The location of the Italian manufacturing industry, 1871-1911: a sectoral analysis

Roberto Basile* Carlo Ciccarelli†

May 10, 2017

Abstract

This study focuses on industrial location in Italy during the period 1871-1911, when manufacturing moved from an artisanal to a factory-based organization of production processes. There is general agreement in the historical and economic literature that factor endowment and domestic market potential represented the main drivers of industrial location. We test the relative importance of the above drivers of industrial location for 12 manufacturing sectors with data at the provincial level. Estimation results reveal that the location of traditional industries tied to the agricultural sector (such as leather and clothing) was mainly driven by water endowment, while the location of more dynamic "high tech" sectors (such as chemicals, cotton, metalmaking, and paper) was driven by domestic market potential and literacy.

Keywords: Market potential, Factor endowment, Concentration, Italy.
Jel codes: R12, R15, N13

*Department of Economics, Second University of Naples, Corso Gran Priorato di Malta, 1 - 81043, Capua (CE), Italy. *Email*: roberto.basile@unina2.it

†Department of Economics and Finance, University of Rome "Tor Vergata", Via Columbia 2, 00133 Rome, Italy. *Email*: carlo.ciccarelli@uniroma2.it

1 Introduction

Economic theory suggests that factor endowment and market access are the key determinants of industrial location. On the one hand, the neoclassical trade theory predicts that regional differences in factor endowments (such as mineral deposits, water supply, and labor skills) contribute to determine regional comparative advantages and, therefore, regional specialization. On the other hand, the New Economic Geography (NEG) literature suggests that an uneven spatial distribution of market access encourages firms to concentrate in regions with higher market potential, to benefit from increasing returns, and to export goods and services to other regions. These centripetal forces tend to contrast market competition (or centrifugal) forces arising from the concentration of firms and inducing to both lower local market prices and higher local factor prices. The equilibrium between these offsetting forces mainly depends on the degree of market integration and transportation costs. Under autarky, each region produces essentially the goods that consumes, the location of industries is stable, and the level of industrial concentration is low. When trade costs decrease and product markets tend to integrate, the neoclassical trade theory predicts that regional specialization will arise as regions produce and export products that are relatively intensive in their abundant resources (comparative advantages). Moreover, when transportation costs decrease, the NEG theory predicts that industrial activities characterized by increasing returns to scale tend to concentrate in the regions with higher demand (*home-market effect*), while the remaining regions suffer deindustrialization (Krugman, 1991). Therefore, a regional division of labor spontaneously arises through a process of uneven development.

The relative effect of factor endowment and market access on industrial location can be hardly quantified using empirical data referring to modern economies, as the two forces tend to coexist and interact in a very complex way along with the effect of (endogenous) policy interventions. Midelfart-Knarvik and Overman (2002) show for instance that the European industrial policy influenced strongly the industrial location patterns across the EU regions. As long as industrial policies are endogenously driven by the actual spatial distribution of economic activities, it can be quite difficult to quantify the genuine (net of industrial policies) effect of comparative advantages and/or market potential on industrial location. However, the use of historical data covering the years following the political unification of a country (that is when the process of domestic market integration was moving its first steps, and no systematic regional industrial policy took place), may provide an opportunity to better contrast the different explanations for the spatial concentration of industry and, in particular, to appreciate the increasing role of the home-market effect. Industrialization processes, the fall in transport costs and the integration of domestic markets may indeed generate the agglomeration forces that change the distribution of economic activity across space and reinforce spatial disparities over time. As a matter of fact, the economic history literature has increasingly drawn on NEG models to analyze national historical experiences and, due to its rising importance, the manufacturing sector has received most of the attention. Examples of studies in this direction are Wolf (2007) for Poland, Klein and Crafts (2012a) for the US, Rosés (2003), Tirado et al. (2002) and Martinez-Galarraga (2012a) for Spain, and Crafts and Mulatu (2005) for Britain. As far as Italy is concerned, A'Hearn and Venables (2013) explores the interactions between external trade and regional disparities since the unification of the country (1861). The authors argue that the economic superiority of Northern regions over the rest of the country was initially based on natural advantages (in particular the endowment of water), while from the late 1880s onwards domestic market access became a key determinant of industrial location, inducing the more dynamic industrial sectors to locate in regions with a large domestic market, that is, essentially, in the North. From 1945 onwards, with the gradual process of European integration, foreign market access represented instead the decisive factor; and the North, again, had the advantage of proximity to these markets.

While we broadly agree with the periodization proposed by A'Hearn and Venables (2013), we believe that it requires further qualification. Specifically, we claim that the relative importance of factor endowment and market potential for industrial location varies according to the technology prevailing in the various sectors. In particular, market potential, in line with Crafts and Mulatu (2006) for 19th century Britain, is expected to be more important in industries characterized by increasing returns to scale (typically, high and medium capital intensive industries), while factor endowment should be more relevant in the remaining manufacturing sectors.

On the basis of these considerations, we analyze the spatial location patterns of the various branches of the Italian manufacturing industry during 1871-1911. Specifically, we assess the relative importance of factor endowment (water abundance and labor skills) and market potential for industrial location behavior in the early phases of Italian industrialization distinguishing among the various manufacturing sectors. Our analysis is based on a recent set of data on value added at 1911 prices at the provincial level (69 provinces) for 10 manufacturing sectors, recently produced by Ciccarelli and Fenoaltea (2013, 2014b). Furthermore, we further disaggregated value added data for the textile and engineering industries into their sectoral components, that is, respectively, cotton, silk, wool, and other natural fibers for textile, and shipbuilding, machinery, and blacksmiths for engineering.

Our results clearly show that as transportation costs decreased and barriers to domestic trade were eliminated, Italian provinces became more and more specialized, and manufacturing activity became increasingly concentrated in a few provinces, mostly belonging to the North-West part of the country. The estimation results corroborate the hypothesis that both comparative advantages and market potential have been responsible of this process of spatial concentration. The location of some traditional industries tied to agriculture and characterized by low capital intensity (such as clothing and leather) was mainly driven by water endowment. The location of more dynamic technology-prone sectors characterized by high capital intensity (such as metalmaking, chemicals, and paper) was driven by the domestic market potential and literacy. Domestic market potential also affected the location of engineering (albeit only in the second part of our sample period) and textile, two sectors characterized by an intermediate level of capital intensity.

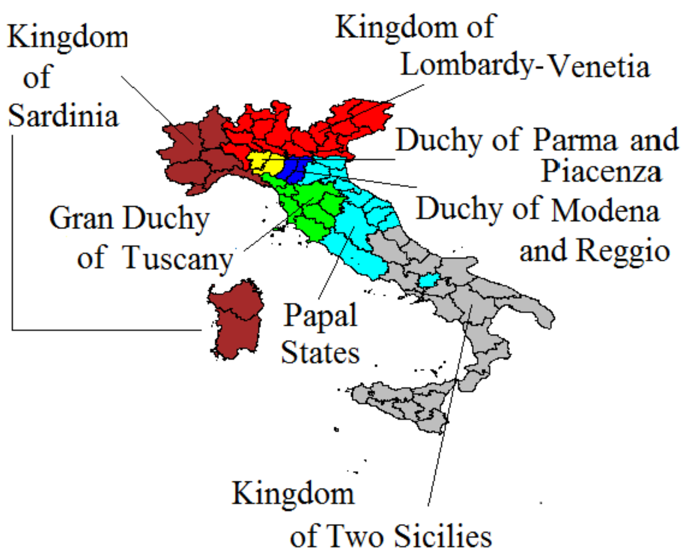
The rest of the paper is organized as follows. Section 2 illustrates the spatial distribution of manufacturing industry over the sample period. Section 3 describes the hypotheses on the key divers of industrial location. Section 4 reports the estimation results. Section 5 concludes.

2 The spatial diffusion of manufacturing activity in Italy: 1871-1911

2.1 Setting the scene

Figure 1 illustrates the division of the Italian territory that roughly prevailed during the 1815-1860 period. The seven pre-unitarian States (Kingdom of Sardinia, Kingdom of Lombardy-Venetia, Duchy of Parma and Piacenza, Duchy of Modena and Reggio, Gran Duchy of Tuscany, Papal States, and the Kingdom of Two Sicilies) were characterized by extremely different institutions and economic policies. The coin, monetary regimes, and trade policies were different. Primary school was mandatory only in certain pre-unitarian States, mainly those in the North.

FIGURE 1: Italian provinces at 1911 borders grouped into pre-unitarian States (1815-1860 ca.)



Italy was unified in 1861, although Venetia and Latium were annexed to the country only in 1866 and 1870, respectively. Between 1861 and 1870 the national capital was moved from Turin to Florence and, finally, to Rome. Soon after the political unification, policy makers realized that there was an urgent need of statistical evidence. Decennial population censuses were established, and dozens of annual reports to the Italian Parliament and other official publications concerning the main economic sectors (public budget and taxation, international trade, railroads, public school system) were regularly produced. The new official historical statistics divided the Italian territory into 16 regions (*compartimenti*, roughly NUTS-2 units), and 69 provinces (*province*, roughly NUTS-3 units). The borders of these administrative units (shown in Figure 1) did not change during 1871-1911.

2.2 The structure of economic activity

Table 1 shows that during 1871-1911 the composition of total gross value added (GVA) changed considerably, with industry (and moderately so the services) gaining shares at the expense of agriculture.[1]

TABLE 1: The changing composition of Italy's GVA: sectoral shares (percentages)

	1871	1881	1891	1901	1911
Agriculture	49.11	47.27	45.72	44.11	38.41
Industry	15.73	16.87	18.20	19.29	24.46
Manufacturing	*12.37*	*13.08*	*14.09*	*15.73*	*18.98*
Services	35.16	35.86	36.08	36.6	37.13
Total	100.00	100.00	100.00	100.00	100.00

Source: authors' elaboration on Fenoaltea (2005), providing annual 1861-1913 national value-added estimates at 1911 prices. Industry includes four major aggregates: 1. extractive, 2. manufacturing, 3. construction, 4, utilities.

Table 2 focuses on manufacturing and its sectoral components, and reports value added percentage shares at benchmark years (1871, 1881, 1901, and 1911) when population censuses were taken. With few exceptions sectors tied to the production of consumption goods (roughly those from 2.1 foodstuffs to 2.6 leather) show a constant reduction in their share. Sectors tied to the production of durable goods (roughly those from 2.7 metalmaking to 2.11 paper) show an opposite long-term trend, with a rapid acceleration in the 1901-1911 decade. The last column of Table 2 summarizes 1871-1911 trends, with numbers below one indicating a reduction of the sectoral share (that is a growth rate of value added smaller than the average one) and vice versa.

In 1911, foodstuffs, textile, and engineering represented alone more the 50 percent of total value added in manufacturing. Effectively, Ciccarelli and Proietti (2013) show that these three sectors alone are able to capture much of the variability of sectoral specialization at the provincial level, and act thus as a sort of "sufficient statistics" for the whole manufacturing industry during 1871-1911. In addition, Fenoaltea (2014) documents how vast and heterogeneous was the engineering sectors and Rosés (2003) bases his result for Spain on separated component of the textile industry. For these reasons, as detailed in the data appendix, we disaggregated the data on value added for the textile and engineering sectors into separated sub-sectors (this was not instead possible for foodstuffs due to lack of historical data). Table 2 shows clearly the importance of this disaggregation. Within textile, only cotton, more suitable to mechanization then other fibers, increased substantially its value added share over time. Within engineering, the size of the machineries component increased substantially. At the same time however, traditional blacksmith activities, despite a declining trend, accounted for about half of value added of the engineering sector even at the end of our sample period. After all, 19th century Italy was much an agricultural country and the maintenance of agricultural tools (such as spades, hoes, and ploughs) represented a substantial part of blacksmiths' traditional activity.

[1] Italy resembles in this respect other Southern European countries. Differently in North-west European countries, above all England, in 1911 ca had only about one fifth of the labor force employed in the agricultural sector (Broadberry et al., 2010).

