufacturing employment as the standardized coefficients, although sporadically significant, are very small (Table 8). Regions producing iron ore, such as the counties of Norrbotten and Kopparberg, contained roughly 1-2 percent of employment in the metals and machinery industry. This suggests that these regions, although endowed with a central natural resource, failed to attract industrial activity that extended beyond basic extraction and refinement. Regional differences in human capital does not seem to have influenced industry location prior to the 1940s. This can be interpreted in two ways. First, as emphasized in the previous literature the early introduction of compulsory schooling promoted an even distribution and high level of human capital across Swedish regions (see, e.g., Sandberg, 1979; Ljungberg and Nilsson, 2009). Thus, marginal differences in human capital levels did not results in important advantages for some regions. Second, as emphasized by Goldin and Katz (1998) the role of skilled labor may have increased during this period due to increasingly capital-intensive production methods and electrification. But regional differences in human capital seems to have become important only in the latter part of the period (Table 8). This squares well with the general expansion and increased importance of administrative and managerial employment from the interwar period (Schön, 2010), but less so with an explanation based on capital- or technology-skill complementarities in the early 20th century. An example of this effect is provided by the chemical industry, the most human capital-intensive (Table 3), that was relatively dispersed in the early 1900s but became increasingly concentrated to the metropolitan regions of Stockholm and Malmöhus, well endowed with human capital and producer services, over this period.

Was Heckscher-Ohlin or New Economic Geography forces the most important determinant of industry location over this period? The bottom two rows in Table 8 presents the sum of a one standard deviation change in each significant New Economic Geography and Heckscher-Ohlin interaction respectively[23]. Clearly, market potential and the corresponding interactions seems to have constituted a more important determination of industry location than the interactions based on factor endowments. In 1900 the sum of the New Economic Geography and Heckscher-Ohlin interactions are 2.6 and 0.7 respectively, although this difference decreased over time.

5 Conclusions

In this paper I have shown, in terms of Gini coefficients and Theil indices, that the Swedish manufacturing industry became increasingly concentrated in the interwar period whereas the period from the outbreak of the Second World War was characterized by strong diffusion of employment. This bell-shaped evolution resulted in a regional distribution of manufacturing that was more equal in the 1960s than during any previous decade of the 20th century. I then proceeded by showing that the concentration patterns, in terms of the coefficient of variation, differed across industries. The overall trend, except for the chemical industry, was one of diffusion during the first two-thirds of the 20th century. To examine the degree of regional specialization I calculated Krugman specialization indices that showed that the industrial composition of Swedish regions became considerably more similar between 1900

[23]This sum is the sum of the absolute value of the significant interactions such that the coefficient of the interaction between market potential and intermediates in gross output enters with a positive sign.

and 1960.

In order to examine the determinants of industry location I then estimated a version of the Midelfart-Knarvik et al. (2000) model where the interaction between region and industry characteristics determine the location of industries. OLS estimations showed that the distribution of industries across Swedish regions, between 1900 and 1960, were jointly determined by factor endowments and market potential. More specifically, industries that intensively used inputs from the agricultural and forestry sectors located in regions abundant in agricultural land and woodland, although the importance of agricultural land diminished from the 1930s. The distribution of iron ore did not have a substantial effect in attracting industries during this period. From the 1940s and onwards, but not before, the regional endowment of human capital was important in attracting skill-intensive industries. Except for in the interwar period, there seems to have been a pull of centrality as industries with increasing returns to scale located in regions with high market potential. Industries where a large share of output was sold as inputs to other industries located in regions with high market potential, this effect being especially important in the interwar period. But high market potential seems to have deterred industries where intermediate consumption was high, although this effect declined over time. In addition, to identify the causal effects of market potential I used an instrumental variables approach (two-step GMM) where I used predetermined levels of market potential and the sum of distances between regional capitals as exogenous instruments of market potential, that confirmed the results obtained by OLS. Using standardized coefficients I then showed that New Economic Geography forces was the most important determinant of industry location across Swedish regions between 1900 and 1960.

