The Industrial Geography of Italy: Provinces, Regions and Border Effects (1871–1911)

Anna Missiaia*

Lund University

Abstract

Italy has been characterized, throughout its unified history, by large regional differentials in the level of industrialization. This paper looks at the distribution of industrial employment in the period 1871–1911. Using provincial industrial employment from Ciccarelli and Missiaia (2013) we address the question whether the change in the industrial employment at provincial level depends on the change in the industrial employment in the neighbouring provinces. The methodology proposed by Overman and Puga (2002) use neighbouring effects, based on neighbours employment, as explanatory variables for change in industrial employment. The same exercise is repeated using pre-unitary borders to define neighbours, in order to assess the persistence of pre-1861 institutions on the industrial geography of unified Italy. The main result of the paper is that regional borders did matter in shaping the industrial geography of Italy. We find that the change in provincial industrial employment is positively affected by the change of the neighbouring provinces belonging to the same region but negatively from the change of the neighbouring provinces belonging to another region. When the pre-unitary borders are used, the findings are basically confirmed.

JEL Classification Numbers: N63, N93, R3, R12.

Keywords: Economic Geography; Economic History of Italy; Market Potential; Border Effect; Industrial Location.

---

*Lund University, Economic History Department, Room 2110, Alfa 1, Scheelevägen 15 B, Lund, Skåne 22363, Sweden, e-mail: *anna.missiaia@ekh.lu.se*. EHS Annual Conference - Wolverhampton, 27-29 March 2015 - Preliminary version. Please do not cite without permission. The author is grateful to Brian A'Hearn, Marco Bertoni, Carlo Ciccarelli, Giovanni Federico, Steve Gibbons and Max Schulze for advice and suggestions.

1 Introduction

Italy's political Unification dates back to 1861. These years correspond to the first industrialization of the country. At the time, Italy was predominantly agricultural, with some early manufacturing in the Northwest, especially in the textile sector (Cafagna, 1989). In this period, all the Italian regions experience industrial growth at least to some extent. By 1911, all the modern sectors had been at some level established (Zamagni, 1990). At the time of Unification, the regions of Italy were not at all homogeneous in terms of their economic structure and performance. Throughout the period of its first industrialization, the country experienced the rise of large regional differentials that are still to be observed in contemporary Italy. The process of divergence across regions in terms of GDP per capita and level of industrialization accelerated after the political Unification, with the South falling behind. Felice (2009) and Brunetti et al. (2011) provide extensive evidence on GDP per capita while the polarization in terms of industrial value added has been considered both at regional and provincial level by Fenoaltea (2003b) and Ciccarelli and Fenoaltea (2013). This paper contains an analysis of the patterns of industrialization across Italian regions and provinces in the period 1871–1911. The goal is first to describe the location patterns using some synthetic indices and then to assess how the evolution of the employment in the industrial sectors relates to the presence of regional borders.

Following a well established practice in both Economic Geography and Economic History, we use the Krugman index, the G index and the Ellison and Glaeser (E-G) index to assess the specialization of the regions and the concentration of the industrial sectors. Our results will be compared to other countries and discussed as applying to the Italian case.

Spatial autocorrelation, which is the degree of spatial interdependence of the observations, will also be analysed through a calculation of the Moran's I for each industrial sector. This measure introduces in this analysis a spatial dimension, which is not considered when measuring the specialization and concentration with the standard Krugman index, G index and E-G index.

The second part of the paper is devoted to a study of the determinants of change in industrial employment over the census years. The paper addresses the question whether (and to what extent) the change over time in industrial employment at the provincial level depended on the change in the neighbouring provinces. Neighbouring provinces, meaning provinces that share a border with a given province, will be divided into two groups: "same region" and "other region" neighbours. The working assumption is that, if industrial employment presented regional border effects, the role of these two groups of neighbours will be different. The methodology used in this paper is the one proposed by Overman and Puga (2002) to test the existence of transnational clusters of unemployment across the EU. To adapt this model to the Italian case we consider the Italian provinces instead of European regions and the Italian regions instead of European countries.[1] The paper tests for both border effects corresponding to regional borders in the period 1871–1911 and for pre-unitary borders which might have had a long term impact perceived even after Unification. This analysis will exploit a newly published provincial level dataset from Ciccarelli and Missiaia (2013). The present study provides for the first time information on employment at the provincial level for all 15 industrial sectors, separating males and females; it also allows us to look at sub-regional units.

The main results of the descriptive analysis of this paper are that regions in this period presented a fairly high level of specialization, mirrored by a concentration of the industrial sectors. Moreover, the spatial autocorrelation among regions was not particularly high, suggesting that regions were fairly independent from each other in terms of employment patterns. These findings are coherent with the results of the regression analysis, where we find that regional borders did matter. In fact, we find that, for a given region, the change in employment of the two types of neighbour had different signs, leading to different effects on employment.

The paper is organized as follows. Section 2 provides an historical account of the Italian local administrations; Section 3 illustrates the empirical framework for both the descriptive indices and the regression model; Section 4 presents the data set used in this work and discusses it in detail; Section 5 provides the empirical results and Section 6 concludes.

2 Pre-unitary states and Italian local administrations: an historical overview

The first attempt to organize the Italian territory into regions dates back to 7 A.D.. Emperor Augustus divided Italy into eleven regions, most named after the ancient populations that had occupied them (Galinsky, 2005, p. 80). These regions were not administrative units but served only as a way to organize population censuses. Figure 1 shows the eleven regions. Since the Barbaric invasions of the 6^{th} century, the different parts

[1] The methodology requires us to use two separate levels of geographical aggregation, one for which the employment is measured and a larger one that imposes borders across the smaller units. Therefore, as a work on the European Union would measure the change in the regions and test the national border effects, our work on Italy measures change in employment in the provinces and tests for regional border effects.

Figure 1: Italian Regions under Augustus, 7 A.D.

of the Italian peninsula did not belong to a single political entity until the unification of 1861. With the political fragmentation of the Middle Ages, the ancient names ceased to be used until approximately the 15^{th} century, when geographers started reviving these terms.[2] As we note, several of these regions, such as Venetia, Aemilia (Emilia), Liguria, Umbria, Etruria (Tuscany), Latium, Sardinia, Apulia, Calabria and Sicily preserve the names and often the borders of modern Italian regions.

After Unification, the pre-unitary states have been the basis for the formation of the Italian regions, in the north and centre of the country in particular. Figure 2 shows the map of Italy before 1861. Figure 3 shows the regions after the annexation of Rome in 1871. The model of administration adopted by the newly established country was, not surprisingly, taken largely from that of Piedmont. However, even though the transition of model was mostly based on the structure of Piedmont, the other pre-unitary states also presented a structure with intermediate elements. For instance, the Kingdom of the Two Sicilies was divided into 22 provinces and 76 districts (Spagnoletti, 1997, p. 162); the Kingdom of Lombardy Venetia was organized in 17 provinces and 218 districts (Meriggi, 1987, p. 34). After the Unification, most of these provinces kept the same name and shape.

At the time of Unification, the two options of centralization and decentralization divided intellectuals and politicians. The former originated from the French model and was basically the one that Piedmont had adopted. It was supported by Cavour, Rattazzi,

[2] Almagià (1935).

Figure 2: Italian states on the eve of Unification, 1861.

Mazzini, Garibaldi and others. The latter implied a federalist approach and was supported by Gioberti and especially Cattaneo. The centralized view, which implied a milder transition and fewer risks for political stability, prevailed.[3] The territory of Unified Italy was divided in 15 regions and 69 provinces. Smaller geographical units, "circondari" and "mandamenti", also existed with limited powers.[4] The smaller unit was the "comune", which represented the "natural delimitation [of the territory]" and had deep historical roots.[5]

Local units in this period had far less autonomy than those created after the Second World War. In the period we are looking at, the province was the main intermediate body between the "comune" and the central state. Regions were mere collections of provinces without powers and without a structure. In spite of some attempts by Crispi in the 1890s to create larger and more independent administrative units, regions would remain so until the new provisions of the Constitution of 1948.[6] However, in this paper we will claim that regions, in spite of the lack of formal powers, still represented meaningful economic units because of their historical connections to pre-unitary states and regions. The next section moves on to the empirical strategy pursued in this paper.

[3] Pavone (1964, p. 195).

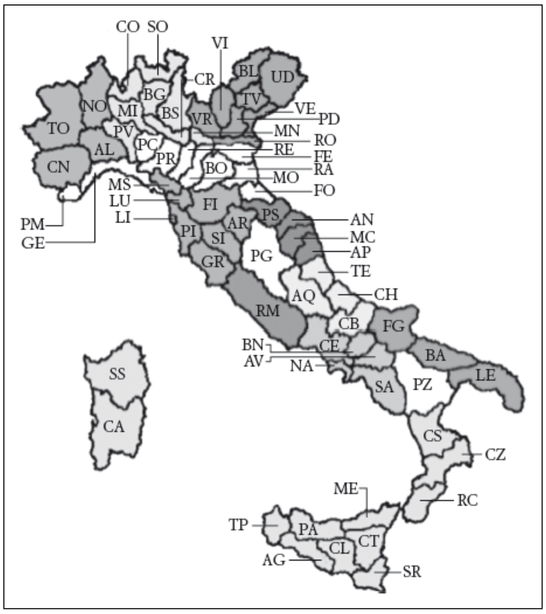
[4] See Antonelli and Palombelli (1995) for a survey of the legislative history of local administrations since 1861.

[5] Antonelli and Palombelli (1995, p. 71).

[6] Bonini (1997).

Figure 3: Italian regions, 1870–1918.

Figure 4: Italian provinces, 1871–1911.



The different shades of grey are used in the map only to indicate regional borders.

3 Spatial distribution of employment: empirical framework

In this section will describe our methodological approach to studying the patterns of spatial distribution of employment. In section 3.1 we propose some measures of concentration of the industrial sectors and specialization of the regions and in section 3.2 we illustrate the calculation of the Moran's I, a measure that takes into account the proximity of the regions and measures the spatial autocorrelation among data points.