TABLE 2: Manufacturing sectors: value added shares[a]

sectors	1871	1881	1901	1911	1911/1871
2.1 foodstuffs	33.5	30.4	25.4	21.5	0.6
2.2 tobacco	1.6	1.3	0.9	0.7	0.4
2.3 textile:	10.3	10.3	12.8	11.1	1.1
2.3.1 cotton	*1.3*	*1.9*	*5.2*	*4.8*	3.7
2.3.2 wool	*1.8*	*2.0*	*2.4*	*2.3*	1.3
2.3.3 silk	*3.9*	*3.8*	*4.3*	*3.3*	0.8
2.3.4 other natural fibers	*3.2*	*2.5*	*0.9*	*0.7*	0.2
2.4 clothing	6.9	7.4	6.8	6.3	0.9
2.5 wood	10.0	9.3	9.7	10.0	1.0
2.6 leather	10.5	11.5	11.4	7.8	0.7
2.7 metalmaking	0.6	1.0	1.7	3.1	5.2
2.8 engineering:	17.5	17.8	18.7	21.5	1.2
2.8.1 shipbuilding	*2.0*	*1.4*	*3.0*	*1.9*	1.0
2.8.2 machinery	*2.9*	*4.3*	*7.4*	*10.5*	3.6
2.8.3 blacksmith	*12.5*	*12.2*	*9.1*	*9.0*	0.7
2.9 non-metallic mineral products	3.6	4.2	4.2	6.6	1.8
2.10 chemicals	2.1	2.4	3.0	4.3	2.0
2.11 paper	2.7	3.5	4.8	6.3	2.3
2.12 sundry	0.7	0.7	0.6	0.7	1.0
2. total manufacturing	100.0	100.0	100.0	100.0	1.0

[a] The table includes 12 manufacturing sectors (numbered 2.1 to 2.12, as it is customary in the national account, where usually 1 refer to the extractive sector, 2 to manufacturing, 3 to constructions, and 4 to the utilities). Textile (2.3) and Engineering (2.8) are further disaggregated into 2.3.1 to 2.3.4, and 2.8.1 to 2.8.3. Numbers need not to add due to rounding.

2.3 Regional specialization and geographical concentration of manufacturing activity

How similar was the industrial structure across different provinces? Did it change during the 1871-1911 period? How concentrated was the manufacturing activity as a whole, and how concentrated was a given industry? Which industries tended to agglomerate, and which ones were instead dispersed? Did concentration increase over time? To address these question we use the disaggregated data on manufacturing value added at 1911 prices in province i, sector k, and time t, $v_i^k(t)$. These are in particular used to compute and analyze three standard indices: the Krugman's specialization index (K_i), the Theil concentration index (C_k), and the Moran I measure of spatial autocorrelation of the location quotient LQ_i^k (see Appendix for details). This latter is defined, at any time t, as $LQ_i^k = s_i^k * (v_i/V)$, where the specialization index has the usual definition $s_i^k = v_i^k/v_i$, and $v_i = \sum_k v_i^k$ and $V = \sum_i \sum_k v_i^k$ represent,respectively, provincial and national value added. At any time t, the index K_i provides a measure of the difference in the specialization (hence the i index) of a given province when compared to the remaining provinces of the country. As it is well known, it is bounded between 0 (no difference) and 2 (when a province has no industry in common with the rest of Italy). The Theil index of concentration (C_k) is based on the normalized location quotient, that is LQ_i^k normalized by the ratio between v_i (manufacturing va in province i) over V (manufacturing va in Italy). The relative Theil index C_k provides useful information about the extent to which industries are concentrated in a limited number of areas, but it does not take into consideration whether those areas are close together or far apart. In other words, it does not take into account the spatial structure of the data. As pointed out by Arbia et al. (2005), Basile and Mantuano (2008), and Arbia et al. (2013), a more accurate analysis of the spatial distribution of economic activities requires the combination of traditional measures of geographical concentration and methodologies that account for spatial dependence, in that they provide different and complementary information about the concentration of the various sectors. Among the spatial dependence measures the most widely used is the Moran's I index based, as it is well known, on a comparison of LQ_i^k at any location with the value of the same variable at surrounding locations. The spatial structure of the data is formally expressed in a spatial weight matrix W (Anselin, 2013) with generic elements w_{ij} (with $i \neq j$) (see Appendix for further important details on the W matrix). Table 3 shows that the three indices increased monotonically over time, reflecting increasing provincial specialization, sectoral concentration, and spatial autocorrelation.

TABLE 3: Specialization, concentration, and spatial dependence: manufacturing industry.

	1871	1881	1901	1911
Krugman	0.32	0.35	0.41	0.42
Theil	0.03	0.04	0.07	0.08
Moran's I	-0.04	0.15	0.00	0.05
	(0.47)	(0.41)	(0.09)	(0.02)

Number in brackets represent p-value (H0: no spatial autocorrelation).

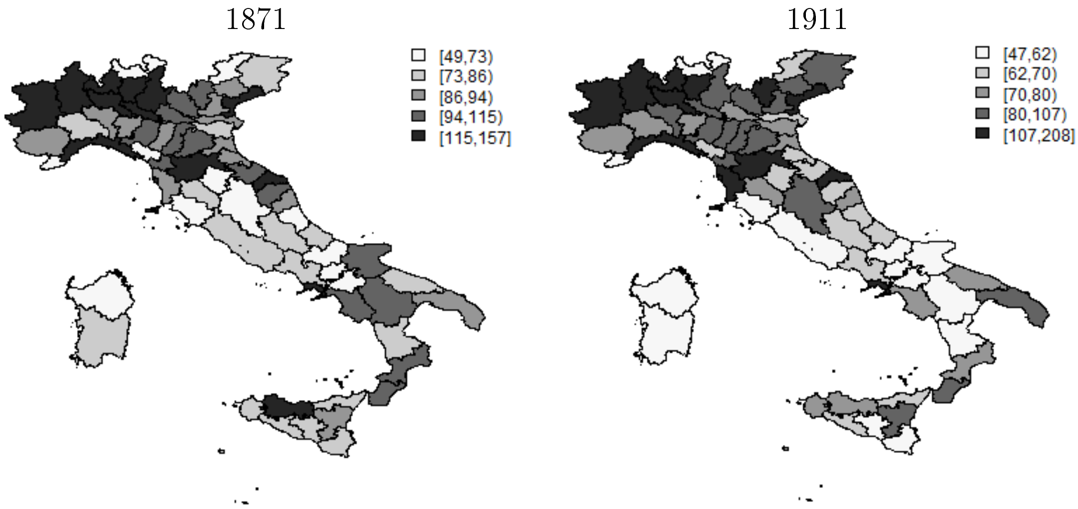
The Theil and Moran'I indices are,as recalled, directly related to the location quotient. For this reason it is interesting to briefly comment on the the choropleth maps of the spatial distribution of LQ values in 1871 and 1911, the end points of our investigation period (see Figure 2). In 1871, manufacturing activity mainly clustered in a few Northern provinces (Milan, Turin, Novara, Bergamo, Como, Brescia, Verona, Vicenza, Genoa, Leghorn and Venice) as well as in Florence, Ancona, Naples and Palermo.[2] Actually, although in 1871 the South as a whole was less industrialized than the North, Naples registered the second highest LQ value in manufacturing (1.41) after Milan (1.57) and higher than Turin (1.35), while Palermo had a LQ value in manufacturing (1.32) higher than Genoa (1.20). In 1911, the North-West confirmed its dominant position, while most of the Southern provinces worsened their relative position.

The consideration made so far are not informative on sectoral developments. Following Guillain and Le Gallo (2010), Figure 3 combines the a-spatial Theil index and the Moran's I index in a scatterplot for 1871 and 1911.[3] The vertical and a horizontal dashed line denote median values and identify four patterns in the distribution of economic activities: (i) HL (high concentration and low spatial dependence), (ii) HH (high concentration and

[2] See Appendix for a map illustrating Italian provinces.

[3] We excluded tobacco and sundry.

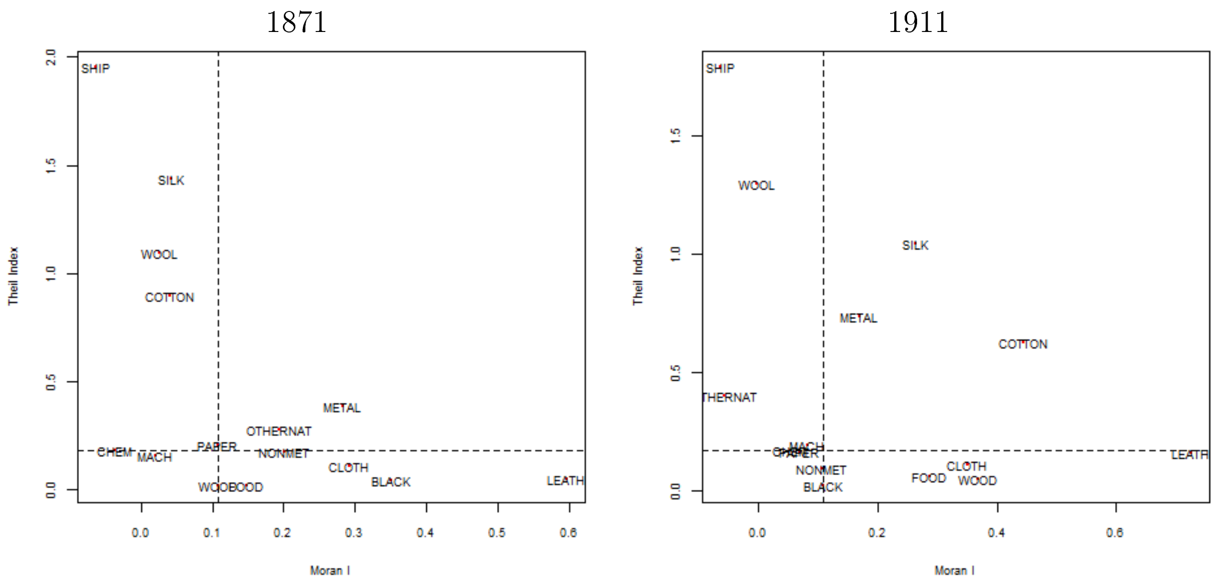
Figure 2: **MAN**: Choropleth maps of LQ values



high spatial dependence), (*iii*) LH (low concentration and high spatial dependence), and (*iv*) LL (low concentration and low spatial dependence).

At this level of sectoral disaggregation, shipbuilding and leather appear the most extreme and specular cases. Shipbuilding belongs both in 1871 and 1911 to spatial pattern *i* (high Theil, high Moran). This means that this industry concentrated its activity in a small number of (coastal) areas that were not close to each other. In this kind of sectors, indeed, economies of scale are reached by increasing the plant size and concentrating the production in a small number of locations. Leather, both in 1871 and 1911 belongs to spatial pattern *iii*, and is thus characterized by low Theil and high Moran. Other sectors kept their position over time. Capital intensive metalmaking belongs in 1871 and 1911 to spatial pattern *ii*, characterized by high concentration and high spatial dependence. Traditional economic activities much tied to the agricultural sector such as clothing, foodstuffs, wood, but also blacksmiths and non-metallic mineral products belong, although with some difference, to spatial pattern *iii*, as already leather. Within textile, silk and cotton, always characterized by above average Theil index, moved from spatial pattern *i* to spatial pattern *ii*, becoming strongly agglomerated in 1911, a feature that characterizes the textile sector even in more recent periods (see Arbia et al., 2013). On the contrary the other natural fibers components moved from spatial pattern *ii* to spatial pattern *i*. The above reallocation of the textile industry in the 1871-1911 period is consistent with Fenoaltea (2011). Within the engineering sector, shipbuilding, machinery and blacksmith differ tremendously in terms of concentration, but also on the spatial autocorrelation dimension. This confirm the importance of considering further disaggregated data for this industry. One further note that fast growing sectors such as machinery, paper, and chemicals present in 1911 about median value of Theil and Moran. Finally, one notice a substantial continuity in the ranking of sectors along the vertical dimension (with traditional sectors like foodstuffs, clothing, blacksmiths, but also wood the most dispersed sectors). Indeed, the ranking of the Theil indices remained very stable over time (the rank correlation between the Theil indices in 1871 and in 1911 is 0.95 with a p-value of 0.000). Interestingly, the stability in the level of concentration across industries

Figure 3: **Spatial concentration**: scatterplot between the a-spatial Theil index of concentration (y-axis), and the global I (x-axis).



Moran index of spatial autocorrelation. Value added data. FOOD="2.1 foodstuffs", COTTON="2.3.1 cotton", WOOL="2.3.2 wool", SILK = "2.3.3 silk", OTHERNAT="2.34 other natural fibers", CLO="2.4 clothing", WOOD="2.5 wood", LEATH="2.6 leather", METAL="2.7 metalmaking", SHIP="2.8.1 shipbuildings", MACH="2.8.2 machinery", BLACK="2.8.3 blacksmith", NONMET="2.9 non-metallic mineral products", CHE="2.10 chemicals and rubber", PAP="2.11 paper".

observed for the 19th century Italy is a pattern also found in Crafts and Mulatu (2006) for 19th century Britain, and in studies concerning more recent periods, both for Italy (Arbia et al., 2013) and other countries (Alonso-Villar et al., 2004; Devereux et al., 2004; Dumais et al., 2002).