References

Amiti, M. 1998. New Trade Theories and Industrial Location in the EU: A Survey of Evidence. *Oxford Review of Economic Policy 14,* 2, 45–53.

Balderston, T. 2010. The Economics of Abundance: Coal and Cotton in Lancashire and the World. *The Economic History Review 63,* 3, 569–590.

Baum, C., Schaffer, M., and Stillman, S. 2003. Instrumental Variables and GMM: Estimation and Testing. *Stata Journal 3,* 1, 1–31.

Berger, T., Enflo, K., and Henning, M. 2012. Industry Location, Urbanization, and Diffusion in Sweden, 1900-1960: Evidence from a New Database. *Forthcoming*.

Betran, C. 2005. Natural Resources, Electrification and Economic Growth from the End of the Nineteenth Century Until World War II. *Revista de Historia Económica (Second Series) 23,* 01, 47–81.

Betran, C. 2011. Regional Specialisation and Industry Location in the Long Run: Spain in the US Mirror (1856-2002). *Cliometrica 5,* 3, 259–290.

Bohlin, J. 2007. Structural Change in the Swedish Economy in the Late Nineteenth and early Twentieth Century - The Role of Import Substitution and Export Demand. *Göteborg Papers in Economic History* 8.

Brülhart, M. 1998. Economic Geography, Industry Location and Trade: The Evidence. *The World Economy 21,* 6, 775–801.

Brülhart, M. and Torstensson, J. 2001. Regional Integration, Scale Economies and Industry Location in the European Union. *CEPR Discussion Paper No. 1435*.

Christaller, W. 1933. *Central Places in Southern Germany.* Englewood Cliffs, NJ: Prentice-Hall.

Combes, P., Lafourcade, M., Thisse, J., and Toutain, J. 2011. The Rise and Fall of Spatial Inequalities in France: A Long-Run Perspective. *Explorations in Economic History 48,* 2, 243–271.

Crafts, N. 2005. 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.

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.

Deaton, A. 1994. *The Analysis of Household Surveys: Microeconometric Analysis for Development Policy.* Princeton University Press.

Edquist, H. 2010. Does Hedonic Price Indexing Change Our Interpretation of Economic History? Evidence from Swedish Electrification. *The Economic History Review 63,* 2, 500–523.

Enflo, K., Henning, M., and Schön, L. 2010. Swedish Regional GDP 1855-2000: Estimations and General Trends in the Swedish Regional System. Working paper, Universidad Carlos III, Departamento de Historia Económica e Instituciones.

Felice, E. 2011. Regional Value Added in Italy, 1891-2001, and the Foundation of a Long-Term Picture. *The Economic History Review 64,* 3, 929–950.

Flam, H. and Flanders, M. 1991. Heckscher-Ohlin Trade Theory. *Cambridge, Mass.: MIT Press 164*, 1–54.

Fridlizius, G. 1963. Sweden's Exports 1850-1960: A Study in Perspective. *Economy and History 6*, 3–100.

Geary, F. and Stark, T. 2002. Examining Ireland's Post-Famine Economic Growth Performance. *The Economic Journal 112,* 482, 919–935.

Gini, C. 1912. Variabilità e Mutabilità. *Reprinted in Memorie di Metodologica Statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi 1.*

Goldin, C. and Katz, L. 1998. The Origins of Technology-Skill Complementarity. *The Quarterly Journal of Economics 113,* 3, 693–732.

HARRIS, C. 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.

HEAD, K. AND MAYER, T. 2004. The Empirics of Agglomeration and Trade. *Handbook of Regional and Urban Economics 4*, 2609–2669.

HENNING, M., ENFLO, K., AND ANDERSSON, F. 2011. Trends and Cycles in Regional Economic Growth: How Spatial Differences Shaped the Swedish Growth Experience from 1860-2009. *Explorations in Economic History 48,* 4, 538–555.

HIPPE, R. AND BATEN, J. 2011. Regional Inequality in Human Capital Formation in Europe, 1790-1880. *Working Paper, Association Française de Cliométrie (AFC).*

JONSSON, P., SANDGREN, F., LINDBERG, E., AND SNÄLL, T. 2009. Towns and Rural Industrialisation in Sweden 1850-1890: A Spatial Statistical Approach. *Scandinavian Economic History Review 57,* 3, 229–251.