The recent paper by Ciccarelli and Proietti (2011) looks at the specialization and the spatial patterns of industrial value added in the Italian provinces in the same period. These authors use a multivariate graphical technique named dynamic specialisation biplot to evaluate the degree of specialization of provinces from Unification until the First World War. The methodology used by Ciccarelli and Proietti (2011) is quite new to economic historians and represents an alternative methodology to the standard one used in this paper to measure specialization. The main difference, however, is in the geographical focus of the two studies: here specialization is measured for regions and it is simply a descriptive exercise to introduce a regression analysis. For Ciccarelli and Proietti (2011) the description of the specialization patterns of provinces is the main goal. The result is that provinces were not particularly specialized at the beginning of the period but they increased their degree of specialization as their industrialization process went on. Moreover, the Industrial Triangle presented a different pattern compared to the rest of the country, with a higher degree of specialization in modern sectors.

In section 3.3 we turn our attention to the the determinants of the distribution of industrial employment. In particular, we adapt a model proposed by Overman and Puga (2002) to describe the changes in the unemployment rate across the EU and use this in relation to the changes in the location of industrial employment in Italy.

3.1 Measuring geographic concentration and regional specialization: the Krugman index, the G index and the E-G index

The first step in the study of the location patterns of Italian industries is to measure the phenomenon of specialization of regions and concentration of industrial sectors within the regions. The location of industries in a given area, divided into subunits such as regions, can be studied looking at both the concentration of industries in the sub-units and the specialization of each subunit in certain industries. These phenomena are different but closely related. It is easy to predict that when industrial concentration is high, regions will be specialized. However, the two measures are not necessarily equal when industries and regions differ in size (Wolf, 2007, p. 25).

The concentration of a given industry k is measured as the ratio of the employment of industry k in region i on the total employment of k, following Wolf (2007):

$$L_{k,i}(t) = \frac{x_{k,i}(t)}{\sum_i x_{k,i}(t)} \,. \tag{1}$$

This measure is not an index and therefore it is not bounded by any two values.

Another index used by Wolf (2007) to overcome this problem is the Krugman index of concentration. It is calculated as follows:

$$K_k(t) = \sum_i |L_{k,i}(t) - \overset{*}{L}_{k,i}(t)| \tag{2}$$

where $\overset{*}{L}_{k,i}(t)$ is equivalent to $L_{k,i}(t)$ except that it excludes industry k. The Krugman index is bounded between 0 and 2. These are very basic measures of concentration that take into account only employment figures by industrial sector by region. Further developments in Economic Geography lead to more complex and refined measures. In the literature, three standard requirements are required for an index to be suitable for measuring the concentration of industries: the measure must be comparable across industries; it must control for the tendency of manufacturing to agglomerate and it must control for the degree of industrial concentration (Duranton and Overman, 2005, p. 1078). The first requirement is met by both the G index and the Krugman index, but the second and third are not. We can therefore introduce another index that controls for the size of plants and the level of geographic aggregation.

$L_{k,i}(t)$ is used to compute another index, the G index, which is the sum of the absolute difference between $L_{k,i}(t)$ and the share of the area of i in the total area:

$$G_k(t) = \sum_i |L_{k,i}(t) - \text{area}_i| \,. \tag{3}$$

The third option for measuring concentration comes from Ellison and Glaeser (1997). They propose the following index:

$$\begin{aligned} \gamma_{E,G}(t) &= \frac{G(t) - (1 - \sum_i x_i(t)^2)H(t)}{1 - \sum_i (x_i(t)^2)(1 - H(t))} \\ &= \frac{\sum_k (s_k(t) - x_i(t)^2) - (1 - \sum_i x_i(t)^2)(\sum_j z_j^2)}{(1 - \sum_i x_i(t)^2)(1 - (\sum_j z_j^2)} \end{aligned} \tag{4}$$

where $s_i(t)$ is the share of industry k employment in area i; $x_i(t)$ is the share of i in the total manufacturing employment; and z_j is the squared plant employment share indexed by

j. Ellison and Glaeser (1997, p. 902) classify their index according to these values: a sector is not very concentrated if the E-G index is smaller than 0.02; it is relatively concentrated if the E-G index is between 0.02 and 0.05 and it is highly concentrated over 0.05. The advantage of using an E-G index is that the size of plants is taken into account as well as the size of the regions throughout the industrial employment. The index calculated for each industry is then compared to a benchmark random distribution of industries. This allows the index not to take value zero if employment is uniformly spread across space but only if the distribution is comparable to a random one (De Dominicis et al., 2007, p. 4). This index requires in theory to know the share of employment of each single plant. These data is not available for any year of our period. The only information available on plants comes from the 1911 industrial census. For each industrial sector and region, the number of plants and workers is recorded. From this, we can work out the mean plant size by region by industry and the corresponding share. This is used in place of the plant employment share. The calculation of the E-G index is possible here for 1911 alone because of data limitations.

Regarding the specialization of regions, the simplest measure is the ratio of the employment of industry k in region i to the total employment of region i:

$$s_{k,i}(t) = \frac{x_{k,i}(t)}{\sum_k x_{k,i}(t)} \,. \tag{5}$$

The Krugman index of specialization can be calculated as follows:

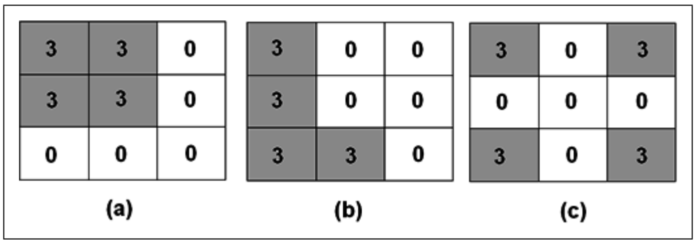
$$K_i(t) = \sum_k |S_{k,i}(t) - \overset{*}{S}_{k,i}(t)| \tag{6}$$

where $s_{k,i}(t)$ is the ratio of employment in industry k in region i over the total employment of all regions except i.

3.2 Moran's I and spatial autocorrelation

The last tool taken from the Economic Geography literature and employed in this paper to describe the localization of Italian industries is the Moran's I. This tool is used to detect whether adjacent regions tend to have closer values in the variable of interest. What differentiates the Moran's I from the previous indices is that it considers each region not as an isolated entity but in relation to the others. This is done through the information provided by a proximity matrix. The Moran's I measures the degree of spatial autocorrelation of the phenomenon studied. Spatial autocorrelation predicts that adjacent observations of the same variable will be more closely correlated than those further away. The notion is similar to standard autocorrelation in econometrics, but it develops across

Figure 5: Three cases of agglomeration.



space instead of time. Previous indices are "a-spatial" in the sense that every spatial unit is treated as isolated from the others. To better explain this point, we propose the example of De Dominicis et al. (2007, p. 16). We consider three possible location scenarios of twelve plants located across nine sub-regions, as shown in Figure 5.

The three scenarios would show the same level of spatial agglomeration when using a-spatial tools such as the Krugman index of concentration, the G index or the E-G index. The Moran's I unlike these can detect that Figure 5a has a higher agglomeration than 5b that has an higher agglomeration than that shown in Figure 5c. Spatial autocorrelation introduces the spatial dimension across regions, considering every region in its position relative to the others through a spatial weight matrix. The elements of the matrix take value one if the two regions are adjacent and zero otherwise. There are two types of Moran's I: the Global Moran's I and the Local Moran's I. The Global Moran's I yields to one index that summarizes the whole study area (in this case, Italy), assuming every region to be internally homogeneous. The Local Moran's I is in contrast calculated for every spatial unit in order to detect clustering within each unit. According to Anselin (1995, p.94), Local Indicators of Spatial Association, such as the Local Moran's I, are proportional to the global indicator of spatial association (the Global Moran's I). In our case, we calculate the global Moran's I. This is because we have no information on the distribution of firms within each region.

The Moran's I is defined in Fotheringham et al. (2000, p. 201) as:

$$I = \frac{\left(\frac{N}{\sum_k \sum_j w_k, j}\right)\left(\sum_k \sum_j w_k, j(x_k - x_a ve)(x_j - x_a)\right)}{\sum_k (x_k - x_a)^2} \tag{7}$$

where N is the number of regions, $w_{k,j}$ is a discrete variable which takes value 1 if regions i and j are adjacent and 0 otherwise, x_k is the characteristic being analysed, in our case employment for each industrial sector and x_a is the average value of the characteristic. For statistical hypothesis testing, Moran's I values can be transformed into Z-scores in which values greater than 1.96 or smaller than -1.96 indicate spatial autocorrelation which is significant at the 5% level. The significance of the spatial autocorrelation of the Moran's I can be tested through a simple z-test.

3.3 Testing for border effects in industrial employment

We now move to the regression analysis on the determinants of change in industrial employment. The methodology is taken from Overman and Puga (2002), who look at the change in unemployment rates across European countries. The empirical strategy is described by the Equation (8):

$$\begin{aligned}\Delta\ \text{Employment}_{r,k,t-(t+1)} = {} & \alpha\ \text{Initial Employment}_{r,k,t} \\ & + \beta \sum_{n} \Delta\ \text{Employment (same region)}_{n,k,t-(t+1)} \\ & + \chi \sum_{m} \Delta\ \text{Employment (other region)}_{m,k,t-(t+1)} \\ & + \gamma\ \%\text{Literacy}_{r,t} + \delta\ \%\text{Agric. Employment}_{r,t} \\ & + \eta\ \%\text{Ind. Employment}_{r,t} + \psi\ \text{Region} \\ & + \phi\ \text{Industry} + \epsilon \end{aligned} \tag{8}$$

The dependent variable is the change in employment rate of province r in sector k between census years t and t+1. The employment rate in each sector is calculated as share of the provincial labour force working in that specific sector. Provinces are defined as neighbouring with respect to province r when they share a border with r. The explanatory variables are the initial industrial employment rate in province r in sector k, the change in the neighbouring provinces employment (indexed by n for "same region" neighbours and m for "other region" neighbours) and three provincial controls at time t: the skill levels (proxied by literacy rate) in province r and share of labour force in agriculture, and industry. Province and industry fixed effects are always included. The two neighbour effects are the weighted averages of the changes in industrial employment rates of neighbouring provinces with provincial labour force as weight.[7] The coefficient of the change in the neighbouring provinces employment represents what is called "neighbouring effect", meaning that if the coefficient of the employment of neighbours is positive and significant, the evolution in the industrial employment of a region tends to be close to that of nearby regions. If industrial clusters locate within regions, this coefficient will be non significant or negative, proving the existence of a border effect. This strategy is also applied by imposing the pre-unitary borders on the post-1861 data. It is done by simply

[7] A matrix where cells take value 1 when two provinces share a border and 0 otherwise was used to work out which provinces to include in the neighbour effect computation.

defining as neighbour "other state" the provinces that belonged to a different state and neighbours "same state" provinces that belonged to the same state before unification.[8]

4 Provincial industrial employment: new insights from the population censuses

The main primary sources upon which this work relies to quantify industrial employment are the population censuses of 1871, 1881, 1901 and 1911 and the industrial census of 1911.[9] The industrial census of 1911 is the only complete one in this period that recorded the number of plants along with other basic information at firm level.[10] The population census of 1891 was not conducted because of budget cuts, leaving us with a twenty-year gap in the data instead of the standard ten-year gap.