3 Drivers of industrial location

In this section we provide possible measures of water endowment, literacy and market potential and posit some specific hypotheses on their role as main determinants of industrial location.

3.1 Factor endowment

3.1.1 Water supply

The central role of natural endowment (water, above all) for industrial location in 19th century Italy (Cafagna, 1989; Bardini, 1998) was recently reiterated by Fenoaltea (2011), providing perhaps the most careful and sharp account of Italian industrialization over the 1861-1913 period. According to Fenoaltea, "the roots of the success of the Northern regions seem [...] environmental rather than historical" (p. 231): factor endowment, thus, more than socioeconomic variables such as social capital or institutions. Modern (or factory-based) system of production replaced gradually artisanal production and "gave a strong advantage to the locations with a year-round supply of water (for power and also in the specific case of textiles, for the repeated washing of the material); and in Italy such locations abound only on the northern edge of the Po valley, where the Alpine run-off offsets the lack of summer rain". And, when analyzing the evolution of the location of the textile industry during 1871-1911, Fenoaltea (2011, p. 232) mentions among the natural advantages of northern regions "the water that flowed in the rivers, the water suspended in the air [...]" and the presence of "mountain glaciers". Beyond representing a source of motive power and adequate climate (as humidity for textile), rivers are important determinant of industrial location even for traditional manufacturing activities. Take the leather industry. Each step of the manufacturing process to obtain leather from the skins of the animals (tanning, retanning, and finishing) requires a considerable amount of fresh water (Sundar et al., 2001).

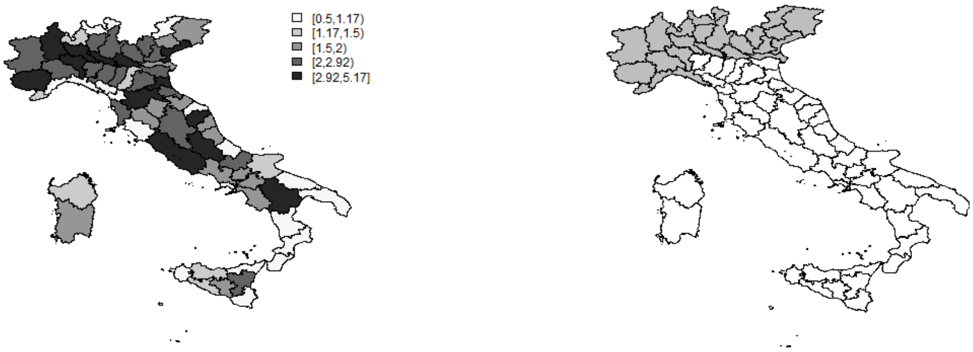
To measure water endowment at provincial level, the present study considers two variables.[4] The first one refers to the economic "relevance" of Italian rivers. The historical source *Annuario Statistico Italiano* (1886, pp. 22-27) provides an exhaustive list of rivers flowing through the 69 Italian provinces. These rivers are here ranked, with a value ranging from 1 to 5, according to the weight assigned to them by the experts of the Italian Military Geographic Institute (IGM) in a publication that culminates a research project started in the 1960s (IGM, 2007). The importance of each river is established on the base of "the length of each river and its socio-economic relevance" (IGM, 2007, preface). In the IGM ranking, the value of 1 means "high relevance", while the value of 5 means "low relevance". Thus, our first measure of water endowment in each province i is the weighted average of the number of rivers in the province using $(1 - IGMrank/6)$ as weights:

$$RIVER_i = \textstyle\sum_{r \in i}(1 - IGMrank_r/6),$$

where r refers to each river flowing through province i.

The second variable is a dummy indicating whether the province belongs to the Alpine region. Italian rivers are essentially of two types: Alpine rivers and Apennine rivers. Alpine rivers typically descend from the Alps, flow from the North into the upper bank of the Po river. The Alps work in a sense as a sponge that absorbs water in autumn and winter, and releases water in spring and summer, when the glaciers melt. For these reason, Alpine rivers are rich in water all year-long, while Apennine rivers are relatively dry during the summer season preventing the regular development of factory-based manufacturing activities not organized on a seasonal basis. Beside that, Alpine rivers also have higher flow rate than Apennines rivers. As we will document in our econometric analysis, the interaction between the continuous variable $RIVER_i$ and the Alpine region dummy variable proved to add significant help in our understanding of industrial location of certain manufacturing sectors, and supports nicely the need for industrial use of a year-round supply of water considered by Fenoaltea (2011). The provincial distribution of these water-related variables is illustrated in Figure 4.

[4]See the Appendix for details on sources and methods.



(a) Rivers (b) Alpine region

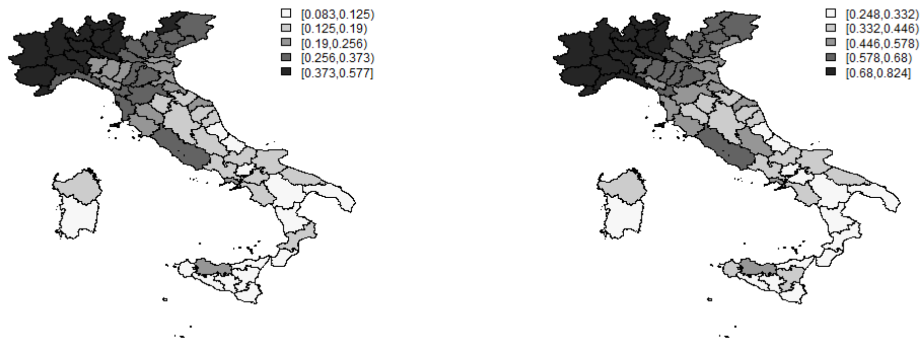
FIGURE 4: **Water endowment**: panel (a) shows the geographical distribution of the $RIVER_i$ variable, a measure of the "relevance" of the rivers flowing through province i; panel (b) shows the provinces belonging to the Alpine region.

3.1.2 Literacy

During the 19th century State support for education became gradually the rule in most European countries. Rulers "recognized the potential of education for promoting national integration, and schooling was used to create citizens loyal to the nation-state [...]. The economic connection between economic growth and formal education became stronger later in the nineteenth century" (Alter and Clark, 2010). Both political and economic motivations behind rising literacy rates during the nineteenth century fits also with the Italian case (Cipolla, 1969; Brian A'Hearn, 2015). On the political side, during the early 19th century, Northern regions were characterized by better governments than Southern ones and a primary public school system developed well before the political unification of the country in 1861. In the Italian Center and South an effective public primary school system only developed after the country's unification. On the economic side, the demand for education was also probably higher in the North than in the South. Quantitative studies referring to the 1830s suggest that in Northern regions a reduction in the gender gap in literacy rates started in those years. Literacy rates in the Center and especially in the South were very low and stagnant before 1861, as if, in addition to poor institutions, there was no particular internal economic need, or demand, for education. This situation changed only after many decades (see Vigo, 1971; Ciccarelli and Weisdorf, 2016, and references therein).

Figure 5 illustrates the geographical distribution of literacy rates in 1871 and 1911 (see Appendix for details on the data). The North-South regional gradient is evident. In 1871 the regions of north-west registered literacy rates of about 50 percent, those of center and north-east of about 30 percent, and southern regions of about 15 percent. In 1911, after four decades of mandatory primary school, literacy rates reached some 75 percent in the north-west, 55 in the north-east and in the center, and about 35 percent in the south. Territorial differences persisted. A recent study (Ciccarelli and Fachin (forthcoming)) estimate a conditional convergence model for manufacturing productivity growth in

Italian provinces during 1871-1911, and provides evidence of a positive contribution to growth of human capital. Interestingly, the authors find no evidence of spatial spillover effects.



(a) 1871

(b) 1911

FIGURE 5: **Literacy rates**

3.2 Market potential

In order to analyze the importance of domestic market access as a key driver of the spatial distribution of economic activity (Fujita et al., 1999; Combes et al., 2008), a sound measure of accessibility to demand is required. In line with Crafts (2005), we constructs market potential estimates for each Italian province i between 1871 and 1911 using Harris (1954)'s formula, that is as a weighted average of GDP (or total value added) of all provinces j[5]:

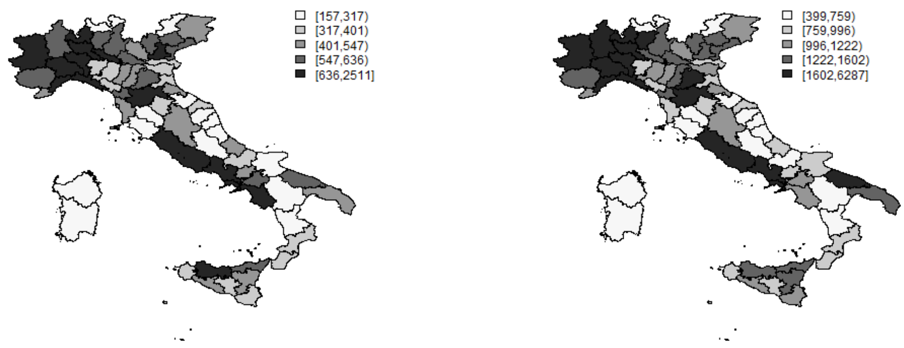
$$MKTPOT_{it} = \sum_{j=1}^{N} GDP_{jt} \times d_{ij}^{-\delta}, \qquad with \qquad \delta = 1 \tag{1}$$

with d_{ij} the great circle distance in km between the centroids of provinces i and j.[6]. In practice, this indicator equates the potential demand for goods and services produced

[5] Harris (1954)'s formula can be derived from a NEG theoretical model (Combes et al., 2008). An alternative way of obtaining a structural estimate of market potential based on NEG models was proposed by Redding and Venables (2004). The latter, however, requires data on bilateral trade flows that are not available at regional level for our sample period.

[6] In order to measure Harris' market potential, we rely on geodesic distances between locations. Recent historical studies (see, e.g. Martínez-Galarraga, 2014) consider more sophisticated measures of bilateral transport costs. These measures take into consideration different transport modes, the routes used in the transportation of commodities by mode, and the respective freight rates; they also take into account that their evolution over time can vary for a number of reasons. The adoption of this approach is beyond the scope of the present paper, since it requires a large amount of historical information not available using current data sources. Rather, it will be considered as an objective in our future research agenda. On the other hand, the assumption of a distance decay parameter δ equal to 1 may also appear as an arbitrary choice. Nevertheless, it should be stressed that the estimation of gravity equation has generated values

in a given location with that location's proximity to consumer markets. Thus, it can be interpreted as the volume of economic activity to which a region has access to after having taken into account the necessary transport costs to cover the distance to reach other provinces.[7] An important point is that historical GDP estimates at the provincial level for the case of Italy are not available. Thus, we proxy for provincial GDP by allocating total regional GDP (NUTS-2 units) estimates for 1871, 1881, 1901, and 1911 to provinces (NUTS-3 units) using the provincial shares of regional population obtained by population census.[8]



(a) Year: 1871 (b) Year: 1911

FIGURE 6: Harris market potential

Figure 6 show that the values of Harris market potential appear to be more and more concentrated in northwestern regions. However, one notes a a substantial amount of within-region heterogeneity providing a neat example of the advantage of using provincial (NUTS-3) over regional (NUTS-2) figures. In Lombardy, the undisputed economic leader among Italian regions, only the provinces of Milan (MI) and Como (CO) neighboring the region of Piedmont (in the very northwest of Italy and bordering France) score particularly well. Similarly, in Liguria only the province of Genoa registers high market potential while Porto Maurizio (PM) does not. But there is more than that. Estimated market potential is also high in Florence, Rome, Naples, and Palermo that is the provinces with the pre-unitarian city capitals of the Gran Duchy of Tuscany, the Papal States, and the Kingdom of Two Sicilies. This last evidence matches perfectly well with the intuition by (Fenoaltea, 2003, pp. 1073-74) that in his study on Italian regional industrialization noticed that in the early 1870s "The industrial, manufacturing regions are those with the former capitals, of the preceding decades and centuries" and that "In such a context the appropriate unit of analysis is not in fact the region, but (in Italy) the much smaller province".

of δ parameter that are often close to 1 (Disdier and Head, 2008; Head and Mayer, 2011). We have also checked the robustness of our econometric results to values of the δ parameter progressively increasing over time from 0.6 in 1871 to 1 in 1911, but relevant differences were not obtained.