JÖRBERG, L. 1961. *Growth and Fluctuations of Swedish Industry 1869-1912: Studies in the Process of Industrialisation.* Vol. 3. Almqvist & Wiksell.

KIM, S. 1995. Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in US Regional Manufacturing Structure, 1860-1987. *The Quarterly Journal of Economics 110,* 4, 881.

KIM, S. 1999. Regions, Resources, and Economic Geography: Sources of US Regional Comparative Advantage, 1880-1987. *Regional Science and Urban Economics 29,* 1, 1–32.

KLEIN, A. AND CRAFTS, N. 2011. Making Sense of the Manufacturing Belt: Determinants of US Industrial Location, 1880-1920. *Journal of Economic Geography*.

KRANTZ, O. AND SCHÖN, L. 2007. *Swedish Historical National Accounts 1800-2000.* Almqvist & Wiksell.

KRUGMAN, P. 1991a. *Geography and Trade.* The MIT Press.

KRUGMAN, P. 1991b. Increasing Returns and Economic Geography. *Journal of Political Economy 99*, 483–499.

KRUGMAN, P. AND VENABLES, A. 1995. Globalization and the Inequality of Nations. *The Quarterly Journal of Economics CX,* 4 (November), 857–880.

LEAMER, E. 1995. *The Heckscher-Ohlin Model in Theory and Practice.* International Finance Section, Department of Economics, Princeton University.

LEE, C. 1971. *Regional Economic Growth in the United Kingdom Since the 1880s.* McGraw-Hill London.

LJUNGBERG, J. AND NILSSON, A. 2009. Human Capital and Economic Growth: Sweden 1870-2000. *Cliometrica 3,* 1, 71–95.

LÖSCH, A. 1940. *The Economics of Location.* New Haven: Yale University Press, 1954.

LUNDH, C. 2002. *Spelets Regler: Institutioner och Lönebildning På Den Svenska Arbetsmarknaden 1850-2000*. SNS förl.

LUNDH, C., SCHON, L., AND SVENSSON, L. 2005. Regional Wages in Industry and Labour Market Integration in Sweden, 1861-1913. *Scandinavian Economic History Review* 3, 71–84.

MADDISON, A. 2003. *The World Economy: Historical Statistics*. Organization for Economic Development.

MARTINEZ-GALARRAGA, J. 2012. The Determinants of Industrial Location in Spain, 1856-1929. *Explorations in Economic History 49*, 255–275.

MIDELFART-KNARVIK, K., OVERMAN, H., REDDING, S., AND VENABLES, A. 2004. *The location of European industry*. Vol. European integration and the functioning of product markets. Edward Elgar Pub.

MIDELFART-KNARVIK, K., OVERMAN, H., AND VENABLES, A. 2000. Comparative Advantage and Economic Geography: Estimating the Determinants of Industrial Location in the EU. Tech. rep., Centre for Economic Policy Research, London School of Economics and Political Science.

NEARY, J. 2001. Of Hype and Hyperbolas: Introducing the New Economic Geography. *Journal of economic Literature 39,* 2, 536–561.

NEFFKE, F., HENNING, M., AND BOSCHMA, R. 2011. How Do Regions Diversify Over Time? Industry Relatedness and the Development of New Growth Paths in Regions. *Economic Geography 87,* 3, 237–265.

NEFFKE, F., HENNING, M., BOSCHMA, R., LUNDQUIST, K., AND OLANDER, L. 2011. The Dynamics of Agglomeration Externalities Along the Life Cycle of Industries. *Regional Studies 45,* 1, 49–65.

NILSSON, J. 2006. Regional Planering i Sverige. *Planering Med Nya Förutsättningar: Ny lagstiftning, Nya Värderingar. Linköping University Interdisciplinary Studies No. 6.*, 91–106.