Although the data on industrial employment at worker level comes entirely from population censuses, these are not readily usable in their original format. Before being able to pool the different years, long and thoughtful reclassification for the four benchmark years is required to create homogeneous industrial sectors. This work was started over forty years ago by Vitali (1970). The resulting dataset started from 1881 and connected each of the professional categories of the censuses to fifteen industrial classes, homogeneous across years and broken down by region. Later on, this work at regional level was extended to 1871 by Fenoaltea (2001), with some adaptation of the industrial sectors to calculate regional value added.[11] In 2013, Ciccarelli and Fenoaltea presented industrial value added estimates for the 69 Italian provinces for the four census years, broken down by industrial sector.[12] This work introduced for the first time a provincial dimension in the study of the Italian industrial sectors. Ciccarelli and Fenoaltea based their estimates on the reclassified employment figures from the censuses but did not provide the underlying numbers. Following this work, Ciccarelli and Missiaia (2013) provided the full dataset of provincial industrial employment broken down by industrial sector and for the first

[8] A similar matrix to those of regions mentioned above was computed to calculate the pre-1861 neighbouring effects.

[9] MAIC (1871, 1881, 1901, 1911a, 1911b). The population census has also been used to obtain the size of each unit. Italy at the time was divided into 16 regions. The population of Italy was about 36,180,000 and its extent was 279,542 km^2. Therefore the average extent of a region was 16,443 sq^2 with 2,128,235 residents on average.

[10] The E-G index is calculated using the first industrial census in Italian history conducted in 1911. The census includes information of the number of firms in each administrative unit and their size in terms of employment.

[11] 1871 is one the most problematic censuses when it comes to standardization of the industrial sectors. On this point see Vitali (1970, p. 3) and Zamagni (1987, p. 37).

[12] Ciccarelli and Fenoaltea (2013).

time showing males and females separately.[13] The main contributions of this work, other than making the numbers available, is to provide an extensive discussion on the state of the art in the reconstructions and their shortcomings. In particular, the availability of both female and male workers allows us to assess what the authors call the "textile-bias-of-the-early-censuses". As previous authors have observed, the female labour force in the early censuses (most notably 1871 and 1881) is severely over represented. According to Vitali (1970), Zamagni (1987), Fenoaltea (2001) and Ciccarelli and Missiaia (2013), the over representation can be entirely imputed to the textile sector. To see this graphically, compare Figures 6 and 7. The former shows the share of industrial labour force in each province, net of textiles. The latter introduces textile workers. It is clear that when textile workers are included, the picture changes dramatically in 1871, 1881 and somewhat less markedly in 1901. According to Ciccarelli and Missiaia (2013, p. 148) the bias arises from the anomalous classification of the occupational activities of female workers. This bias is particularly severe in the Southern regions, such as Calabria, Apulia and Basilicata, where women in textiles reached as much as 80% of the industrial workforce of the region. Moreover, textile employment in the South included many more part-time and seasonal workers than in the North. Figure 8 shows the employment of female workers in textiles as a share of the total industrial employment in the different provinces. Their over representation in the South appears quite extreme, leading us to conclude that females in textiles bring in more bias than can be than can be used in our analysis.

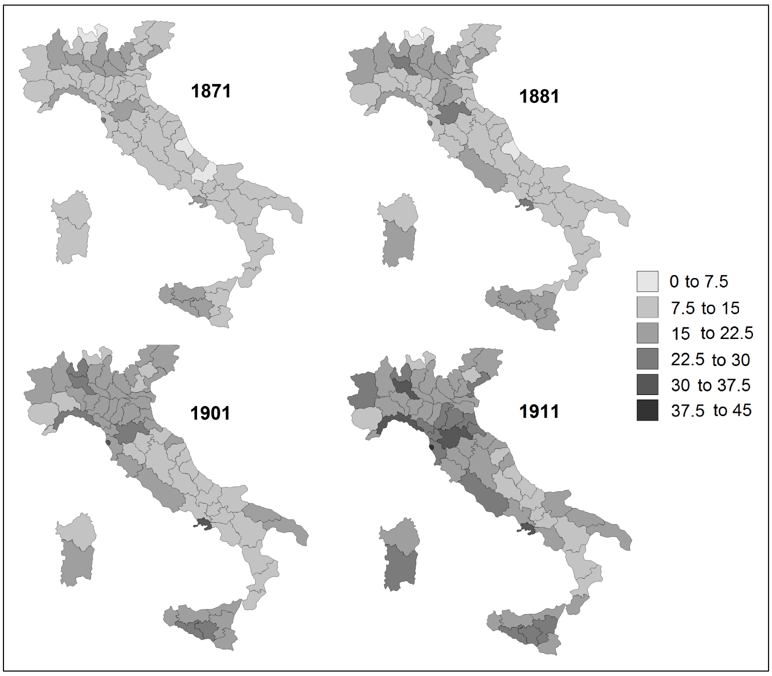
Scholars agree that the picture that arises from the employment figures in the census is so distorted that no analysis can proceed without a correction. In the literature two main corrections have been proposed. The first is the one by Zamagni (1987, p. 37–43); it is based on comparing the information contained in the population censuses with corresponding information in the industrial census of 1911 and other sources at firm level for 1876–1881 and 1901–1903.[14] It should be noted that industrial censuses (or official publications on industries, as for 1876–1881 and 1901–1903) report much lower figures than the population censuses do; therefore the assumption here is that the "true" number of workers lies somewhere between the lower bound of the industrial censuses and the upper bound of the population censuses.[15] The methodology also relies on the fact that the

[13] Although they were unpublished, the underlying figures for the industrial value added estimates were already completed, for the total industrial labour force, by Ciccarelli and Fenoaltea (2013). Ciccarelli and Missiaia (2013) extends the data further by including males and females separately.

[14] The two sources used are Ellena (1880) which contains information on some industries for 1876 and MAIC(1906) which provides a summary of the "industrial conditions" in the country.

[15] To give a sense of the difference, in the textile sector Ellena (1880) reports for 1876/1881 295,700 workers and the population census of 1881 reports 1,337,108; for 1901/1903 MAIC (1906) reports 408,404

Figure 6: Share of provincial industrial employment, 1871–1911 (net of employment in textiles).



Source: our calculations using employment data from Ciccarelli and Missiaia (2013) and MAIC (1874, 1883, 1902, 1914).

relationship between industrial censuses and population censuses becomes from the 1930s onward somewhat stable: around 110% of the figure in the population census equals 100% of the figure from the industrial census. Therefore, Zamagni applies a 110% coefficient to the industrial census data nearest in time and takes these as the value for textile workers whenever they do not exceed the population census figures.

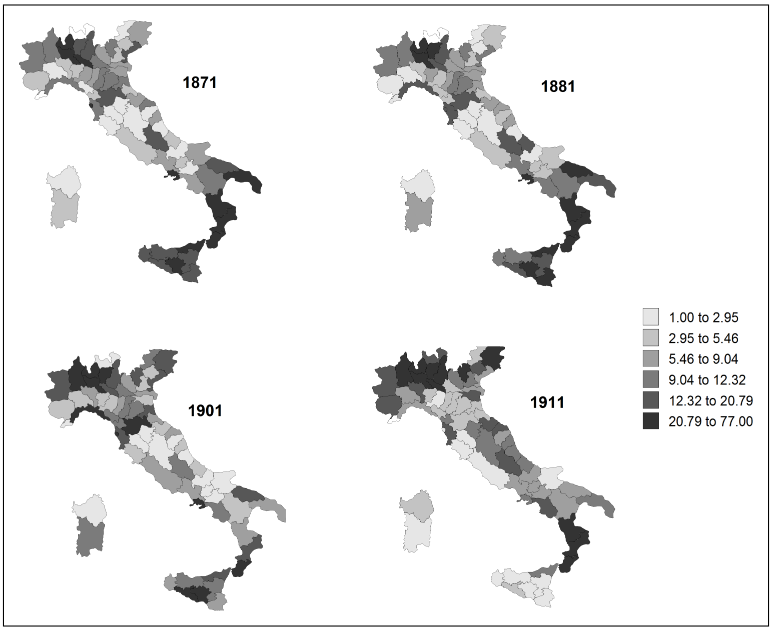
On the other front, Fenoaltea (2003b, p. 1083) corrects the textile figures calculating the number of workers in each region as the sum of males plus females capped at four females for each male. The 4:1 ratio is approximately the proportion of males to females in industrial employment in other sectors at the end of this period.[16] In this work we decided to follow the method by Fenoaltea for two main reasons. First of all, because we start our analysis in 1871, the first year of the dataset would not have been covered by an alternative industrial source.[17] Second, we are interested in quantifying the location of workers, which is simply a measure of where people physically are. This is a very different exercise from using employment data to produce, say, a value added estimate

and the 1901 population census reports 783,253; for 1911 the difference between the two censuses is 508,076 vs. 673,968.

[16] To give a sense of the difference between the two methods, following Zamagni we get for 1881 325,270 textile workers while following Fenoaltea we get 555,684; for 1901 the numbers are 449,244 vs. 514,285. For 1911, 558,883 vs. 502,920.

[17] Ellena (1880) starts in 1876, which being closer to 1881 than 1871 is used as a source for 1881.

Figure 7: Share of provincial industrial employment, 1871–1911 (without correction).



Source: our calculations using employment data from Ciccarelli and Missiaia (2013) and MAIC (1874, 1883, 1902, 1914).