[7] The trade cost to access the market internal to the province is assumed to be proportional to two third of the radius of a circle with an area equal to that of the province, i.e. $d_{ii} = 2/3\sqrt{Area_i/\pi}$.

[8] See Appendix for further details.

3.3 Expected sectoral effects of market potential, water endowment and literacy

Crafts and Mulatu (2006) show that patterns of industrial location in Britain were rather persistent and regional specialization changed relatively little during 1871-1911. Factor endowment (coal abundance in the case of Britain) had a stronger effect on overall industrial location than proximity to markets.[9] However, the latter was an attraction for industries with large plant size, above all shipbuilding and textile. This finding is important in that it supports the view that the the effect of market potential may depend on the degree of scale economies prevailing in the various sectors. Effectively, Economic geography theory predicts that economic activities with increasing returns to scale tend to establish themselves in regions that enjoy good market access, while the location of economic activities with constant returns technologies is mainly influenced by factor endowment. The effect of of factor endowment (especially of water in the Italian case) on industrial location may depend on the prevailing technology and thus, ultimately, on the industrial activity considered. Sectoral differences may also exist in the effect of regional skill endowment on industrial location. Standard neoclassical trade theory (Heckscher-Ohlin model) predicts that high-skilled labor intensive industries tend to be concentrated in regions with higher endowment of high-skilled labor. Crafts and Mulatu (2006) show that in 19th century Britain educated workers where an important input for the chemical sector, but not for textile.

Based on these economic geography arguments, we expect that the location of sectors with high and medium capital to labor ratio (K/L) was relatively more influenced by market potential, while the location of low K/L sectors was relatively more influenced by the water endowment. In addition, as far as water endowment is concerned, we expect that, as suggested by the economic history literature, the emergence of factory-based systems of production "gave a strong advantage to the locations with a year-round supply of water" (Fenoaltea, 2011). As we will illustrate shortly, the inclusion of the interaction terms between the continuous variable *river* and the dummy variable *Alpine* in our empirical model attempts to capture this important determinant of industrial location. We also expect, other things being equal, that the effect of literacy is relatively more important for the location of medium and high K/L sectors than for low K/L sectors.

Given the above conjectures, the capital to labor ratio should plays an important role in the analysis of industrial location. For this reason, even though our empirical results refer to each single manufacturing sector, we summarize the classification proposed by Italy's economic historians. Table 4 groups various manufacturing sectors into light and heavy industries depending on their capital intensity. It reports in particular data on horse-power per worker (HP/L), often used to proxy capital intensity (Rostas, 1948; Broadberry and Crafts, 1990; Broadberry and Fremdling, 1990; de Jong, 2003).

The table is mostly based on the data from the first industrial census of 1911 given in Zamagni (1978, 1993), and the 0.6 for 1911 threshold for "K/L" inferred from Federico (2006).[10] It also uses the disaggregated data for the engineering sector by Fenoaltea

[9] Fernihough and O'Rourke (2014) consider the importance of geographical proximity to coal as a factor underpinning comparative European economic development during the Industrial revolution.

[10] Federico (2006) distinguishes between into *light industries* and *heavy industries*. The former are

TABLE 4: Horse-power per worker (HP/L) in 1911

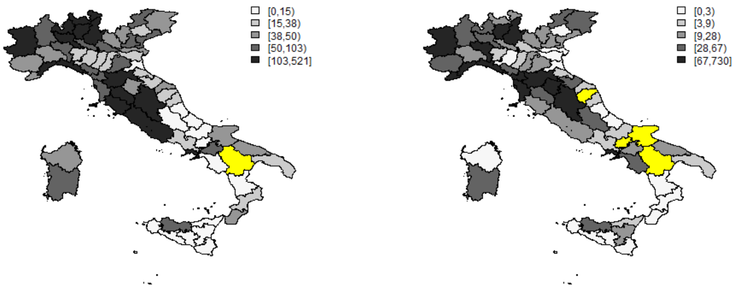
High HP/L	
Metalmaking	2.62
Chemicals	1.30
Foodstuffs[a]	0.94
Textile:	
cotton	0.85
wool	0.78
Paper	0.73
Engineering:	
machinery	0.61
shipbuilding	0.60
Low HP/L	
Non-metallic mineral products	0.36
Wood	0.23
Engineering:	
blacksmiths	0.20
Textile:	
silk	0.11
Leather	0.09
Clothing	0.07

Source: Zamagni (1978), and Fenoaltea (2014) for machinery, shipbuilding, and blacksmiths. The "K/L" figure for machinery is in particular the average of rail-guided vehicles, other heavy equipments, and other ordinary machinery as reported in Fenoaltea (2014), Table 2. [a] Foodstuffs is net of sugar (with K/L=2.18)

(2014). There is little argument in the literature that metalmaking, paper and chemicals represent high "K/L" sectors, while non-metallic mineral products, wood, leather, and clothing low "K/L" ones. Within the textile industry, cotton and wool belong to the high "K/L" group, while silk to the low one. The production of raw silk, an activity at the boundary between agriculture and manufacturing, is a very labour-intensive activity suitable for rich-water and densely populated area, such as the North-West of Italy. In addition, silk represented a leading component of Italian exports towards North-Western Europe, and being a luxury good, beyond the reach of most of 19th century Italians.((Federico and Tena-Junguito, 2014)). As far as the engineering industry is concerned, Fenoaltea (2014) shows that the average value of $K/L = 0.33$ for the whole industry, includes actually values as low as 0.20 for blacksmith and as high as 0.60 for shipyards, reflecting the fact that the manufacturing of major naval vessels (but also

those that “are featured by a (relatively) low capital intensity and by the prevailing orientation towards the final consumer”, the latter “are more capital-intensive and produce mainly inputs for other sectors”. The author (see in particular Table 2.6 therein), uses the industrial census of 1927 and consider as heavy industries those with a value of "K/L" above 0.7. Our data refer to 1911, and a 0.6 threshold for "K/L" dovetails better with the data at hand. Zamagni (1993) groups manufacturing industries into *advanced*, *intermediate*, and *traditional* ones, depending on their capital intensity and number of workers per plant steam locomotives) was far more sophisticated from a technological point of view than the maintenance of agricultural tools by blacksmiths.[11] A final note on the Foodstuff sector is that, as warned by Zamagni (1978), the relatively high level of horse-power per worker should not be misinterpreted, since it was largely due to the traditional flour-milling industry.

This section ends by illustrating the geographical distribution of manufacturing sectors. For reason of space, we consider metalmaking and leather as representative of, respectively "high" (metalmaking, machinery, shipbuilding, chemical, papers, cotton, wool) and "low" K/L and "low" (non-metallic mineral products, blacksmiths, leather, silk, wood, and clothing) K/L sectors. We consider instead separately the main sectoral components of the textile industry (silk, cotton, and wool).



(a) 1871 (b) 1911

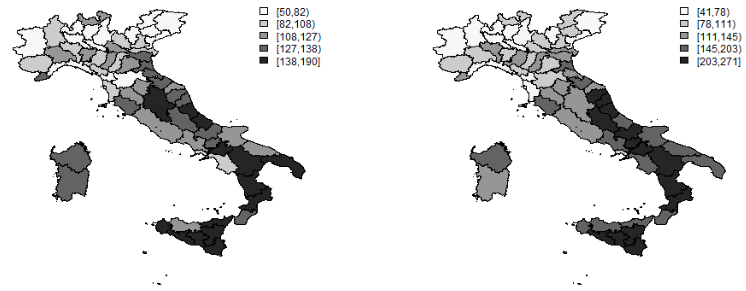
FIGURE 7: **Metalmaking (high K/L)**: Choropleth maps of LQ values

Figure 7 illustrates the case of "high" K/L sectors in 1871 and 1911. There is a clear predominance of the West side of Central and Northern Italy. Within northern regions, a marked difference between sub-Alpine provinces and those along the Po valley is evident. This shows well the advantage of using NUTS-3 instead of NUTS-2 data to analyze industrial location. At the same time, however, it is also evident some reallocation within the Center, with a shift away from the coastal Tyrrhenian regions (Latium and Southern Tuscany). In the South light colors prevail in 1871 and 1911, especially so along the backbone represented by the Central and Southern Apennines provinces down to Calabria, at the toe of Italy's boot.

Figure 8 considers "low" K/L sectors in 1871 and 1911. There is little dynamic here. In 1871, and even more in 1911, a clear North-South gradient or spatial trend is present. It is interesting to note that Figure 7 and Figure 8 essentially complement each other, suggesting possibly alternative driving forces for their industrial location.

Figures 9 and 10 refer finally to the geographical distribution of cotton and silk, components of the textile industry, and representing, together with metalmaking, the most agglomerated sectors in 1911 (see again Figure (3). The predominant role of Alpine

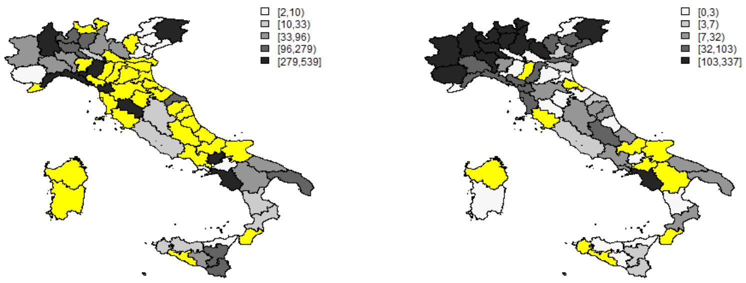
[11] The historical data and methodology used in Zamagni (1978) and Fenoaltea (2014) for the engineering sector differ, so that the average "K/L" value for the year 1911 proposed by Zamagni (1978)) ("K/L"=0.51) and by Fenoaltea (2014) ("K/L"=0.33) are difficult to compare.



(a) 1871 (b) 1911

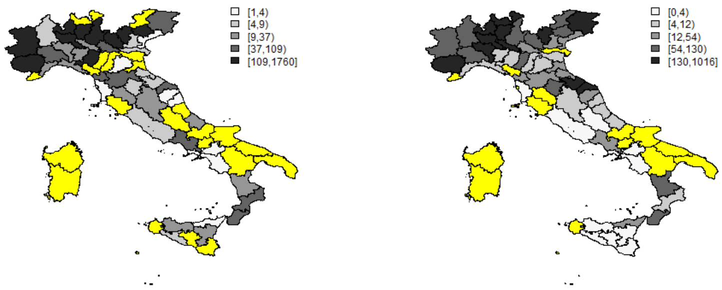
Figure 8: **Leather (low K/L)**: Choropleth maps of LQ values

regions emerges clearly, with Piedmont and Lombardy alone accounted for about 75% of value added in textile in 1913(Fenoaltea, 2004). However the sectoral reallocation between 1871 and 1911, with the whitening of the maps for the southern provinces of Campania and Sicily is also evident, especially for the cotton industry. These facts are largely consistent with the progressive mechanization of the cotton industry, also sustained by rising protection, with import duties increased during the late 1870s and 1880s (Fenoaltea, 2004). It is also interesting to note that silk producers were instead traditionally favourable to free trade (Cafagna, 1989; Fenoaltea, 2011).



(a) 1871 (b) 1911

Figure 9: **Cotton (high K/L)**: Choropleth maps of LQ values



(a) 1871 (b) 1911

FIGURE 10: **Silk (low K/L)**: Choropleth maps of LQ values

4 Econometric analysis

This section examines the relevance of factor endowment and (domestic) market potential in shaping the location of industries across Italian provinces during the period 1871-1911. Industrial location is measured in relative terms, that is using the log of the location quotient, $\ln(LQ)$. Market potential is measured by the log of Harris (1954) formula ($\ln(Mktpot)$) (see section 3.2). As for factor endowment, we focus on labor skills and water abundance. Labor skills are measured by the log of literacy, $\ln(Literacy)$, i.e. the share of literate population in the province's total population. Water supply is measured by the continuous, but time-invariant, variable $\ln(River)$ described in section 3.1, and its interaction with the dummy variable *Alpine*, indicating whether the province belongs to the Alpine region.

Following Combes and Gobillon (2015), we test the effect of market potential, literacy, and natural advantages in two steps. In the first step, we exploit the panel structure of the data (69 provinces for 4 time periods) to assess the effect of time-varying variables (i.e. $\ln(Mktpot)$ and $\ln(Literacy)$), while in the second step we estimate the effect of time-invariant variables (i.e. water abundance).