OLSSON, F. 2007. Järnhanteringens Dynamik: Produktion, Lokalisering och Agglomerationer i Bergslagen. Ph.D. thesis, Umeå University, Faculty of Social Sciences, Economic History.

ÖRTENBLAD, T. 1914. *Sweden Historial and Statistical Handbook, Second Part: Industries*. Stockholm: Government Printing Office. P. A. Norstedt och Söner, Chapter Forestry.

OTTAVIANO, G. AND THISSE, J. 2004. Agglomeration and Economic Geography. *Handbook of Regional and Urban Economics 4*, 2563–2608.

POLLARD, S. 1981. *Peaceful Conquest: The Industrialization of Europe 1760-1970*. Oxford: University Press Oxford.

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.

Redding, S. and Venables, A. 2004. Economic Geography and International Inequality. *Journal of International Economics 62,* 1, 53–82.

Rosés, J. 2003. Why Isn't the Whole of Spain Industrialized? New Economic Geography and Early Industrialization, 1797-1910. *The Journal of Economic History 63,* 4, 995–1022.

Ross, E. 1896. The Location of Industries. *The Quarterly Journal of Economics 10,* 3, 247–268.

Sandberg, L. 1979. The Case of the Impoverished Sophisticate: Human Capital and Swedish Economic Growth Before World War I. *Journal of Economic History 39,* 1, 225–241.

Schön, L. 1997. Internal and External Factors in Swedish Industrialization. *Scandinavian Economic History Review 45,* 3, 209–223.

Schön, L. 2000. Electricity, Technological Change and Productivity in Swedish Industry, 1890-1990. *European Review of Economic History 4,* 2, 175–194.

Schön, L. 2010. *Sweden's Road to Modernity: An Economic History.* SNS.

Schulze, M. 2007. Regional Income Dispersion and Market Potential in the Late Nineteenth Century Hapsburg Empire. *Working Paper, LSE*.

Shea, J. 1997. Instrument Relevance in Multivariate Linear Models: A Simple Measure. *Review of Economics and Statistics 79,* 2, 348–352.

Söderberg, J. 1984. A Long-Term Perspective on Regional Economic Development in Sweden, Ca. 1550-1914. *Scandinavian Economic History Review 32,* 1, 1–16.

Söderberg, J. and Lundgren, N. 1982. *Ekonomisk och Geografisk Koncentration 1850-1980.* LiberFörlag.

Stobart, J. 2000. In Search of Causality: A Regional Approach to Urban Growth in Eighteenth-Century England. *Geografiska Annaler. Series B. Human Geography*, 149–163.

Stock, J. H. and Yogo, M. 2005. *Testing For Weak Instruments in Linear IV Regression.* Identification and Inference for Econometrics Models. Essays in Honor of Thomas Rothenberg. Cambridge: Cambridge University Press.

Sundbärg, G. and Åmark, K. 1914. *Sweden Historial and Statistical Handbook, Second Part: Industries.* Stockholm: Government Printing Office. P. A. Norstedt och Söner, Chapter Manufacturing Industries.

Theil, H. 1967. *Economics and Information Theory.* Amsterdam: North-Holland.

Tirado, D., 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.

Törnlund, E. and Östlund, L. 2006. Mobility Without Wheels: The Economy and Ecology of Timber Floating in Sweden, 1850-1980. *The Journal of Transport History 27,* 1, 48–70.

von Thünen, J. 1826. *The Isolated State (English Translation, P.G. Hall, 1966).* New York: Pergamon Press.

White, H. 1980. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

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.

A Data Appendix

In this appendix I provide some additional details regarding the construction of the dataset (see section three).

Regional Manufacturing Employment
Data on regional manufacturing employment was obtained from the recently constructed *Swedish Manufacturing Industry Database* (SMID), which in turn is based on the official industrial statistics (*BiSOS-D* and *SOS: Industri*) published by Statistics Sweden (see Berger et al., 2012). I have excluded the mining industries as these were not included in the industrial statistics prior to 1913. Similarly, the power industry is excluded since it is not included in the industrial statistics from the 1960s and onwards.