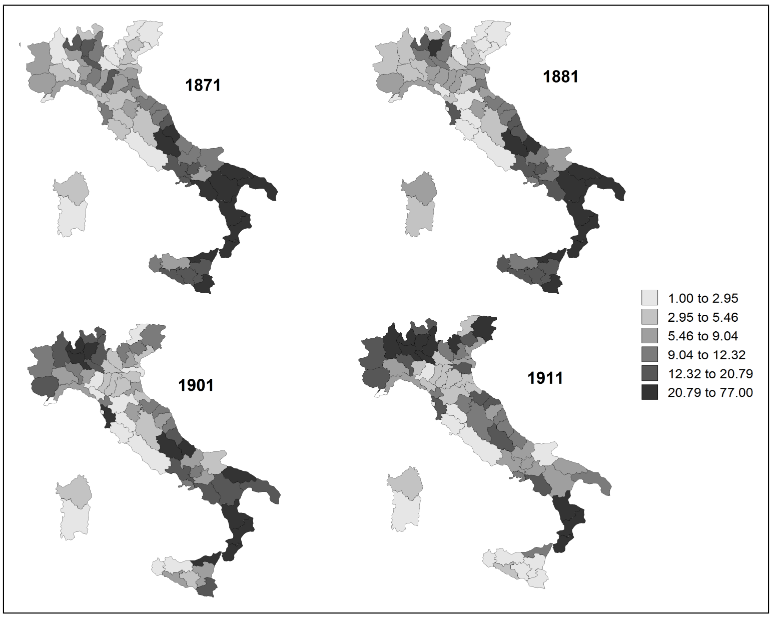
or number of hours worked.[18] These two measures, and in general any measure affected by hours per worker or labour productivity, are very sensitive to the inclusion of workers who might be seasonal, part-time or simply less productive. In our case, the inclusion of non-full time and low-productivity workers is expected, as long as they are actually working in a particular sector. Finally, other than Fenoaltea, the correction capping the number of females to four times the number of males has also been adopted by A'Hearn and Venables (2011) in their work on internal geography and external trade in the long run, which is probably the most similar work to ours in the literature on Italian Economic History literature. The resulting employment rates, with the correction for textiles are shown in Figure 9. These are the ones ones in this paper.

Other than industrial employment figures, this work uses two other variables that are unpublished at provincial level. The first one is literacy rates by province, which has been computed following the same methodology of A'Hearn et al. (2011) for the regions.[19] Figure 10 shows a map of the literacy rates at provincial level.

[18]For examples of census data used to assess the former, see the works by Fenoaltea and Ciccarelli (Fenoaltea (2001), Fenoaltea (2003b), Fenoaltea (2003a) and Ciccarelli and Fenoaltea (2013)) and the estimates of regional GDP by Felice and Brunetti (Felice (2009) and Brunetti et al. (2011)); for an example of estimation of total hours worked based on censuses see Giordano and Giugliano (2012).

[19]The literacy rates are computed on a population of at least 13 years of age for 1871 and 15 years for 1881, 1901 and 1911. The age group is not the same throughout the sample because of limitations in the sources; for a full discussion see A'Hearn et al. (2011, p. 205).

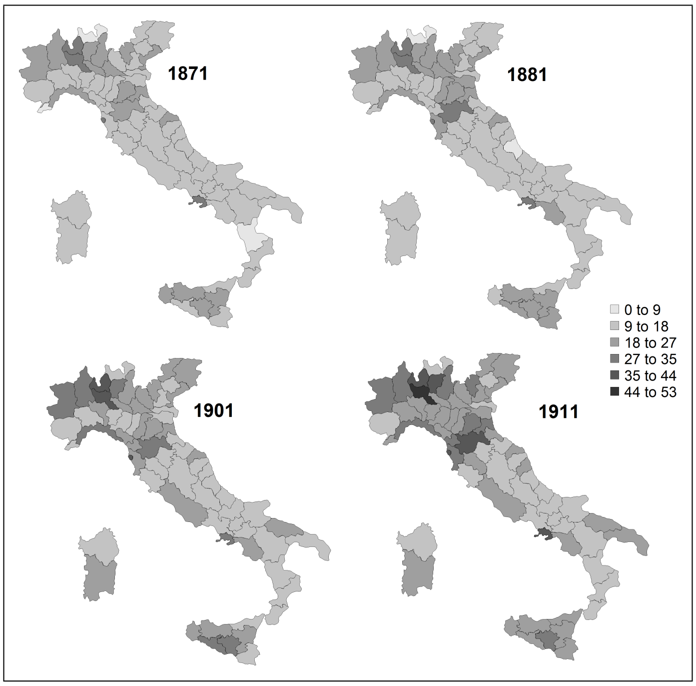
Figure 8: Female employment in textiles as a share of total industrial employment, 1871–1911.



Source: our calculations using employment data from Ciccarelli and Missiaia (2013) and MAIC (1874, 1883, 1902, 1914).

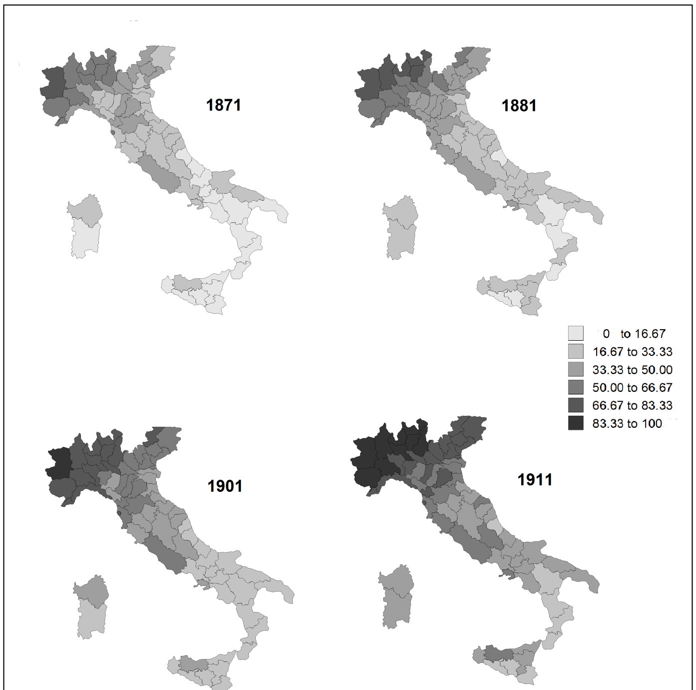
The last variable computed using population censuses is the share of the labour force in agriculture. The reference for this the classification by Vitali (1970, p. 298) for the years 1881–1911. For 1871 we included the the entire category of agriculture (categoria I), which appeared similar enough to the corresponding category of the other censuses. The results are shown in Figure 11.

Figure 9: Share of provincial industrial employment, 1871–1911 (with correction).



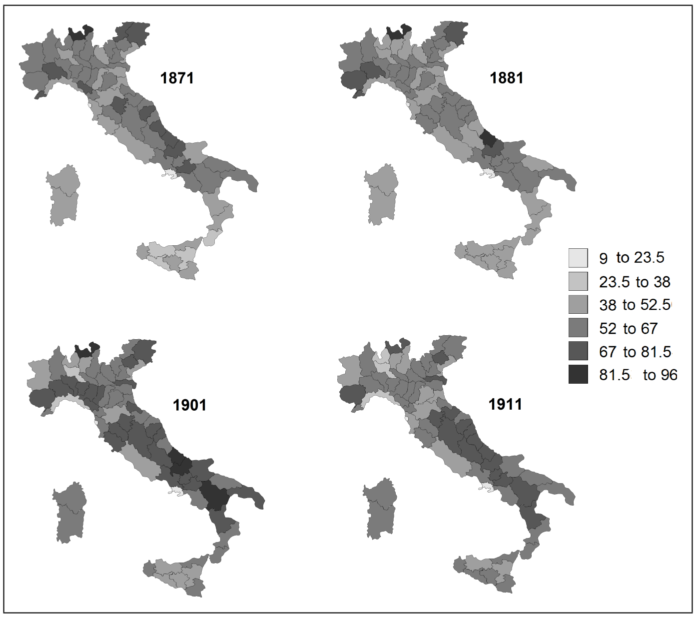
Source: our calculations using employment data from Ciccarelli and Missiaia (2013) and MAIC (1874, 1883, 1902, 1914).

Figure 10: Literacy rates in the Italian provinces, 1871–1911.



Source: MAIC (1874, 1883, 1902, 1914).

Figure 11: Share of provincial labour force in agriculture, 1871–1911.



Source: MAIC (1874, 1883, 1902, 1914).

5 Pattern of industrial employment and border effects: empirical results

This section illustrates the empirical results of both the descriptive indices and the model of the determinants of change in industrial employment.

5.1 Concentration

We start the analysis by looking at the concentration measures. Table 1 and Figure 12 show the Krugman concentration index for all benchmark years and all 15 industrial sectors. The Krugman concentration index indicates for each industrial sector, where it stands between two bounds: it takes value 2 when the sector is concentrated in one region and value 0 when it is equally distributed in all regions. In all benchmark years the values have a minimum value of about 0.70 and a maximum value of 1 for metalmaking. These numbers show a relevant degree of concentration throughout the period, with quite similar values across sectors.

A second index of concentration is the G index. The values are presented in Table 2 and Figure 13.

Unlike the Krugman index, the G index controls for the size of region, avoiding bias from size differences which would bias the results (Wolf, 2007, p. 31). The G index shows much more heterogeneity than the Krugman index. This difference is probably driven by the fact that the G index controls for the area of the regions when the Krugman controls only for the share of employment of a sector out of the total. All sectors present some concentration, with values going from a minimum of about 0.30 to a maximum of about 1. Sectors such as construction or foodstuff show lower concentration, since as expected, we see them present to some extent in all regions. Other sectors, more closely linked to local resources, such as mining, show a persistently higher value. There is on average an upward trend in the index, showing a mild increase of concentration through time. In some cases there is a much sharper increase in the concentration. This is, for example, the case of metalmaking. The reason for this is that at the beginning of the period this sector was quite small, composed of small and dispersed plants. Only in 1884 did a large steelworks company based in Terni (Umbria) start its activity, boosted by public funding Zamagni (1990, p. 128). An inverse path is followed by utilities, among which we find electric power production. In this case, Italy experiences an increase of the power installed from 86,175 to 1,286,883 kW Zamagni (1978, p. 89). The greater part of this power came from hydroelectric plants. The development of this sector was made possible through

Table 1: Krugman concentration index, 1871–1911.