4.1 The effect of market potential and literacy

4.1.1 Empirical strategy

In this section we discuss the estimated effects of the two time-varying variables (i.e. $\ln(Mktpot)$ and $\ln(Literacy)$). Differently from Combes and Gobillon (2015), however, in this first step we control for time-invariant unobserved heterogeneity by using a simple semiparametric model with a smooth spatial trend (the so-called *Geoadditive Model*) (Lee and Durbán, 2011), rather than by introducing spatial fixed effects. More formally, for each sector k, each province i, and each time period t, the model for the first step is specified as:[12]

$$\begin{aligned}
\ln\left(LQ_{i,t}^{k}\right) &= \alpha+\beta_1\ln\left(Literacy_{i,t}\right)+\beta_2\ln\left(Mktpot_{i,t}\right) \\
&\quad +f_1\left(Lat_i\right)+f_2\left(Long_i\right)+f_{12}\left(Lat_i,Long_i\right)+\gamma_t+\epsilon_{i,t} \\
\varepsilon_{i,t} &\sim iid\mathcal{N}\left(0,\sigma_{\varepsilon}^{2}\right) \qquad i=1,...,N \qquad t=1,...,T
\end{aligned} \tag{2}$$

Time fixed effects (γ_t) are introduced in the model to control for time-related factor biases. Moreover, the geoadditive terms, i.e. the smooth effect of the latitude - $f_1(Lat_i)$, of the longitude - $f_2(Long_i)$, and of their interaction - $f_{12}(Lat_i, Long_i)$ - work as control functions to filter the spatial trend out of the residuals, and transfer it to the mean response in a model specification. Thus, they allow to capture the shape of the spatial distribution of the dependent variable, conditional on the determinants included in the model. These control functions also isolate stochastic spatial dependence in the residuals, that is spatially autocorrelated unobserved heterogeneity (see also McMillen, 2012; Basile et al., 2014). Thus, they can be regarded as an alternative to individual regional dummies (spatial fixed effects) to capture unobserved spatial heterogeneity as long as the latter is smoothly distributed over space. Regional dummies peak significantly higher and lower levels of the mean response variable. If these peaks are smoothly distributed over a two-dimensional surface (i.e. if unobserved spatial heterogeneity is spatially auto-correlated), the smooth spatial trend is able to capture them.[13] We simply demonstrate the validity of these statements by estimating the two competing models without explanatory variables (see Appendix 1).

A complication with the estimation of model (2) is given by the presence of endogenous variables - $\ln(Mktpot_{i,t})$ and $\ln(Literacy_{i,t})$ - on the right hand side. As for $\ln(Mktpot_{i,t})$, New Economic Geography models describe a process characterized by reverse causality in which market potential, by attracting firms and workers, increases production in a particular location, and this, in turn, raises its market potential. $\ln(Literacy_{i,t})$ may also be an endogenous variable. On the one hand, the availability of literate workers may foster the concentration of industrial activities in certain regions. On the other hand, however, more industrialized regions may provide an incentive to achieve education that is generally lacking in backward areas of the country.

To address these issues, we extend the REML methodology to estimate the parameters of model (2) in a 2-step "control function" (CF) approach (Blundell and Powell, 2003; Wooldridge, 2010), that is an alternative to standard instrumental variable/two-stage-least-square (IV-2SLS) methods. In the first step, each endogenous variable is regressed

[12] This specification is different from the one used in the related literature (for example by Wolf, 2007; Ellison and Glaeser, 1999; Midelfart-Knarvik et al., 2000a,b). These authors pool the data by regions, sectors and time, and regress the location quotient on a set of interactions between the vectors of location characteristics (factors endowment and market potential) and a vector of industry characteristics (measuring industries' factor intensities and the share of intermediate inputs in GDP). Our dataset is reach enough to avoid considering the above interaction and pooling of the data, and it allows us to estimate separate models for "homogeneous" groups of industrial activity.

[13] The smooth components of the spatial trend in equation 2 - f_1, f_2 andf_{12} - are approximated by tensor product smoothers (Wood, 2006; Basile et al., 2014). As is well known, any semiparametric model can be expressed as a mixed model and, thus, it is possible to estimate all the parameters using restricted maximum likelihood methods (REML) (Ruppert et al., 2003). To estimate this model, we used the method described by Wood (2006) which allows for automatic and integrated smoothing parameters selection through the minimization of the REML. Wood has implemented this approach in the R package *mgcv*.

on a set of conformable instrumental variables (Z), using a semiparametric model. The residuals from the first steps are then included in the original model (2) to control for the endogeneity of $\ln Mktpot_{i,t}$ and $\ln Literacy_{i,t}$. Since the second step contains generated regressors (i.e. the first-step residuals), a bootstrap procedure is used to compute p-values (see Basile et al., 2014, for details on the bootstrap procedure).

This procedure requires finding good instruments, i.e. variables that are correlated with the endogenous explanatory variables but not with the residuals of the regression. To control for the endogeneity of market potential, we follow the main empirical literature in using a measure of centrality of the region ($Centrality = \sum_i d_{ij}^{-1}$) (Head and Mayer, 2006), and the geographical distance from the main economic center (i.e. the distance from Milan, $DistMilan_i$) (Redding and Venables, 2004; Klein and Crafts, 2012b; Martinez-Galarraga, 2012b; Wolf, 2007) as instrumental variables. To control for the endogeneity of $\ln Literacy_{i,t}$, we use its time lag ($\ln Literacy_{i,t-10}$).[14]

4.1.2 Evidence for the whole manufacturing sector

For the case of the whole manufacturing activity, we report in Table 5 the estimation results of the semiparametric CF approach, along with the estimation results of a fully parametric 2SLS. Obviously, we cannot use the within-group version of the 2SLS estimator, since two important instruments ($DistMilan_i$ and $Centrality_i$) are time invariant, while the third one ($\ln Literacy_{i,t-10}$) would be correlated with the within-group transformed error term. Thus, in order to control for spatial heterogeneity, we included a parametric nonlinear spatial trend (i.e. Lat, Lat^2, $Long$, $Long^2$, $Lat \times Long$) on the r.h.s. of the pooled 2SLS model.

All in all, the diagnostic tests of the 2SLS model provide encouraging evidence in favor of the chosen set of instruments. First, the Wu-Hausman test confirms that $\ln Literacy_{i,t}$ and $\ln Mktpot_{i,t}$ are endogenous. Second, the weak instrument tests confirm that the instrumental variables are strongly correlated to the endogenous variables. Third, the Sargan test of overidentifying restrictions suggests that the excluded instruments ($DistMilan_i$, $Centrality_i$, and $\ln Literacy_{i,t-10}$) are valid instruments, i.e. they are uncorrelated with the error term, and thus they are correctly excluded from the estimated equation. Nevertheless, the estimated parametric 2SLS model does not properly control for unobserved spatial heterogeneity. From Table 5 it emerges indeed that the spatial variables (Lat, Lat^2, $Long$, $Long^2$, $Lat \times Long$) are weakly significant.

The spatial trend in the data (i.e. the smooth spatial heterogeneity) is much better captured by the semiparametric geoadditive model as indicated by the highly significance of the smooth terms f_1, f_2 and f_{12} (in column Semiparametric CF). [15] The smooth functions of the residuals from the two first steps of the CF approach ($h_1(Res_1)$ and $h_2(Res_2)$) work as control functions to correct the estimated parameters for the endogeneity bias. The statistical significance of the these smooth terms can also be used as endogeneity tests. In Table 5 we also report a test of the joint significance of the $h_1(Res_1)$ and $h_2(Res_2)$, similar to the Wu-Hausman test in the parametric 2SLS approach, as well as

[14] See Appendix for detail on literacy data.

[15] For the nonparametric smooth terms (i.e. the spatial trend and the first step residuals), Table 5 also shows the estimated degree of freedom (*edf*), a broad measure of nonlinearity (an *edf* equal to 1 indicates linearity, while a value higher than 1 indicates nonlinearity).

Variable	2SLS	Semiparametric CF
Parametric terms		
Intercept	5.003	-1.833***
	(0.613)	(0.001)
$\ln(Mktpot)$	0.610***	0.468***
	(0.000)	(0.000)
$\ln(Literacy)$	0.015	0.235**
	(0.934)	(0.027)
Lat	-0.487	
	(0.238)	
Lat^2	0.006	
	(0.153)	
$Long$	0.268	
	(0.456)	
$Long^2$	-0.009**	
	(0.027)	
$Lat \times Long$	-0.001	
	(0.940)	
Non-parametric terms		
$f_1(Lat)$		7.760***
		(0.000)
$f_2(Long)$		10.259***
		(0.000)
$f_{12}(Lat, Long)$		20.047***
		(0.000)
$h_1(Res_1)$		3.007***
		(0.078)
$h_2(Res_2)$		2.933***
		(0.002)
Diagnostics		
Weak instr.-Ln(Mktpot)	21.208***	24.461*
	(0.000)	(0.000)
Weak instr.-Ln(Literacy)	405.489***	210.072***
	(0.000)	(0.000)
Wu-Hausman	10.535***	33.056***
	(0.000)	(0.000)
Sargan	0.436	3.727
	(0.509)	(0.155)

TABLE 5: Whole manufacturing. Estimation results of the parametric instrumental variable (2SLS) approach and of the semiparametric control function (CF) approach. Coefficients, Edf and bootstrap p-values (in parenthesis). Time fixed effects are included in both models. Number of observations: 276.

two weak-instrument tests using the results of the semiparametric first steps (they confirm that the excluded instruments are strongly correlated to the endogenous variables, thus rejecting the hypothesis of weak instruments). Unfortunately, there is not a well-known and widely accepted test for the validity of the conditional mean restrictions imposed by the CF approach (overidentification test). A practically feasible way of testing such restrictions consists of including the excluded instruments in the CF estimate, and test their significance. They should not be significant since the CF should pick up all of the correlation between the structural error term and (X, Z). In our case, all the external instruments turned out to be strictly exogenous.

The CF estimation results for whole *manufacturing* show that, after controlling for unobserved heterogeneity and after correcting for the endogeneity bias, the variables $\ln(Literacy)$ and $\ln(Mktpot)$ enter the model significantly and positively, thus proving to be important drivers of industrial location during the sample period. In particular, the positive effect of market potential corroborates the hypothesis that even during the early stage of Italy's industrialization firms tended to settle in regions with the highest market potential to minimize costs. And, as discussed in Section 2.1, the North West was the area with increasing market potential during 1871-1911, as if the country's unification improved its relative position within the domestic market. Moreover, the elasticity of $Mktpot$ (0.47) is much higher than that of $Literacy$ (0.24).

4.1.3 The heterogenous effect of market potential across sectors

Table 6 looks within manufacturing and provides the estimation results of (the second step of) the model (2) for each industrial sector. Again, the model includes a smooth spatial trend to control for unobserved spatial heterogeneity, time dummies to control for unobserved time heterogeneity, and the smooth terms of the first-step residuals as control functions to correct the inconsistency due to the endogeneity of $\ln(Mktpot)$ and $\ln(Literacy)$.[16] The results were obtained by pooling the data over the 69 provinces and the four census years (1871, 1881, 1901, and 1911). For each parametric term, we report the estimated coefficient and the corresponding bootstrapped p-value.

In line with our expectations, during the period 1871-1911 market potential turned out to be a key driver of the industrial location in the case of *high* HP/L industries. The coefficient associated to the variable $\ln(Mktpot)$ is indeed positive and strongly significant in the case of Metalmaking, Chemicals, Paper, Machinery, Cotton, Wool, and Other natural fibers. Particularly high the elasticity of this variable in the case of Cotton (2.8) and Metalmaking (1.5). Shipbuilding is the only scale intensive sector with a non-significant effect of market potential. But in this case, the industrial location was obviously driven by the presence of a port, independently of the market potential of the region. The effect of $\ln(Mktpot)$ was instead negative (or not significant) in the case of *low* HP/L industries (Blacksmith, Foodstuffs, Leather, and Wood). Silk represent an isolated, yet important, exception. It was a labour intensive sector for which the estimated effect of market potential is sizeable (2.2), and highly significant. Recall however that silk production was increasingly concentrated in norther provinces and largely exported to Norther European countries. It is thus likely that that proximity to North Europe and reducing international transport cost associated with railway development, more than domestic market potential, where at the hearth of the relocation to Northern Italian provinces of the silk activity.[17]

These findings are consistent with the predictions of the Core-Periphery NEG model (Krugman, 1991), according to which economic activities with increasing returns to scale tend to establish themselves in regions that enjoy good market access. A region with a larger market potential is also characterized by a more generous compensation of local

[16] The results for these controls are not reported in Table **??**, but are available upon request.