Human Capital
Data on enrollment rates in secondary schooling (*BiSOS-P*), 1870-1940, was kindly provided by Kerstin Enflo. For the benchmark years 1950 and 1960 I have used the enrollment rates for 1940. Regarding the skill-intensity of industries, the industrial statistics (*BiSOS-D* and *SOS: Industri*) did not report white-collar employment prior to the 1920s. Therefore I have used the share of white-collar employment for each industry in the year 1920 for 1900 and 1910.

Regional Production of Iron Ore
As data on iron ore production was not readily available for the year 1960 I have used the output for 1950. Given the relative stability of regional shares of production of iron ore this should have neglible effects on the results.

Market Potential
The geodesic distances between domestic regional capitals, and foreign nodes, was obtained from an electronic atlas (http://www.distancefromto.net/). Since Maddison (2003) does not provide GDP data for Poland in 1920 and 1940 and the Czech Republic in 1940 I have interpolated these years based on the GDP per capita in adjacent years and the population figures provided by Maddison. When calculating market potential including foreign markets the regional GDPs provided by Enflo et al. (2010) was converted into 1990 $GK using the GDP figures for Sweden provided in Maddison (2003). Following the suggestion of Crafts (2005) I have used *nominal* regional GDPs when constructing market potentials solely based on regional markets as these were the prices that contemporary agents responded to.

Data Sources
Statistics Sweden. *Bidrag till Sveriges Officiella Statistik: Fabriker och Manufakturer* 1900/10
Statistics Sweden. *Bidrag till Sveriges Off. Statistik: Jordbruk och Boskapsskötsel, 1900/10*
Statistics Sweden. *Bidrag till Sveriges Officiella Statistik: Skogsväsendet, 1900/10*
Statistics Sweden. *Bidrag till Sveriges Officiella Statistik: Undervisningsväsendet, 1900/10*
Statistics Sweden. *Sveriges Officiella Statistik: Industri*, 1920/30/40/50/60
Statistics Sweden. *Statistical Yearbook of Sweden,* 1920/30/40/50/60/90

B Robustness Checks

In this section I present various robustness checks of the main results in section four.

B.1 OLS With Market Potential Including Foreign GDPs

Including foreign GDPs in the measure of regional market potential is not a trivial task (see Schulze, 2007). Foreign markets are often weighted by the distance-equivalent to tariffs, obtained from gravity estimates, and actual tariff rates (Crafts, 2005; Martinez-Galarraga, 2012), but this data is not readily available for Sweden. Therefore I have calculated market potentials of Swedish regions where foreign GDPs are weighted by the geodesic distances between regional capitals and foreign nodes (see section three). In Table 9 I present the estimations of equation (8) by OLS, where I include foreign GDPs in the market potential interactions. Although the results are similar to the ones presented in Table 7 these estimates seem to suggest that the Swedish manufacturing industry was more responsive to domestic rather than foreign markets.

B.2 OLS With Size Controls, Region-, and Industry Characteristics

I have estimated equation (2) by OLS using the method of White (1980), including size controls, region, and industry characteristics. The interactions are presented in Table 10. The interactions capturing increasing returns to scale and forward linkages become less significant suggesting that the pull of centrality mainly worked through backward linkages. The other interactions remain very similar to the ones presented in Table 7.

B.3 OLS With Cluster-Robust Errors

I have estimated equation (2) by OLS using cluster-robust errors, thus taking into account the potentially unobserved cluster-effect arising from the fact that the sample is constructed for regions. The results are presented in Table 11. Although the New Economic Geography interactions become less prevalent the overall results are very similar to the ones presented in Table 7.

B.4 Two-Step GMM With Sum of Distances as Instruments

I have estimated equation (8) by two-step GMM solely using the geographical distances between regional capitals as an instrument for market potential. The results are presented in Table 12. Except for the fact that the increasing returns to scale interaction loses its significance the results remain very similar to the ones presented in Table 7.