	1871	1881	1901	1911
Mining	0.98	0.98	0.97	0.97
Foodstuffs	0.88	0.89	0.91	0.92
Tobacco	1.00	1.00	1.00	0.99
Textile	0.72	0.75	0.81	0.85
Clothing	0.78	0.73	0.82	0.79
Leather	0.87	0.87	0.87	0.90
Wood	0.88	0.88	0.89	0.88
Metalmaking	1.00	1.00	0.99	0.99
Engineering	0.91	0.91	0.89	0.87
Non-metallic mineral products	0.96	0.96	0.96	0.94
Chemicals, rubber	1.00	0.99	0.99	0.99
Paper, printing	0.99	0.99	0.98	0.98
Sundry manufacturing	0.99	0.99	0.99	0.99
Construction	0.88	0.88	0.81	0.80
Utilities	1.00	1.00	1.00	0.99

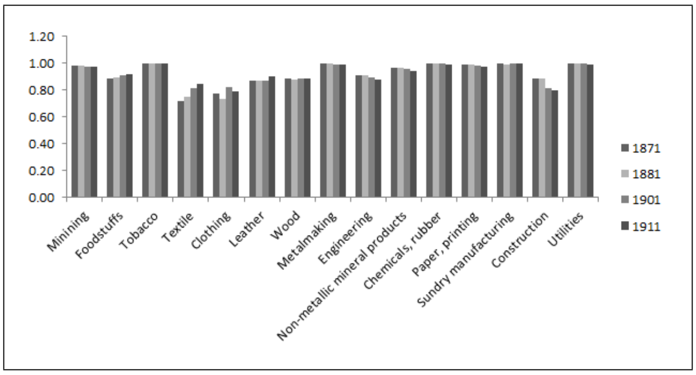
Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

the opening of new plants rather than enlargement of the existing ones, explaining the reduced concentration in the period. Summing up, both indices show a fair amount of concentration, in the mining industry in particular (which is quite predictable, given the characteristics of point resource extraction). The Krugman index shows less heterogeneity while the G index shows more differences across sectors.

As noted in Section 3.1, the two indices presented so far do not meet two of the three standard requirements prescribed by Duranton and Overman (2005, p. 1078) for an index to be suitable for measuring the concentration of industries. They are comparable across industries but they do not control for the tendency of manufacturing to agglomerate and for the degree of industrial concentration. This means that different localization schemes may be represented by the same concentration measure. The Economic Geography literature provides another index that controls for the size of plants and the level of geographic aggregation. The third option for measuring concentration comes from Ellison and Glaeser (1997). Their index can be calculated only for 1911, since data on the number of plants are taken from the 1911 industrial census.[20] This tool is useful in taking into account

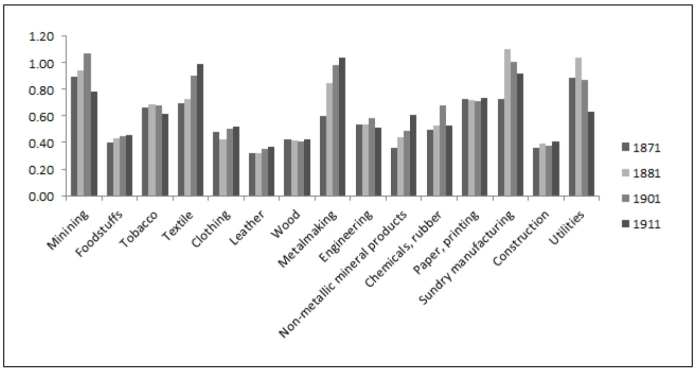
[20] To calculate E-G indices, ideally we should know the size of each plant. Here we replace plant size with the mean size of the observations for each region. This is because we do not have firm-level data.

Figure 12: Krugman concentration index, 1871–1911.



Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

Figure 13: G index of concentration, 1871–1911.



Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

the size of the plants as well as the size of the regions (in terms of their total industrial employment). It allows us to distinguish between the industrial concentration caused by market concentration (a few large plants) from concentration from agglomerative forces (many small plants co-located). According to Ellison and Glaeser (1997, p. 890), the former phenomenon cannot be regarded as proper concentration. These authors provide the example of the vacuum cleaner industry in the US, where 75% of the employees work in one of the four largest plants: according to the authors, this does not mean that the industry is concentrated. This is because they read the phenomenon of concentration as firms locating close to each other and claim that having few big plants may simply be related to economies of scale.

Table 3 and Figure 14 show the values of the index for the 15 industrial sectors.

Ellison and Glaeser (1997, p. 902) classify their index as follows: a sector is not very concentrated if the E-G index is smaller than 0.02; it is relatively concentrated if the E-G index is between 0.02 and 0.05 and it is highly concentrated over 0.05. The index here

This procedure is not optimal but still it is an improvement over indices that do not take into account the size of plants at all.

Table 2: G index of concentration, 1871–1911.

	1871	1881	1901	1911
Mining	0.89	0.95	1.07	0.78
Foodstuffs	0.40	0.44	0.45	0.46
Tobacco	0.66	0.69	0.68	0.62
Textile	0.70	0.73	0.90	0.99
Clothing	0.48	0.43	0.51	0.52
Leather	0.32	0.32	0.35	0.37
Wood	0.42	0.42	0.41	0.42
Metalmaking	0.60	0.84	0.98	1.04
Engineering	0.53	0.53	0.58	0.51
Non-metallic mineral products	0.36	0.44	0.49	0.61
Chemicals, rubber	0.50	0.53	0.68	0.53
Paper, printing	0.73	0.72	0.71	0.74
Sundry manufacturing	0.73	1.10	1.01	0.92
Construction	0.36	0.39	0.38	0.41
Utilities	0.89	1.04	0.87	0.63

Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

indicates a fairly high level of concentration. The index takes value zero if it deviates from what would be expected given a random distribution across space.

The results support the idea that Italy had a high concentration of industries at least in the final year of our period. All sectors are well above the 0.05 threshold of concentration. Mining, as expected, is the most concentrated industry. This is because mining is not necessarily organized in large plants but it tends to locate, in the case of Italy, in the few regions to benefit from natural resources endowment. Metalmaking, looks much less concentrated than the previous indices. This is due to the correction for the plant size proposed by Ellison and Glaeser (1997). Metalmaking was mostly carried on by a few large firms and the index eliminates the "few-large-plants" effect already described. We can now move to the specialization measures.

5.2 Specialization

The index used for specialization is a simple Krugman index as described in the previous section. For the specialization of regions we do not need to correct for plant size, so this task is somewhat simpler. Table 4 and Figure 15 show the values of the index.

Table 3: Ellison-Glaeser index, 1911.

	Ellison-Glaeser Index, 1911.
Mining	0.42
Foodstuffs	0.69
Tobacco	0.33
Textile	0.16
Clothing	0.15
Leather	0.26
Wood	0.26
Metalmaking	0.18
Engineering	0.28
Non-metallic mineral products	0.17
Chemicals, rubber	0.16
Paper, printing	0.14
Sundry manufacturing	0.18
Construction	0.14
Utilities	0.16

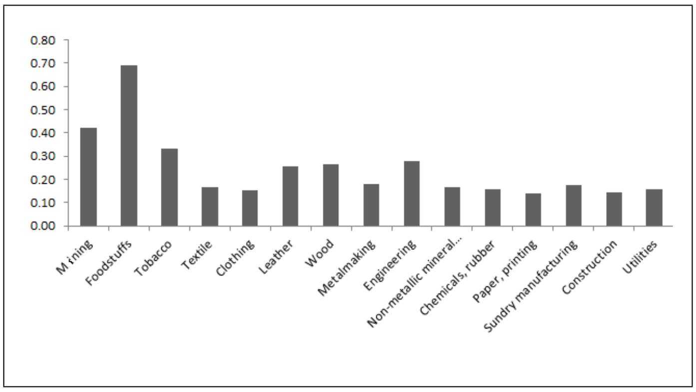
Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

The Krugman index of specialization is bounded by 0 (no specialization) and 2 (complete specialization). This index has been employed in Wolf (2007) to show patterns of specialization in interwar Poland. The index indicates for each region the degree of specialization with respect to the rest of the country. For the case of interwar Poland, the values range from a minimum of about 0.7 to a maximum of slightly less than 1. Italy shows similar values and all regions seen to some extent specialized in all the benchmark years. Values are between 0.84 and 1 in all years and the great majority of regions has values between 0.90 and 1. This index shows the picture of fairly high and constant levels of specialization by Italian regions over time. This index therefore confirms that the distribution of industrial activity in Italy in the post unification period was fairly far from being homogeneous.

5.3 Moran's I

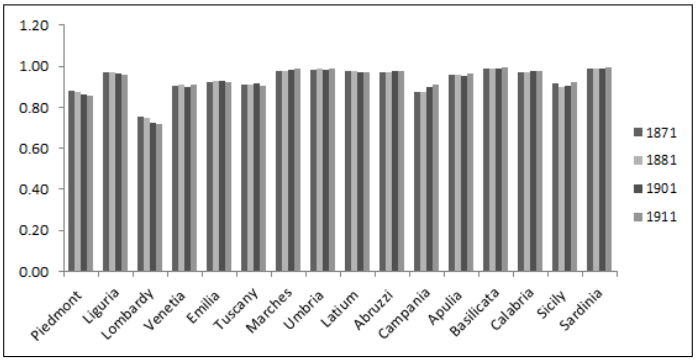
The tools provided so far look only at concentration in regions that we define as part of a broader area but with no any relationship with one another. However, these tools do not take into account the position of the regions with respect to each other. As

Figure 14: Ellison-Glaeser index, 1911.



Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

Figure 15: Krugman specialization index, 1871–1911.



Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

discussed in Section 3.1, the Moran's I is an index of spatial autocorrelation that introduces a spatial dimension across regions. Table 5 and Figure 16 below show the Moran's Is for all industrial sectors and benchmark years. The Moran's I shows positive spatial autocorrelation when the values are higher than the expected value. The expected value of the Moran's I is $E(I) = \frac{-1}{N-1}$ with N number of regions (Arbia et al. (2006, p. 17).). The Moran's I is bounded between E(I)-1 (perfect negative correlation with the neighbours) and E(I)+1 (perfect positive correlation with the neighbours). The statistical significance of the Moran's I is tested with a standard Z test at a 5% level. The numbers in Table 5 and Figure 16 are standardized so that E(I)=0. The results above show predominantly positive spatial autocorrelation.