[17] Up to the late 1880s, when Italy started a nonsense trade-war with France, nearly a half of Italy's total exports, albeit rather limited in size, were directed to the France market (Federico et al., 2011, pp. 42-43).

Sectors	Market potential	Literacy
Whole manufacturing	0.468***	0.235**
	(0.000)	(0.027)
High HP/L sectors		
Metalmaking	1.502***	1.976***
	(0.000)	(0.002)
Chemicals	0.653***	0.670**
	(0.000)	(0.044)
Paper	0.969***	0.320
	(0.000)	(0.217)
Shipbuilding	0.475	2.659**
	(0.469)	(0.015)
Machinery	0.687***	0.105
	(0.000)	(0.629)
Cotton	2.815***	-0.244
	(0.000)	(0.817)
Wool	0.775*	-1.475*
	(0.073)	(0.058)
Other nat.fibers	0.614**	0.220
	(0.018)	(0.701)
Low HP/L sectors		
Silk	2.199***	-1.035
	(0.000)	(0.321)
Blacksmith	-0.222***	-0.184*
	(0.000)	(0.054)
Foodstuffs	-0.286***	0.272***
	(0.000)	(0.002)
Clothing	0.017	0.126
	(0.876)	(0.406)
Leather	-0.453***	-0.005
	(0.000)	(0.963)
Wood	-0.191***	0.199**
	(0.001)	(0.019)
Non-met.min.prod.	0.023	-0.770***
	(0.862)	(0.001)

TABLE 6: Marginal effects of Ln(Mktpot) and Ln(Literacy). Coefficients and bootstrap p-values (in parenthesis). The number of observations for each sector is 276 (69 province by 4 time points.)

factors, while small regions which are far from the large market will have lower local wages compared to regions close to the industrial core. Thus, if everything else is unchanged, and if the firms all achieve constant returns to scale (as in the case of *low* HP/L industries), the increase in the price of local factors (labor, land and so on) will reduce the profitability of all the firms at that location.

Wihtin *high HP/L* sectors, Shipbuidings, Metalmaking, and Chemicals proved to be knowledge-intensive sectors requiring educated labor force, mainly located in the north-western regions. While for most traditional *low HP/L* sectors, often tied to agriculture, literacy has a negative impact or is not significant. With Foodstuffs and Wood representing the only exceptions.

4.2 The effect of water endowment

4.2.1 Empirical strategy and evidence for the whole manufacturing sector

Similarly to Combes and Gobillon (2015), in the second step of our empirical strategy we assess the effect of time-invariant variables (i.e. *Alpine*, *lnRiver*, and their interaction). However, instead of regressing the estimated fixed effects, $\widehat{\alpha_i}$, on $Alpine_i$, $\ln(River_i)$ and $Alpine_i \times \ln(River_i)$, we regress the estimated values of the spatial trend $(\widehat{f_{spt_i}})$ on the three variables. That is, we estimate the following interaction model using OLS:

$$
\begin{aligned}
\widehat{f_{spt_i}} &= \alpha + \beta_1 Alpine_i + \beta_2 \ln(River_i) + \\
&\quad +\beta_3 Alpine_i \times \ln(River_i) + \epsilon_i \\
\varepsilon_i &\sim iid\mathcal{N}\left(0, \sigma_\varepsilon^2\right) \qquad i = 1, ..., N \\
\widehat{f_{spt_i}} &= \widehat{f_1}(Lat_i) + \widehat{f_2}(Long_i) + \widehat{f_{12}}(Lat_i, Long_i)
\end{aligned}
\tag{3}
$$

As expected, however, the OLS residuals turned out to be spatially autocorrelated. For the case of the Whole manufacturing sector, the standardized Moran I test statistic on the residuals is equal to 4.539 and its pvalue is 0.000. In fact, $Alpine_i$, $\ln(River_i)$, and $Alpine_i \times \ln(River_i)$ capture only a portion of the large spatial heterogeneity which characterizes the spatial distribution of the location quotient net of the effect of market potential and literacy, i.e. the spatial distribution of $\widehat{f_{spt_i}}$ (the adjusted R^2 of the model is 0.31). Heterogeneity among provinces, induced by an uneven distribution of immobile resources (such as natural harbors) and amenities (climate) may also be at the origin of a variety of comparative advantages. This unobserved heterogeneity may also be spatially correlated, thus introducing spatial autocorrelation in the residuals.

In order to control for spatial autocorrelation, we use a semiparametric spatial filter (Tiefelsdorf and Griffith, 2007). As well known, spatial filtering uses a set of spatial proxy variables, which are extracted as eigenvectors from the spatial weights matrix, and implants these vectors as control variables into the model. These control variables identify and isolate the stochastic spatial dependencies among observations, thus allowing model building to proceed as if the observations were independent. Specifically, we used the function `SpatialFiltering` from the `R` library `spdep`. This function selects eigenvectors in a semiparametric spatial filtering approach (as proposed by Tiefelsdorf and Griffith, 2007) to removing spatial dependence from linear models. The optimal subset of eigenvectors is identified by an objective function that minimizes spatial autocorrelation, i.e. by finding the single eigenvector which reduces the standard variate of Moran's I for regression residuals most, and continuing until no candidate eigenvector reduces the value by more than a tolerance value. This subset of eigenvectors is a proxy either for those spatially autocorrelated exogenous factors that have not been incorporated into a model, or for an underlying spatial process that ties the observations together. Furthermore, incorporation of all relevant eigenvectors into a model should leave the remaining residual component spatially uncorrelated. Consequently, standard statistical modeling and estimation techniques as well as interpretations can be employed for spatially filtered models.

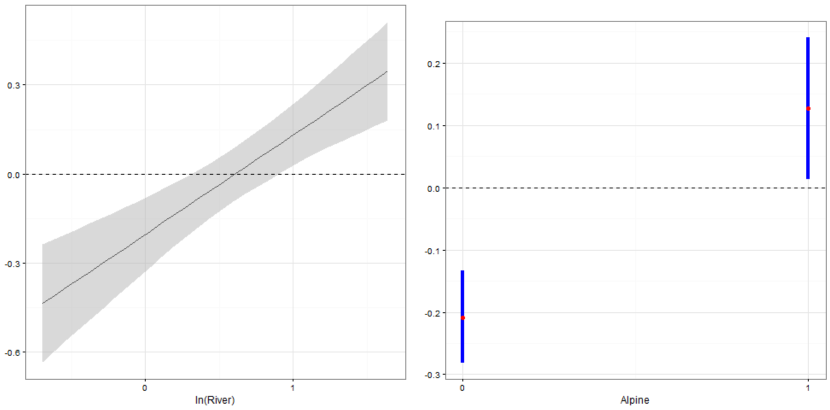
The spatially filtered OLS estimates of model (3) confirm the widely accepted idea that, over the analyzed period, the whole manufacturing activity was mainly located in proximity to Alpine rivers. Figure 11 displays in panel (a) the changes in the marginal effect of *Alpine* conditional on the values of $\ln River$, that is

$$\frac{\partial \widehat{f_{spt}}}{\partial Alpine} = \beta_1 + \beta_3 \ln River$$

while it shows in panel (b) the changes in the marginal effect of the continuous variable $\ln River$ conditional on the values of the dummy variable *Alpine*, that is

$$\frac{\partial \widehat{f_{spt}}}{\partial \ln River} = \beta_2 + \beta_3 Alpine$$

The two plots, created using the `R` function `interplot`, also include simulated 95% pointwise confidence intervals (obtained using the simulation function from the **R** package **arm** of Gelman and Hill, 2006)) around these marginal effects (see also Brambor et al., 2006, for a methodological discussion on the measurement of the marginal effects of interaction terms).[18]



(a) Estimated coefficient of Alpine by ln(River)

(b) Estimated coefficient of ln(River) by Alpine

FIGURE 11: Whole manufacturing. Marginal effect of ln(River) and Alpine with simulated 95% confidence intervals.

The plot in panel (a) shows that, with increasing importance of the river (along the horizontal axis), the magnitude of the marginal effect of Alpine on the location of the whole manufacturing sector also increases (along the vertical axis). The "dot-and-whisker" plot in panel (b) shows the effect of $\ln River$ (measuring the importance of the

[18] Let us consider, for example, the computation of the simulated confidence intervals for the marginal effect of $\ln(River)$ on $\widehat{f_{spt}}$. Let $\widehat{ME}_x^0 = \widehat{\beta_2}$ be shorthand notation for $\partial\widehat{f_{spt}}/\partial \ln River$ when $Alpine = 0$, and $\widehat{ME}_x^1 = \widehat{\beta_2} + \widehat{\beta_3}$ be shorthand notation for $\partial\widehat{f_{spt}}/\partial \ln River$ when $Alpine = 1$. The statistical significance of these marginal effects can be assessed in two different ways. First, we may use the appropriate analytic formula to calculate the variance of $\widehat{ME}_x^0$ and $\widehat{ME}_x^1$ using the variance-covariance matrix of the estimated regression: $var(\widehat{ME}_x^0) = var(\widehat{\beta_2})$ and $var(\widehat{ME}_x^1) = var(\widehat{\beta_2}) + var(\widehat{\beta_3}) + 2cov(\widehat{\beta_2}, \widehat{\beta_3})$. This enables us to calculate a pointwise 95% confidence interval using the critical t-statistic for a two-tailed $\alpha = 0.05$ test in the usual way. Second, we can simulate 5000 draws out of the asymptotic (multivariate normal) distribution of β_s for the regression, calculate $\widehat{ME}_x^0$ and $\widehat{ME}_x^1$ for each draw, and select the 2.5th and 97.5th percentiles of those calculations to form a 95% confidence interval.

river) conditional on the dummy variable *Alpine*. On the one hand, when $Alpine = 1$ (that is when the province belongs to the Alpine region),[19] the simulated 95% confidence interval around the marginal effect of $\ln River$ is above (and does not contain) the horizontal zero line. On the other hand, when $Alpine = 0$ (that is when the province belongs to the Apennine region), the simulated 95% confidence interval around the marginal effect of $\ln River$ is below (and does not contain) the horizontal zero line. Thus, for the manufacturing as a whole, it emerges a positive effect of Alpine rivers, as widely suggested by previous qualitative studies of historical nature stressing the importance of the water as a source of motive power and, more generally, to sustain manufacturing activities that required water throughout the year.

4.2.2 The heterogenous effect of water endowment across sectors

Table 7 shows that, once market potential and literacy are accounted for, the marginal effect of $\ln River$ when $Alpine = 1$ is indeed positive and statistically significant for six out fifteen industries. Consistently with our hypotheses, they are mostly *low HP/L industries* (Foodstuffs, Clothing, Silk, and Leather), even though the set includes two *high HP/L sector* (Chemicals, and Other natural fibers). In the case of Chemical products we find in particular the same result holding for the whole manufacturing (i.e. a positive marginal effect of Alpine rivers and a negative effect of Apennine rivers). The estimated effect is instead negative or insignificant for the remaining industries. Apennine rivers have instead a positive effect in the cases of Wool, Silk, Blacksmith and Leather.

Sectors	Alpine = 0	Lower bound	Upper bound	Alpine = 1	Lower bound	Upper bound
Manufacturing	-0.208*	-0.281	-0.133	0.127*	0.013	0.241
High HP/L sectors						
Metalmaking	-0.294*	-0.532	-0.059	-0.283	-0.632	0.074
Chemicals	-0.198*	-0.333	-0.064	0.430*	0.221	0.646
Paper	-0.095	-0.215	0.025	-0.225*	-0.403	-0.052
Shipbuilding	-1.460*	-2.254	-0.661	-0.557	-1.730	0.643
Machinery	-0.040	-0.098	0.015	0.031	-0.054	0.114
Cotton	-0.200	-0.459	0.049	-1.023*	-1.389	-0.656
Wool	0.712*	0.256	1.162	-0.002	-0.679	0.681
Other natural fibers	0.029	-0.130	0.189	0.528*	0.292	0.767
Low HP/L sectors						
Blackstmith	0.089*	0.046	0.132	-0.033	-0.101	0.035
Foodstuffs	0.055	-0.018	0.130	0.150*	0.041	0.257
Clothing	0.078	-0.036	0.188	0.189*	0.018	0.361
Leather	0.075*	0.009	0.141	0.156*	0.062	0.250
Wood	-0.026	-0.091	0.039	0.051	-0.051	0.152
Silk	0.492*	0.016	0.961	0.731*	0.034	1.452
Non-met.min.prod.	0.045	-0.074	0.164	-0.397*	-0.589	-0.204

TABLE 7: Marginal effect of ln(River) conditional on Alpine. Coefficients and bootstrap confidence intervals. The number of observations for each sector is 69.