	(1) 1900	(2) 1910	(3) 1920	(4) 1930	(5) 1940	(6) 1950	(7) 1960
VARIABLES	*ln(Share of manufacturing industry i in region j)*						
IRS	0.112**	0.074**	0.005	0.004	0.019	0.007	0.003
	(0.056)	(0.032)	(0.020)	(0.010)	(0.012)	(0.011)	(0.006)
Intermediates	-5.653	-5.401*	-7.468***	-4.588**	-2.435*	-2.311*	-1.310*
	(4.300)	(2.995)	(2.853)	(1.913)	(1.313)	(1.319)	(0.690)
Sales to Ind.	7.414	3.439	9.309**	4.786**	1.921	3.417*	2.676**
	(5.572)	(3.764)	(3.664)	(2.327)	(1.872)	(1.899)	(1.086)
Iron Ore	1.118	-0.477	1.207***	0.246	0.073	0.266*	0.282*
	(2.359)	(0.702)	(0.358)	(0.188)	(0.161)	(0.139)	(0.147)
Human Capital	3.877	4.225	5.487	0.398	1.099*	0.811***	0.298***
	(8.025)	(5.201)	(3.761)	(0.578)	(0.585)	(0.299)	(0.113)
Arable Land	9.962**	12.122***	11.343***	11.636***	6.189*	4.814	5.575*
	(4.883)	(4.108)	(4.159)	(3.798)	(3.170)	(3.498)	(2.976)
Woodland	4.307	5.511	3.948	10.025***	10.742***	6.717	5.993*
	(3.181)	(3.423)	(4.070)	(3.793)	(3.754)	(4.310)	(3.529)
Constant	-7.387***	-5.167***	-3.679***	-2.331*	-3.589***	-3.975***	-4.289***
	(1.947)	(1.363)	(1.295)	(1.236)	(1.069)	(1.213)	(1.210)
Observations	174	178	183	184	182	187	189
R-squared	0.563	0.565	0.580	0.583	0.590	0.628	0.639

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

Table 9: OLS with Market Potentials Including Foreign GDPs

	(1) 1900	(2) 1910	(3) 1920	(4) 1930	(5) 1940	(6) 1950	(7) 1960
VARIABLES	*ln(Share of manufacturing industry i in region j)*						
IRS	0.058	0.022	0.001	-0.001	0.002	0.001	0.001
	(0.041)	(0.020)	(0.003)	(0.002)	(0.002)	(0.001)	(0.000)
Intermediates	-2.552*	-1.443	-0.243	-0.397	-0.182	-0.083	-0.027
	(1.452)	(1.034)	(0.246)	(0.348)	(0.166)	(0.074)	(0.034)
Sales to Ind.	4.950*	3.047*	0.932***	1.002***	0.322	0.236**	0.129**
	(2.646)	(1.819)	(0.354)	(0.373)	(0.237)	(0.110)	(0.050)
Iron Ore	0.841	-0.270	1.064***	0.263	0.040	0.259*	0.274*
	(2.130)	(0.693)	(0.400)	(0.193)	(0.174)	(0.137)	(0.141)
Human Capital	2.731	2.241	3.659	0.383	0.894*	0.606**	0.226**
	(7.434)	(5.001)	(3.356)	(0.572)	(0.483)	(0.259)	(0.102)
Arable Land	7.054**	10.234***	6.596**	7.179***	2.812	1.589	2.053
	(3.264)	(2.994)	(2.721)	(2.752)	(1.973)	(1.841)	(1.678)
Woodland	7.926***	8.694***	8.984***	13.591***	14.358***	9.930***	9.573***
	(2.373)	(2.473)	(2.975)	(2.774)	(2.887)	(2.956)	(2.578)
Constant	2.793	1.563	3.991	2.563	1.941	2.922**	2.567**
	(2.357)	(2.439)	(2.479)	(2.047)	(1.629)	(1.376)	(1.266)
Observations	174	178	183	184	182	187	189
R-squared	0.533	0.542	0.553	0.546	0.557	0.608	0.621