All the values are significant to the Z test at a 5% level. The expected value in this case is 0.0625; therefore the maximum value that the Moran's I could have is 1.0625. Looking at Table 5, the values are almost all below 0.20. Although all the Moran's I values are significant, these values are not particularly high compared to other cases in which there

Table 4: Krugman specialization index, 1871–1911.

	1871	1881	1901	1911
Piedmont	0.88	0.88	0.86	0.86
Liguria	0.97	0.97	0.96	0.96
Lombardy	0.75	0.75	0.73	0.72
Venetia	0.90	0.91	0.90	0.91
Emilia	0.92	0.93	0.93	0.92
Tuscany	0.91	0.91	0.92	0.90
Marches	0.98	0.98	0.98	0.99
Umbria	0.99	0.99	0.99	0.99
Latium	0.98	0.98	0.97	0.97
Abruzzi	0.97	0.97	0.97	0.98
Campania	0.88	0.87	0.90	0.91
Apulia	0.96	0.96	0.95	0.96
Basilicata	0.99	0.99	0.99	0.99
Calabria	0.97	0.97	0.97	0.98
Sicily	0.91	0.90	0.90	0.92
Sardinia	0.99	0.99	0.99	0.99

Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

is a high spatial autocorrelation.[21] The level of spatial autocorrelation is also different in the various industrial sectors. Some sectors have quite low values, around 0.05. This is for example the case with mining, which is a sector that is very concentrated in a few regions (notably in Sardinia, which has all zeros in the spatial weight matrix since it is an island). Other sectors with lower levels of autocorrelation are leather and sundry manufacturing. In the latter case, this could be explained by the fact that employment in the sector is generally low and not widespread enough to create transregional clusters. The Moran's I is higher in sectors such as metalmaking and engineering, which were probably developed enough to cross regional borders. Utilities has higher values as well, probably because of the production of hydroelectric power across the Alpine regions. There are some cases of sectors having relatively large differences in different years. Given the generally low level of spatial autocorrelation and in some cases the low level of employment (as for the sundry manufacturing), small changes in absolute terms can cause relatively large swings in the Moran's I.

[21] See Arbia et al. (2006, p. 27), the Moran's I they obtain for manufacturing and services in the 1990s in Italy are higher, mostly between 0.10 and 0.30.

Table 5: Moran's I index, 1871–1911.

	1871	1881	1901	1911
Mining	0.06	0.06	0.06	0.03
Foodstuffs	0.12	0.10	0.13	0.17
Tobacco	0.13	0.16	0.12	0.04
Textile	0.06	0.08	0.06	0.13
Clothing	0.07	0.04	0.14	0.07
Leather	0.06	0.05	0.04	0.03
Wood	0.13	0.12	0.12	0.15
Metalmaking	-0.01	0.18	0.18	0.16
Engineering	0.13	0.16	0.17	0.23
Non-metallic mineral products	0.07	0.07	0.08	0.17
Chemicals, rubber	0.04	0.10	0.13	0.16
Paper, printing	0.16	0.13	0.14	0.13
Sundry manufacturing	0.08	-0.02	0.00	0.07
Construction	0.15	0.14	0.16	0.18
Utilities	0.15	0.14	0.23	0.17

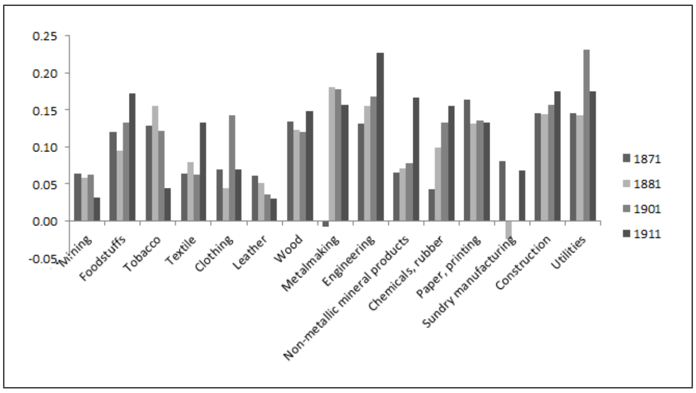
Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

The last point to make is on the relationship between concentration, specialization and spatial autocorrelation. The previous measures show a generally higher concentration and specialization and generally lower autocorrelation. These two results are compatible with a scenario of the high concentration of industries within the boundaries of the traditional regions. In general, it looks as though industries tended to concentrate at regional level and regions tended to specialize within their borders. This result is in line with the working hypothesis of the next section in which we test the regional border effects in the change of industrial employment.

5.4 The determinants of changes in industrial employment, 1871–1911

In this section we finally move to the regression model. Table 6 starts with a simple cross sectional OLS. The three periods are presented first with robust standard errors and no clustering and then in the second column with clustering at provincial level. All coefficients are in logs. All provinces and all sectors are pooled. All specifications include province and sector fixed effects. First, we notice that quite predictably the change in industrial employment in a province is largely explained by the starting employment rate in that province.

Figure 16: Moran's I index, 1871–1911.



Source: our calculations using employment data from Ciccarelli and Missiaia (2013).

The R^2 of all specifications and years is well above 60%, proving that we do indeed capture much of the variation. The second insight is that the two types of neighbour effects have different signs in all years. This confirms our working hypothesis of there being different effects of the change in employment in neighbouring provinces, defined as the provinces sharing a border with a given province, depending on whether they belong to the same region or not. In terms of changes in time, we see that the neighbour effects are stronger in earlier periods in terms of the size of the coefficient and also in terms of significance. If we believe that regions matter not because of the post-unitary arrangements but because of their historical meaning, the fact that earlier years show a stronger neighbour effect is predictable. In the next table, we show the same exercise pooling the sample and with panel regression. The reason why we are interested in pooling all the years when running the model is that, in spite of a fairly large overall sample (over 500 province-industry pairs), the number of cases of neighbouring provinces is a great deal more limited. It should be noted that the variables of interest for us are the neighbour effects rather than the controls. Therefore, pooling the three periods allows us to increase the number of cases under scrutiny.

Table 7 shows the pooled and panel specification with no clustering in the first column and region, province and industry clustering in the following sequence of columns in turn. The first thing that we notice is that the coefficients increase for the neighbour effects, in particular for the "same region" ones. The level of significance also increases to 1% for all specifications. Comparing the pooled regressions with the panel regression, we notice that the results on neighbour effects are very similar whereas the controls change. In particular, the share of the labour force in industry is negative and significant in the pooled regressions while the same result arises for the share of labour force in agriculture but not in industry in the panel. Also, in the panel, the literacy rate is positive and

Table 6: The determinants of changes in industrial employment, 1871–1911 (cross sectional OLS).

Log Change ind. employ.	1871–1881		1881–1901		1901–1911	
Log employment rate	0.803$^{(***)}$	0.803$^{(***)}$	0.678$^{(***)}$	0.678$^{(***)}$	0.525$^{(***)}$	0.525$^{(***)}$
	(0.0688)	(0.0831)	(0.0650)	(0.0708)	(0.0624)	(0.0597)
Log neighbour effect (same region)	0.0645$^{(*)}$	0.0645$^{(**)}$	0.102$^{(***)}$	0.102$^{(***)}$	0.0399	0.0399
	(0.0370)	(0.0314)	(0.0383)	(0.0350)	(0.0333)	(0.0280)
Log neighbour effect (other region)	-0.110$^{(***)}$	-0.110$^{(**)}$	-0.0516	-0.0516	-0.0416	-0.0416
	(0.0402)	(0.0432)	(0.0424)	(0.0371)	(0.0281)	(0.0287)
Log literacy	0.0887	0.0887	0.0476	0.0476	0.106	0.106
	(0.320)	(0.178)	(0.347)	(0.127)	(0.308)	(0.0978)
Log ind. LF	0.862$^{(*)}$	0.862$^{(***)}$	0.168	0.168	0.330	0.330$^{(***)}$
	(0.520)	(0.0657)	(0.409)	(0.135)	(0.345)	(0.0602)
Constant	-4.727$^{(***)}$	-4.727$^{(***)}$	-0.939	-0.939	-2.622$^{(*)}$	-2.622$^{(**)}$
	(1.469)	(1.147)	(1.210)	(0.719)	(1.370)	(1.090)
Clustering	no	province	no	province	no	province
Observations	534	534	576	576	536	536
R^2	0.678	0.678	0.687	0.687	0.624	0.624

Notes: Heteroskedastic robust standard errors in parentheses. $^{(*)}$ $^{(**)}$ and $^{(***)}$ correspond to a coefficient significantly different from zero with a 10%, 5% and 1% confidence level respectively. The dependent variable is the difference in the logs of the employment rates by industry by province. All explanatory variables are in logs. Neighbour effects are defined as the weighted sum of changes in employment in the regions (with weights equal to the relative size of the industry/province labour force). Two provinces are neighbours if they share a border. Neighbours belonging to the same region are separate from those belonging to another region. Literacy, the agricultural labour force and industrial labour force are expressed as rates.

significant. These differences may be due to some collinearity issue between the share of industry and agriculture in the labour force in the different sectors, but they may also be due to issues between these and literacy.[22]

To summarize our findings so far, what we observed is that the change in industrial employment of a province is mainly explained by the initial industrial employment level in the province plus some controls, such as the share of labour force in each sector, literacy rates and province and industry fixed effects. We also included the change in the employment of the neighbouring provinces, defined as provinces that share a border with the given province, and, following Overman and Puga (2002), we separated the neighbours belonging to the same region from the neighbours belonging to another region. We found that these have different signs. The interpretation for this result goes back to the specialization of the regions, which we found was quite high in the case of Italy. When

[22]We decided to keep both the share of the labour force in industry and that in agriculture in the regression to follow the example of Overman and Puga (2002) who also include a measure of human capital. As these are not necessarily the variables of interest but simple controls, whichever of these is actually included does not affect the analysis.

Table 7: The determinants of changes in industrial employment, 1871–1911 (pooled and panel OLS).