Overall, the results shed a new light on the interpretation of the much debated role of

[19] Essentially, the Alpine region include the provinces on the left bank (that is at the north) of the Po river.

water endowment on industrial location in 19th century Italy. The historical literature stressed the comparative advantage of northern regions over southern ones stemming from water endowment during the early development of the Italian manufacturing industry. Effectively, even when the effect of market potential and literacy is duly accounted for, there is still room left for water endowment as a driver of industrial location. However this role appear to be inherently tied to the nature of the industrial sector considered, and cannot be simply attributed to a generic North-South divide. Consider the case of Silk and Leather industries. The descriptive analysis showed that Silk was increasingly agglomerated in the Alpine regions of Northern Italy, while Leather spatially diffused in the Italian South. Both industries made an intensive use of fresh water, not just as a source of motive power, but also and more generally in the various steps of their production process. For both industries, the empirical results of Table 7 show effectively that the proximity to a river was relevant independently of its position (either Alpine or Apennine).

The set of empirical results also allow us to analyse the net effect of Alpine regions on industrial location as a function of river. These are discussed in the Appendix.

5 Conclusions

Using a new set of provincial (NUTS-3 units) data on industrial value added at 1911 prices, this paper analyzed the spatial location patterns of manufacturing activity in Italy during the period 1871-1911. Specifically, we tested the relative effect of tangible factors, such as market size and factor endowment (water abundance and literacy), as the main drivers of industrial location, ruling out the role of intangible factors such as knowledge spillovers. The results show that, after the political and economic unification of the country of 1861, Italian provinces became more and more specialized, and manufacturing activity became increasingly concentrated in a few provinces, mostly belonging to the north-west part of the country. The estimation results corroborate the hypothesis that both comparative advantages (water endowment effect) and market potential (home-market effect) have been responsible of this process of spatial concentration.

The "K/L" ratio provided a useful guidance in the analysis of sectoral developments. The role of market potential as a driver of industrial location emerged clearly for high "K/L" sectors mostly tied to the production of durables (Metalmaking, Chemicals, Machinery, Paper), but also of consumption goods (Cotton, and Wool). In addition, the location of Metalmaking and Chemicals was positively related to the availability of more educated labor force. Furthermore, once the effect of market potential and literacy is accounted for, we find evidence that the location of some traditional industries characterized by a low capital-labor ratio (such as Silk and Leather) was mainly driven by water endowment. There are of course more industrial location developments in heaven and earth than in our theoretical and empirical models. For instance, the case of shibuilding, where only literacy emerged as important driver of industrial location, doesn't fit well within our interpretive scheme. Wit these important caveats in mind, our results show the importance of using local (regional or provincial) data that are further disaggregated at the sectoral level when investigating the main factors influencing industrial location.

Our findings are quite in line with the Core-Periphery NEG model (Krugman, 1991) that predicts increasing polarization and regional specialization as a result of economic integration. In this model, agglomeration economies derive from the interaction among economies of scale, transportation costs, and market size, while intangible external economies (such as information spillovers) do not play any role. Today NEG models are strongly criticized on the base of the observation that they focus on forces and processes that were important a century ago but are much less relevant today. The NEG approach seems less and less applicable to the actual location patterns of advanced economies. Nevertheless, the current economic geography of fast-growing countries like Brazil, China, and India is highly reminiscent of the economic geography of Western World countries during the nineteenth century, and it fits well into the NEG framework. Historical analyses like the one presented in this paper might thus help understanding the current process of economic modernization of developing countries.

Appendix

A1 Map of Italy's provinces at 1911 borders

The name of each region (roughly NUTS 2 units), in bold, is followed by the name (and tag) of its provinces (roughly NUTS 3 units). Different shadows of gray simply highlight regional borders. There were 16 regions and 69 provinces. The list follows a broad North to South ordering.
PIEDMONT: Alessandria (AL), Cuneo (CN), Novara (NO), Turin (TO)
LIGURIA: Genoa (GE), Porto Maurizio (PM)
LOMBARDY: Bergamo (BG), Brescia (BS), Como (CO), Cremona (CR), Mantua (MN), Milan (MI), Pavia (PV), Sondrio (SO)
VENETIA: Belluno (BL), Padua (PD), Rovigo (RO), Treviso (TV),Udine (UD), Venice (VE), Verona (VR), Vicenza (VI)
EMILIA: Bologna (BO), Ferrara (FE), Forlì (FO), Modena (MO), Parma (PR), Piacenza (PC), Ravenna(RA), Reggio Emilia (RE)
TUSCANY: Arezzo (AR), Florence (FI), Grosseto (GR), Leghorn (LI), Lucca (LU), Massa Carrara (MS), Pisa (PI), Siena (SI)
MARCHES: Ancona (AN), Ascoli Piceno (AP), Macerata (MC), Pesaro (PE)
UMBRIA: Perugia (PG)
LATIUM: Roma (RM)

ABRUZZI: Aquila (AQ), Campobasso (CB), Chieti (CH), Teramo (TE)
CAMPANIA: Avellino (AV), Benevento (BN), Caserta (CE), Naples (NA), Salerno (SA)
APULIA: Bari (BA), Foggia (FG), Lecce (LE)
BASILICATA: Potenza (PZ)
CALABRIA: Catanzaro (CZ), Cosenza (CS), Reggio Calabria (RC)

SICILY: Caltanissetta (CL), Catania (CT), Girgenti (AG), Messina (ME), Palermo (PA), Syracuse (SR), Trapani (TP)
SARDINIA: Cagliari (CA), Sassari (SS)

A2 Data

A2.1 Value added for manufacturing sectors

In this paper the location quotient is defined in terms of provincial value added data at 1911 prices. Provincial value added data for the following manufacturing sectors: 2.1 foodstuffs, 2.2 tobacco, 2.4 clothing, 2.5 wood, 2.6 leather, 2.7 metalmaking, 2.9 non-metallic mineral products, 2.10 chemicals and rubber, 2.11 paper, and 2.12 sundry are from Ciccarelli and Fenoaltea (2013). The data are available at `http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0289.2011.00643.x/full` (see in particular the supplementary material). Ciccarelli and Fenoaltea obtained the provincial figures as follows. They first produced annual 1861-1913 regional value added data at 1911 prices (or estimates) as a result of a long term project sponsored by the bank of Italy. The regional figures, disaggregated by industrial sector, are available at `https://www.bancaditalia.it/pubblicazioni/altre-pubblicazioni-storiche/produzione-industriale-1861-1913/`. Once the regional value added data (NUTS-2 units) were obtained, the authors allocated them to provinces (NUTS-3 units), by using the provincial labor force shares ($LFSHPROV$) of regional totals, separately for each of industrial sector. For each given region and time, provincial value added in province i and sector j has been obtained as:

$$VAPROV_{i,j} = VAREGIO_j * LFSHPROV_{i,j}$$

As a consequence, while annual 1861-1913 regional estimates ($VAREGIO$) are based on a rich set of detailed historical sources, provincial estimates ($VAPROV$) rest essentially on the information on labor force as reported in the population censuses and are thus only available for the years 1871, 1881, 1901, and 1911. (See also Ciccarelli and Fachin (forthcoming) for further details.)

Following this approach in this paper we used the population censuses of 1871, 1881, 1901, and 1911 to further disaggregated at the the provincial level the regional value added data for the 2.3 textile and 2.8 engineering industries. Detailed regional value added data for the textile industry and its sectoral component (including wool, silk, cotton, and other natural fibers) are those contained in Fenoaltea (2004). We obtained the provincial data for the sectoral component of the engineering industry in exactly the same way. The regional value added data in this case are those provided in Ciccarelli and Fenoaltea (2014a). We in particular disaggregated the textile industry into 2.31 cotton, 2.32 wool, 2.33 silk, 2.34 other natural fibers, and the engineering industry into 2.81 shipbuilding, 2.82 machinery, and 2.83 blacksmith. While a finer disaggregation would have been in principle possible, census data revealed that the it was particularly important to separate out the blacksmith component of the engineering industry (Fenoaltea (2014), Fenoaltea (2016)).

It is important to notice that population censuses do not generally report information on industrial sectors, rather on individual professions. These latter, as it is customary, must be mapped into industrial sector, what always leaves a certain degree of arbitrariness. The details are too many to be even partially illustrated here. However, just to give the basic idea, we allocated, province by province, and census year by census year, to the cotton sectors professions such as cotton spinner or cotton weaver, and to the blacksmith sector professions such as blacksmith or coppersmith.

A2.2 Gross domestic product

Historical GDP estimates at the provincial level for the case of Italy are not available. Thus, we proxy for provincial GDP by allocating total regional GDP (NUTS-2 units) estimates for 1871, 1881, 1901, and 1911 to provinces (NUTS-3 units) using the provincial shares of regional population obtained by population censuses. Data on GDP at historical borders (NUTS 2 units) were kindly provided by E. Felice. GDP includes of course industry, agriculture and services. It is important to stress that the industrial component of the regional GDP estimates by Felice is largely based on the statistical reconstructions by Ciccarelli and Fenoaltea considered before (we refer the reader to Felice (2013) for further details).

A2.3 Literacy

Data on literacy (population of age 6 and above) at the provincial level for the years 1871, 1881, 1901, and 1911 are those reported in the population censuses. Data on literacy rates for 1861 (used as an instrument for literacy rates in 1871) are from the population census. However, in 1861 Latium and Venetia were not part of Italy and census data are not available. To fill this gap, we proceeded as follows. Estimates of literacy rates for the provinces of Venetia in 1861 were obtained by assuming a constant 1871 to 1861 ratio of literacy rates in the regions of Lombardy and Venetia (both part of the Habsburg Empire during 1815-1860 ca). Similarly, the estimates of literacy rates for Latium in 1861 assume a constant 1871 to 1861 ratio of literacy rates of Latium on the one hand, and of the macro-area formed by Emilia, Umbria, and Marches on the other hand. The underlying assumption, is that between 1861 and 1871 literacy rates of the regions forming the Papal States (during 1815-1860 ca) evolved similarly.

A3 Geoadditive model vs. fixed effects model

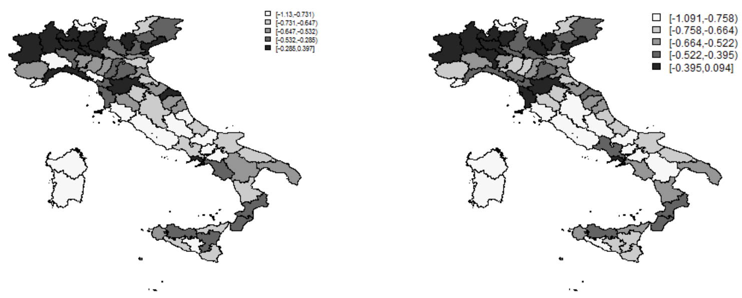
In this appendix we compare the estimation results of the geoadditive model in *Step(1)* and the corresponding fixed effects model using as dependence variable the location quotient for the whole manufacturing sector. Starting from a model without explanatory variables, we report in figure A1 the choropleth map of the estimated fixed effects, $\widehat{\alpha_i}$, from a simple model like

$$\ln\left(LQ_{i,t}^{k}\right) \quad = \quad \alpha_i + \epsilon_{i,t}$$

along with the choropleth map of the predicted values of the spatial trend, $\widehat{spt_i} = \widehat{f_1} + \widehat{f_2} + \widehat{f_{12}}$, obtained from the estimation of a simple geoadditive model without explanatory variables,

$$\ln\left(LQ_{i,t}^{k}\right) \;=\; \alpha + f_1\left(no_i\right) + f_2\left(e_i\right) + f_{12}\left(no_i, e_i\right) + \epsilon_{i,t}$$

The evidence clearly shows that both models well capture the spatial distribution of the location quotient for the whole manufacturing activity.