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

Table 10: OLS Including Size Controls, Region, and Industry Characteristics

	(1) 1900	(2) 1910	(3) 1920	(4) 1930	(5) 1940	(6) 1950	(7) 1960
VARIABLES	*ln(Share of manufacturing industry i in region j)*						
IRS	0.086	0.045*	0.001	-0.000	0.002	0.001	0.000*
	(0.051)	(0.025)	(0.002)	(0.002)	(0.002)	(0.001)	(0.000)
Intermediates	-5.006**	-3.535**	-0.601	-0.642	-0.345	-0.167	-0.059
	(2.047)	(1.271)	(0.379)	(0.530)	(0.252)	(0.133)	(0.045)
Sales to Ind.	5.639**	3.164	1.114**	1.166***	0.445	0.293	0.149
	(2.318)	(2.229)	(0.404)	(0.415)	(0.299)	(0.180)	(0.091)
Iron Ore	1.079	-0.223	1.193***	0.325**	0.105	0.283***	0.290***
	(1.928)	(0.811)	(0.321)	(0.133)	(0.125)	(0.091)	(0.098)
Human Capital	3.512	2.801	4.158	0.350	0.935	0.620**	0.220***
	(8.072)	(6.580)	(4.115)	(0.701)	(0.630)	(0.222)	(0.058)
Arable Land	8.485**	12.157***	8.052**	8.041**	3.536	2.355	2.703
	(3.232)	(3.566)	(3.848)	(3.760)	(2.691)	(2.768)	(2.319)
Woodland	7.444**	7.884**	8.295**	13.142***	13.935***	9.595**	9.331***
	(2.764)	(3.098)	(3.741)	(3.349)	(3.740)	(4.067)	(3.248)
Constant	-7.557***	-5.787***	-5.097***	-3.452**	-3.724***	-4.487***	-4.918***
	(2.328)	(1.128)	(1.398)	(1.355)	(1.026)	(1.305)	(1.157)
Observations	174	178	183	184	182	187	189
R-squared	0.571	0.582	0.583	0.582	0.590	0.640	0.653

Robust standard errors in parentheses

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

Table 11: OLS With Errors Clustered at the Regional Level

	(1) 1900	(2) 1910	(3) 1920	(4) 1930	(5) 1940	(6) 1950	(7) 1960
VARIABLES	*ln(Share of manufacturing industry i in region j)*						
IRS	0.077	0.031	0.003	-0.002	0.001	0.001	0.000
	(0.055)	(0.037)	(0.004)	(0.003)	(0.003)	(0.001)	(0.001)
Intermediates	-7.526***	-3.904**	-0.993***	-1.354***	-0.741***	-0.372***	-0.121***
	(2.337)	(1.724)	(0.332)	(0.423)	(0.213)	(0.097)	(0.036)
Sales to Ind.	7.860**	2.911	0.971**	1.604***	0.885***	0.554***	0.281***
	(3.328)	(2.217)	(0.402)	(0.422)	(0.311)	(0.148)	(0.070)
Iron Ore	1.260	-0.260	1.166***	0.448***	0.249	0.400***	0.415***
	(1.616)	(0.655)	(0.290)	(0.157)	(0.153)	(0.117)	(0.138)
Human Capital	3.069	4.084	5.560	0.270	0.843	0.442	0.159
	(7.095)	(4.703)	(3.702)	(0.566)	(0.563)	(0.310)	(0.121)
Arable Land	9.895***	12.362***	9.155***	10.109***	5.488**	5.038**	4.509**
	(2.863)	(3.165)	(3.251)	(3.356)	(2.419)	(2.462)	(2.096)
Woodland	7.198***	7.955***	7.707***	12.314***	13.059***	8.319***	8.293***
	(2.118)	(2.409)	(2.728)	(2.549)	(2.558)	(2.523)	(2.308)
Constant	-6.404***	-4.311**	-3.823***	-1.590	-2.955***	-4.269***	-5.218***
	(2.374)	(1.703)	(1.427)	(1.360)	(1.052)	(1.461)	(1.566)
Observations	174	178	183	184	182	187	189
R-squared	0.565	0.578	0.576	0.570	0.579	0.621	0.636

Robust standard errors in parentheses

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

Table 12: Two-Step GMM with Sum of Distances as Instrument