Log Change ind. employ.	Pooled				Panel			
Log employment rate	0.667$^{(***)}$	0.667$^{(***)}$	0.667$^{(***)}$	0.667$^{(***)}$	0.394$^{(***)}$	0.394$^{(***)}$	0.394$^{(***)}$	0.394$^{(***)}$
	(0.0368)	(0.0334)	(0.0265)	(0.0456)	(0.0661)	(0.0794)	(0.0810)	(0.126)
Log neighbour effect (same region)	0.0852$^{(***)}$	0.0852$^{(***)}$	0.0852$^{(***)}$	0.0852$^{(***)}$	0.0948$^{(***)}$	0.0948$^{(***)}$	0.0948$^{(***)}$	0.0948$^{(***)}$
	(0.0186)	(0.0191)	(0.0214)	(0.0221)	(0.0249)	(0.0239)	(0.0256)	(0.0279)
Log neighbour effect (other region)	-0.0799$^{(***)}$	-0.0799$^{(***)}$	-0.0799$^{(***)}$	-0.0799$^{(***)}$	-0.101$^{(***)}$	-0.101$^{(***)}$	-0.101$^{(***)}$	-0.101$^{(***)}$
	(0.0198)	(0.0176)	(0.0250)	(0.0208)	(0.0250)	(0.0240)	(0.0300)	(0.0208)
Log literacy	0.425	0.425	0.425	0.425	1.971$^{(***)}$	1.971$^{(***)}$	1.971$^{(***)}$	1.971$^{(***)}$
	(0.671)	(0.716)	(0.572)	(0.546)	(0.303)	(0.251)	(0.246)	(0.620)
Log agric. LF	-0.110	-0.110	-0.110	-0.110	-0.867$^{(**)}$	-0.867$^{(**)}$	-0.867$^{(*)}$	-0.867
	(0.469)	(0.358)	(0.433)	(0.625)	(0.423)	(0.328)	(0.437)	(0.726)
Log ind. LF	-1.062$^{(**)}$	-1.062$^{(**)}$	-1.062$^{(***)}$	-1.062$^{(*)}$	-0.357	-0.357	-0.357$^{(*)}$	-0.357
	(0.465)	(0.408)	(0.217)	(0.600)	(0.437)	(0.415)	(0.196)	(0.579)
Constant	0.622	0.622	0.622	0.622	-4.573$^{(**)}$	-4.573$^{(***)}$	-4.573$^{(**)}$	-4.573
	(3.210)	(3.107)	(2.477)	(3.063)	(1.781)	(1.691)	(1.770)	(3.246)
Clustering	no	region	province	industry	no	region	province	industry
Observations	1646	1646	1646	1646	1646	1646	1646	1646
R^2	0.599	0.599	0.599	0.599	0.164	0.164	0.164	0.164

Notes: Heteroskedastic robust standard errors in parentheses. $^{(*)}$ $^{(**)}$ and $^{(***)}$ correspond to a coefficient significantly different from zero with a 10%, 5% and 1% confidence level respectively. The dependent variable is the difference in the logs of the employment rates by industry by province. All explanatory variables are in logs. Neighbour effects are defined as the weighted sum of changes in employment in the regions (with weights equal to the relative size of the industry/province labour force). Two provinces are neighbours if they share a border. Neighbours belonging to the same region are separate from those belonging to another region. Literacy, the agricultural labour force and industrial labour force are expressed as rates.

a region tends to specialize, all its provinces will tend to have similar trends in the change of industrial employment, whereas provinces belonging to other regions will tend to have opposite trends to those of their competing neighbours. In the next tables we split the sample by macro area and then by industrial sector in order to go into depth about the relationship between industrial employment and border effects.

Table 8 shows the model run separately for the three macro areas.[23] The result of these three separate regressions is that most of the effect we observe when we pool all provinces comes from the North-East-Centre, where the coefficient is larger for the neighbour effect(in the same region) than in the other areas (over 0.1 when the North-West has coefficients below 0.05 and the South below 0.07) and is highly significant. The South presents a less pronounced neighbour effect but one still significant in the pooled regression. However, which basically represents the Industrial Triangle of Italy, does not show any significant neighbour effect. The explanation for this is two fold. First of all, Liguria and Piedmont were part of the same pre-unitary state; therefore it is expected that they would preserve their ties after Italy's Unification. However, Lombardy is also part of the North-West but this does not seem to drive any neighbour effect. The phenomenon can be explained through the similar economic trajectory that all three regions of the Industrial Triangle followed during the first Italian industrialization. Regarding the South, although the regions in this macro area were all part of the Kingdom of the Two Sicilies, it appears that the economic heterogeneity they experienced allowed for some neighbouring effect, albeit not as strong as in the North-East-Centre, where most of the variation is nested.[24]

The last variation on the model that we show is the one with pre-unification borders. Here the pre-1861 borders are applied to the post-1861 provinces. Therefore, two provinces are neighbours (in the same state) if they shared a border in 1871 and if they belonged to the same pre-unitary state before 1861. But, if they shared a border in 1871 but did not belong to the same pre-unitary state, they would fall under the heading of neighbours (in another state).[25] The regressions here are ran as pooled and panel OLS. Again the first column is with no clustering and the following three with three different types of clustering.

[23]The three areas are North-West, North-East-Centre and South, as in Felice (see Felice (2007) and subsequent works on the regional development of Italy.

[24]It should also be noted that because of the way that the three macro areas are designed, the variation among them is not evenly distributed. Therefore, the North-East-Centre embraces far more provinces and, most importantly, far more borders. The stronger results here are therefore expected not only for historical reasons but also because of the way that the three sub-samples are constructed.

[25]To illustrate with an example: the provinces of Alessandria and Genova do share a border but they belong to two different post-1861 regions (Alessandria is in Piedmont and Genova in Liguria). Therefore, in all the previous tables, their neighbour effects fall under neighbours (in another region). With pre-unitary borders they on the contrary fall under neighbours (in the same state), since both provinces belonged to the Kingdom of Sardinia before 1861. The same applies to all cases. To work out the matrix we used information on borders variations from ISTAT (2001).

Table 8: The determinants of changes in industrial employment, 1871–1911, by macro area (pooled and panel OLS).

Log Change ind. employ.	North–West		North–East–Centre		South	
	Pooled	Panel	Pooled	Panel	Pooled	Panel
Log employment rate	0.648$^{(***)}$	0.446$^{(***)}$	0.665$^{(***)}$	0.337$^{(***)}$	0.629$^{(***)}$	0.384
	(0.0591)	(0.0832)	(0.0419)	(0.109)	(0.109)	(0.251)
Log neighbour effect (same region)	0.0265	0.0424	0.113$^{(***)}$	0.125$^{(***)}$	0.0694$^{(*)}$	0.0601
	(0.0773)	(0.0858)	(0.0242)	(0.0281)	(0.0357)	(0.0414)
Log neighbour effect (other region)	-0.0814	-0.0696	-0.0645$^{(**)}$	-0.0821$^{(**)}$	-0.0858$^{(**)}$	-0.103$^{(*)}$
	(0.0650)	(0.0605)	(0.0269)	(0.0334)	(0.0368)	(0.0509)
Log literacy	-1.256	2.649$^{(***)}$	1.152	1.897$^{(***)}$	4.053$^{(*)}$	1.509$^{(***)}$
	(1.370)	(0.603)	(1.126)	(0.562)	(2.212)	(0.408)
Log agric. LF	0.120	-0.547	0.847	-0.0725	-0.179	-0.780
	(0.827)	(1.251)	(1.048)	(0.749)	(0.869)	(0.445)
Log ind. LF	-1.599$^{(*)}$	-0.531	-0.857	-0.330	-0.613	-0.752
	(0.847)	(0.723)	(0.813)	(0.929)	(0.824)	(0.731)
Constant	7.440	-9.205	-5.806	-6.879$^{(*)}$	-11.94	-2.351
	(5.002)	(5.585)	(7.161)	(3.282)	(7.934)	(1.743)
Clustering	province	province	province	province	province	province
Observations	486	486	722	722	438	438
R^2	0.635	0.210	0.585	0.201	0.625	0.090

Notes: Heteroskedastic robust standard errors in parentheses. $^{(*)}$ $^{(**)}$ and $^{(***)}$ correspond to a coefficient significantly different from zero with a 10%, 5% and 1% confidence level respectively. The dependent variable is the difference in the logs of the employment rates by industry by province. All explanatory variables are in logs. Neighbour effects are defined as the weighted sum of changes in employment in the regions (with weights equal to the relative size of the industry/province labour force). Two provinces are neighbours if they share a border. Neighbours belonging to the same region are separate from those belonging to another region. Literacy, the agricultural labour force and industrial labour force are expressed as rates.

The main result here is that the neighbour effect (same state) is positive and significant, while the neighbour effect (other state) is non significant. In terms of the other controls, we get similar results compared to the specification with post-1861 neighbour effects (share of industrial employment negative and significant for for the pooled and literacy positive and significant for the panel). Table 9 basically confirms the results obtained using post-unitary borders, although the neighbours in other states in this case seem to behave independently rather than in the opposite direction.

To conclude, this section has shown that the change in industrial employment is mainly explained by the initial industrial employment level in the province plus some controls (share of labour force in each sector; literacy rates; and province and industry fixed effects). Once we have controlled for all these factors, we can include in the model the change in the employment of the neighbouring provinces. Neighbours are defined as provinces that share a border with the given province. We took into account two types of neighbour: those

Table 9: The determinants of changes in industrial employment in Italy, 1871–1911, pre-unification neighbour effect (pooled and panel OLS).

Log Change ind. employ.	Pooled				Panel			
Log employment rate	$0.672^{(***)}$	$0.672^{(***)}$	$0.672^{(***)}$	$0.672^{(***)}$	$0.362^{(***)}$	$0.362^{(***)}$	$0.362^{(***)}$	$0.362^{(**)}$
	(0.0426)	(0.0396)	(0.0287)	(0.0530)	(0.0745)	(0.0872)	(0.0889)	(0.125)
Log neighbour effect (same state)	$0.0585^{(**)}$	$0.0585^{(**)}$	$0.0585^{(*)}$	$0.0585^{(**)}$	$0.0747^{(**)}$	$0.0747^{(**)}$	$0.0747^{(**)}$	$0.0747^{(**)}$
	(0.0238)	(0.0270)	(0.0306)	(0.0271)	(0.0290)	(0.0286)	(0.0253)	(0.0323)
Log neighbour effect (other state)	-0.0189	-0.0189	-0.0189	-0.0189	-0.0385	-0.0385	-0.0385	-0.0385
	(0.0226)	(0.0247)	(0.0274)	(0.0264)	(0.0280)	(0.0318)	(0.0294)	(0.0313)
Log literacy	0.614	0.614	0.614	0.614	$2.061^{(***)}$	$2.061^{(***)}$	$2.061^{(***)}$	$2.061^{(**)}$
	(0.822)	(0.697)	(0.707)	(0.660)	(0.363)	(0.340)	(0.303)	(0.724)
Log agric. LF	0.254	0.254	0.254	0.254	-0.569	-0.569	-0.569	-0.569
	(0.563)	(0.451)	(0.573)	(0.633)	(0.486)	(0.399)	(0.537)	(0.762)
Log ind. LF	$-1.156^{(**)}$	$-1.156^{(**)}$	$-1.156^{(***)}$	-1.156	-0.309	-0.309	-0.309	-0.309
	(0.588)	(0.559)	(0.354)	(0.833)	(0.546)	(0.590)	(0.390)	(0.752)
Constant	-0.968	-0.968	-0.968	-0.968	$-5.753^{(***)}$	$-5.753^{(**)}$	$-5.753^{(***)}$	-5.753
	(3.971)	(3.128)	(1.707)	(3.962)	(2.144)	(2.192)	(1.679)	(3.650)
Clustering	no	region	province	industry	no	region	province	industry
Observations	1201	1201	1201	1201	1201	1201	1201	1201
R^2	0.588	0.588	0.588	0.588	0.186	0.186	0.186	0.186

Notes: Heteroskedastic robust standard errors in parentheses. $^{(*)}$ $^{(**)}$ and $^{(***)}$ correspond to a coefficient significantly different from zero with a 10%, 5% and 1% confidence level respectively. The dependent variable is the difference in the logs of the employment rates by industry by province. All explanatory variables are in logs. Neighbour effects are defined as the weighted sum of changes in employment in the regions (with weights equal to the relative size of the industry/province labour force). Two provinces are neighbours if they share a border. Neighbours belonging to the same region are separate from those belonging to another region. Literacy, the agricultural labour force and industrial labour force are expressed as rates.

belonging to the same region and those belonging to another region and we included them separately in the regression. We found that these two have different signs, suggesting that regional borders do matter in attempts to explain the patterns of regional specialization. By splitting the sample into three macro areas we have shown that most of the strength of these border effects comes first from the North-East-Centre and second from the South. The Industrial Triangle seems to be acting as a unique "economic" region in terms of the evolution of its industrial sectors.

We believe that the post-unitary administrative arrangements alone cannot explain this border effect. The first Italian state, unlike today's, was quite centralized and if intermediate bodies had power, they were the "comune" in the first place and then the province. Regions did not have specific administrative powers and were mere collections of provinces for census purposes. Therefore, to test whether the effect we observe originates from the years before unification, we created some "counterfactual" border effects, imposing the definition of neighbour according to pre-1861 borders on the post-1861. The result is similar to the one with post-1861 borders, with a positive and significant effect of neighbours in the same state but no effect of neighbours in other states.

6 Conclusions

In the previous sections, we followed two main steps in order to study the geographical patterns of industries in the Italian regions in the period between the Unification and First World War.

We first presented different indices (Krugman, G and E-G indices) to measure the concentration of industries, the specialization of regions and spatial autocorrelation. The general results are that Italy experienced both a concentration of industries and specialization of its regions. These phenomena are, to some extent, present in all industrial sectors and all regions. The purpose of the exercise is first of all to establish whether there are some location patterns in the way that the different industrial sectors located themselves. Looking at standard measures of concentration and specialization, this appears to be the case. It is also useful to measure concentration and specialization at regional level in order to assess the role of regional borders. High levels of concentration and specialization suggest that borders do matter in the location patterns. The results for the third measure, the Moran's I, show a relatively low level of spatial autocorrelation among industries. Spatial autocorrelation has been introduced in the analysis to relate the dynamics of industrial location within the regions with the same dynamics in neighbouring regions. Spatial autocorrelation tells us whether regions tend to have more similar industrial sectors when they are closer to each other. The result of low spatial autocorrelation suggests that Italian industrial sectors tend to cluster more within regions than across regions.

The second step was to run a regression model to test whether the change in employment in a given province depended, all else being equal, on the change in neighbouring provinces. We sorted neighbouring provinces according to whether they belonged to the same region or not and found that provinces belonging to the same region had a positive and significant effect on employment while provinces belonging to different regions had a negative and significant effect. The interpretation of these results is connected with the results on specialization and concentration, for both steps of our methodology confirm that regions did matter in the location patterns for industrial sectors. We have claimed, in Section 2 that the role of post unification regions was minor compared to other smaller geographical units. This suggests that the importance of regional borders stems from pre-unitary arrangements. To test this, we imposed the pre-unitary borders on the post-unitary provincial industrial employment and repeated the exercise. The results were basically confirmed.

The findings presented in this paper are important for three reasons. First, they bring some insights onto the location patterns of industries using newly published provincial level data. The availability of these data allows us for the first time to look into a lower geographical unit and opens the door to further research in this direction. Second, it brings some historical insights into the impact of pre-unitary institutions on the post-1861 industrial patterns. And last, it assures us that the use of regions as unit of analysis in the literature has an economic rationale and is not merely a technical choice in the empirical analysis.

References

A'Hearn, B., C. Auria, and G. Vecchi (2011): *In ricchezza e in povertà*, Il Mulino, Bologna, chap. Istruzione.

A'Hearn, B. and A. Venables (2011): "Internal Geography and External Trade: regional disparities in Italy, 1861-2011," Economic History Working Papers 12, Bank of Italy.

Almagià, R. (1935): *Enciclopedia Italiana*, Treccani, Roma, chap. Regione.

Antonelli, P. and G. Palombelli (1995): *Amministrazioni pubbliche e territorio in Italia*, Il Mulino, Bologna, chap. Le province: la storia, il territorio.

Arbia, G., L. D. Dominicis, and H. D. Groot (2006): "Spatial Distribution of Economic Activities in Local Labour Market Areas: the Case of Italy," ERSA conference papers 06p497, European Regional Science Association.

Bonini, F. (1997): *Francesco Crispi e l'unità: da un progetto di governo un ambiguo mito politico*, Bulzoni, Roma.

Brunetti, A., E. Felice, and G. Vecchi (2011): *In ricchezza e in povertà*, Il Mulino, Bologna, chap. Reddito.

Cafagna, L. (1989): *Dualismo e sviluppo nella storia d'Italia*, Marsilio, Padova.

Ciccarelli, C. and S. Fenoaltea (2013): "Through the magnifying glass: provincial aspects of industrial growth in post-Unification Italy," *Economic History Review*, 66, 57–85.

Ciccarelli, C. and A. Missiaia (2013): "The industrial labor force of Italy's provinces: estimates from the population censuses, 1871–1911," *Rivista di Storia Economica*, 2, 141–192.

Ciccarelli, C. and T. Proietti (2011): "Patterns of industrial specialisation in post-Unification Italy," *Scandinavian Economic History Review*, 61, 1–28.

De Dominicis, L., G. Arbia, and H. de Groot (2007): "The Spatial Distribution of Economic Activities in Italy," Tinbergen Institute Discussion Papers 07-094/3, Tinbergen Institute.

Duranton, G. and H. G. Overman (2005): "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies*, 72, 1077–1106.

Ellison, G. and E. Glaeser (1997): "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy*, 105, 889–927.

Felice, E. (2007): *Divari regionali e intervento pubblico: per una rilettura dello sviluppo in Italia*, Il Mulino, Bologna.

——— (2009): "Estimating regional GDP in Italy (1871-2001): sources, methodology and results," Working Papers in Economic History 0907, Universidad Carlos III.

Fenoaltea, S. (2001): *La crescita industriale delle regioni d'Italia dall'Unità alla Grande Guerra: una prima stima per gli anni censuari*, Quaderni dell'Ufficio Ricerche Storiche, Bank of Italy.

——— (2003a): "Notes on the Rate of Industrial Growth in Italy, 1861-1913," *Journal of Economic History*, 63, 695–735.

——— (2003b): "Peeking Backward: Regional Aspects of Industrial Growth in Post-Unification Italy," *Journal of Economic History*, 63, 1059–1102.

Fotheringham, A., C. Brunsdon, and M. Charlton (2000): *Quantitative Geography: Perspectives on Spatial Data Analysis*, SAGE Publications, London.

Galinsky, K. (2005): *The Cambridge Companion to the Age of Augustus*, Cambridge University Press, Cambridge.

Giordano, C. and F. Giugliano (2012): "A Tale of Two Fascisms: Labour Productivity Growth and Competition Policy in Italy, 1911-1951," Economic History Working Papers 28, Bank of Italy.

Meriggi, M. (1987): *Il Regno Lombardo-Veneto*, UTET, Torino.

Overman, H. G. and D. Puga (2002): "Unemployment clusters across Europe's regions and countries," *Economic Policy*, 17, 115–148.

Pavone, C. (1964): *Amministrazione centrale e amministrazione periferica*, Giuffrè, Milano.

Spagnoletti, C. (1997): *Storia del Regno delle Due Sicilie*, Il Mulino, Bologna.

Vitali, O. (1970): *Aspetti dello sviluppo economico italiano alla luce della ricostruzione della popolazione attiva*, Failli, Roma.

Wolf, N. (2007): "Endowments vs. market potential: What explains the relocation of industry after the Polish reunification in 1918?" *Explorations in Economic History*, 44, 22 – 42.

Zamagni, V. (1978): *Industrializzazione e squilibri regionali in Italia: bilancio dell'età giolittiana*, Il Mulino, Bologna.

——— (1987): "A Century of Change: Trends in the Composition of the Italian Labour Force, 1881-1981," *Historical Social Research*, 44, 36–97.

——— (1990): *Dalla periferia al centro: la seconda rinascita economica dell'Italia (1861-1990)*, Il Mulino, Bologna.