(a) Estimated fixed effects (b) Estimated spatial trend

FIGURE A1: Whole manufacturing. Comparing estimated fixed effects and spatial trend

Extending the two models to include the two time-varying variables (the log of Harris market potential, and the log of labor skills), we get the results reported in table A1 which confirm that the two approaches can be used as alternative specifications to control for unobserved spatial heterogeneity. The dependent variable is still the location quotient for the whole manufacturing activity. The sign of the coefficient estimated with the two models is the same, although the magnitude and level of significance is a bit different. However, the model with spatial trend has an important advantage over the fixed effects model. In particular, the geoadditive model allows us to save degrees of freedom. The fixed effect model uses $69 + 2 = 71$ degrees of freedom (one coefficient for each of the 69 regional dummies, plus two parameters for the two explanatory variables), while the model with the spatial trend uses 45 effective degrees of freedom (42 e.d.f. for the spatial trend plus 3 parameters for the intercept and the two explanatory variables). To estimate the geoadditive model, we used a number of knots equal to 14 for each of the two univariate components of the spatial trend - f_1 and f_2 - and 7 knots for the bivariate term - f_{12}. Thus, in total we used $14{+}14{+}49 = 77$ knots, i.e. a large number of knots (larger than the number of fixed effects!) to better capture nonlinearities in the spatial trend. However, the penalized spline method reduced the number of effectively used parameters (i.e. the effective degrees of freedom) to 45 (exactly 45.123). It is worth noticing that that the degree of smoothing, that is the degree of penalization, is automatically determined within the REML estimation procedure, and thus there is no arbitrary choice from the researcher. This suggests that we do not need to remove all the between-group variability in the model to filter the spatial unobserved heterogeneity. What really matters is to remove any systematic spatial pattern from the residuals which might be correlated to the explanatory variables (i.e. the source of endogeneity). We conclude that the spatial trend model must be preferred to the fixed effects model as a way to control spatial heterogeneity and to get reasonable results of the effects of the explanatory variables.[20]

Variable	Fixed effects	Geoadditive model
Intercept		-1.290***
		(0.444)
Ln(Mktpot)	0.096**	0.244***
	(0.048)	(0.047)
Ln(Literacy)	-0.262***	-0.363***
	(0.054)	(0.056)

TABLE A1: Whole manufacturing. Estimation results of the parametric fixed effect model and of the semiparametric geoadditive model. Coefficients and standard errors (in parenthesis). Number of observations: 276 (69 provinces by 4 points in time.)

A4 Indices of specialization, concentration and spatial dependence

As indicated in the main text, we use disaggregated data on manufacturing value added at 1911 prices in province i, sector k, and time t, $v_i^k(t)$. These are in particular used to compute and analyze three standard indices: the Krugman's specialization index (K_i), the Theil concentration index (C_k), and the Moran I measure of spatial autocorrelation of the location quotient LQ_i^k. This latter is defined, at any time t, as $LQ_i^k = s_i^k * (v_i/V)$, where the specialization index has the usual definition $s_i^k = v_i^k/v_i$, and $v_i = \sum_k v_i^k$ and $V = \sum_i \sum_k v_i^k$ represent,respectively, provincial and national value added.

A4.1 Krugman specialization index

We can address the question of how specialized were Italian provinces by using Krugman's specialization index K_i, that for a given point in time (t) is defined as:

$$0 \leq K_i = \sum_k |(s_i^k - s_i^{-k})| \leq 2 \tag{4}$$

where $s_i^k = \dfrac{v_i^k}{\sum_k v_i^k}$ and s_i^{-k} is the share of industry k in the national total net of province i. Thus, the Krugman index provides a measure of the difference in the specialization of a given province when compared to the remaining provinces of the country. It takes the value of zero if there is no difference, and the value of two if a province has no industries in common with the rest of Italy.

A4.2 Theil index of concentration

The Theil index of concentration (C_k) is based on the normalized location quotient, that is LQ_i^k normalized by the ratio between v_i (manufacturing va in province i) over

[20] As expected, the residuals from the estimation results of the model in equation $Step(1)$ are not spatially autocorrelated due to the inclusion of the spatial trend in the model.

V (manufacturing va in Italy). The relative Theil index C_k provides useful information about the extent to which industries are concentrated in a limited number of areas, but it does not take into consideration whether those areas are close together or far apart. In other words, it does not take into account the spatial structure of the data. Every region is treated as an island, and its position in space relative to other regions is not taken into account. Thus, the relative Theil index, $C_k(t)$, is an a-spatial measure of concentration: the same degree of concentration can be compatible with very different localization schemes. For example, two industries may appear equally geographically concentrated, while one is located in two neighboring regions, and the other splits between the northern and the southern part of the country. A more accurate analysis of the spatial distribution of economic activities requires the combination of traditional measures of geographical concentration and methodologies that account for spatial dependence, in that they provide different and complementary information about the concentration of the various sectors.

A4.3 Moran's I index of spatial dependence

Spatial autocorrelation is present when the values of one variable observed at nearby locations are more similar than those observed in locations that are far apart. More precisely, positive spatial autocorrelation occurs when high or low values of a variable tend to cluster together in space and negative spatial autocorrelation when high values are surrounded by low values and vice versa. Among the spatial dependence measures the most widely used is the Moran's I index based, as it is well known, on a comparison of LQ_i^k at any location with the value of the same variable at surrounding locations. the most widely used is the Moran's I index (Moran, 1950):

$$I = \left(\frac{N}{\sum_i \sum_j w_{ij}} \right) \left(\frac{\sum_i \sum_j w_{ij} \left(LQ_i - \overline{LQ} \right) \left(LQ_j - \overline{LQ} \right)}{\sum_i \left(LQ_i - \overline{LQ} \right)^2} \right) \tag{5}$$

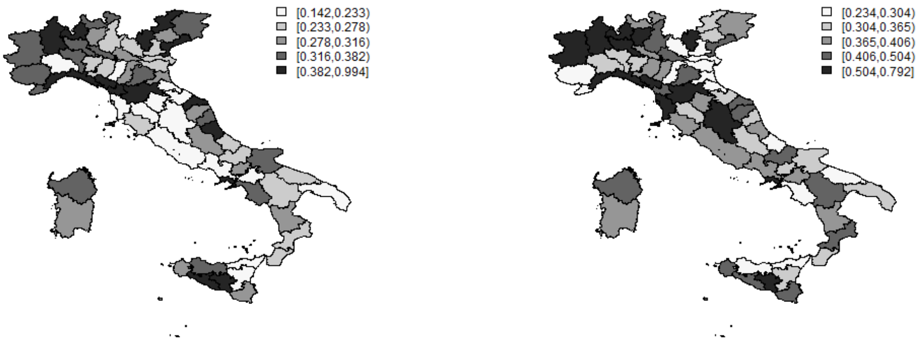
where N is the total number of provinces, LQ_i and LQ_j are the observed values of the location quotient for the locations i and j (with mean $\overline{LQ}$), and the first term is a scaling constant. This statistic compares the value of a continuous variable at any location with the value of the same variable at surrounding locations. The spatial structure of the data is formally expressed in a spatial weight matrix W (Anselin, 2013) with generic elements w_{ij} (with $i \neq j$). In the rest of the paper we will employ row-standardized spatial weights matrix (W), whose elements w_{ij} on the main diagonal are set to zero whereas $w_{ij} = 1$ if $d_{ij} < \overline{d}$ and $w_{ij} = 0$ if $d_{ij} > \overline{d}$, with d_{ij} the great circle distance between the centroids of region i and region j and $\overline{d}$ a cut-off distance (equal to 112 km, corresponding to the minimum distance which allows all provinces to have at least one neighbor).

Table A2 reports the calculated value of the Theil and Moran indices for the years 1871, 1881, 1901 and 1911 with data disaggregated by sector. The shipbuilding sector, for instance, register the highest values of the Theil index and no spatial autocorrelation (high p-values in parentheses). At the opposite side, the leather industry registers extremely low concentration and high spatial dependence.

Table A2: Industry concentration: Theil index and Moran I

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Theil				Moran			
	1871	1881	1901	1911	1871	1881	1901	1911
2.1 foodstuffs	0.02	0.02	0.04	0.06	0.15	0.31	0.25	0.29
					(-0.02)	(0.00)	(0.00)	(0.00)
2.2 tobacco	1.23	1.05	1.04	0.81	-0.04	0.00	0.02	0.01
					(-0.63)	(-0.41)	(-0.32)	(-0.38)
2.3 textile:	0.26	0.30	0.45	0.45	0.33	0.38	0.47	0.49
					(0.00)	(0.00)	(0.00)	(0.00)
2.31 cotton	*0.90*	*0.91*	*0.70*	*0.63*	*0.04*	*0.08*	*0.39*	*0.44*
					(0.23)	*(0.11)*	*(0.00)*	*(0.00)*
2.32 wool	*1.09*	*1.21*	*1.40*	*1.30*	*0.03*	*0.02*	*0.00*	*0.00*
					(0.27)	*(0.27)*	*(0.42)*	*(0.41)*
2.33 silk	*1.44*	*1.45*	*1.14*	*1.04*	*0.04*	*0.04*	*0.13*	*0.26*
					(0.04)	*(0.08)*	*(0.00)*	*(0.00)*
2.34 other natural fibers	*0.27*	*0.37*	*0.23*	*0.40*	*0.19*	*0.45*	*0.07*	*-0.06*
					(0.00)	*(0.00)*	*(0.14)*	*(0.73)*
2.4 clothing	0.11	0.13	0.10	0.11	0.29	0.30	0.40	0.35
					(0.00)	(0.00)	(0.00)	(0.00)
2.5 wood	0.02	0.02	0.03	0.05	0.11	0.10	0.22	0.37
					(-0.05)	(-0.06)	(0.00)	(0.00)
2.6 leather	0.05	0.06	0.11	0.16	0.60	0.62	0.69	0.73
					(0.00)	(0.00)	(0.00)	(0.00)
2.7 metalmaking	0.38	0.57	0.86	0.74	0.28	0.26	0.17	0.17
					(0.00)	(0.00)	(0.00)	(0.00)
2.8 engineering	0.05	0.03	0.06	0.07	0.17	0.19	0.12	0.02
					(-0.01)	(0.00)	(-0.03)	(-0.34)
2.81 shipbuildings	*1.95*	*1.56*	*1.67*	*1.79*	*0.00*	*-0.01*	*-0.02*	*-0.06*
					(0.34)	*(0.49)*	*(0.55)*	*(0.78)*
2.82 machinery	*0.16*	*0.09*	*0.22*	*0.19*	*0.02*	*0.35*	*0.06*	*0.08*
					(0.32)	*(0.00)*	*(0.16)*	*(0.10)*
2.83 blacksmith	*0.04*	*0.03*	*0.04*	*0.02*	*0.35*	*0.37*	*0.43*	*0.11*
					(0.00)	(0.00)	(0.00)	(0.05)
2.9 non-metallic mineral products	0.17	0.17	0.19	0.09	0.20	0.08	0.12	0.11
					(0.00)	(-0.03)	(0.00)	(-0.02)
2. 10 chemicals and rubber	0.18	0.16	0.23	0.17	-0.04	0.15	0.00	0.05
					(-0.61)	(-0.02)	(-0.40)	(-0.19)
2.11 paper and printing	0.21	0.21	0.17	0.16	0.11	0.07	0.06	0.07
					(-0.05)	(-0.12)	(-0.17)	(-0.13)
2.12 sundry	0.33	0.89	0.62	0.46	0.05	0.00	0.01	0.15
					(-0.19)	(-0.37)	(-0.36)	(-0.01)
2. total manufacturing	0.03	0.04	0.07	0.08	0.05	0.00	0.01	0.15
					(-0.47)	(-0.41)	(-0.09)	(-0.02)

. p-values in parenthesis.



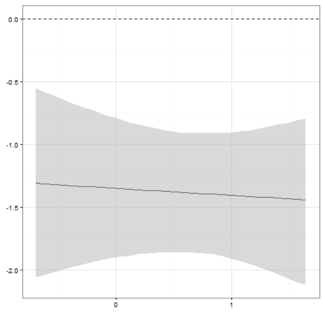
(a) Year: 1871 (b) Year: 1911

FIGURE A2: **Specialization**: Krugman index

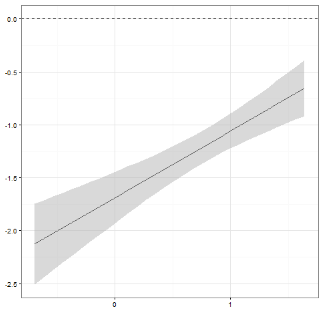
A5 The sectorial effects of Alpine regions as a function of river

The empirical model allows us to estimated the marginal effect of Alpine regions, conditional on the effect of market potential and literacy, as a function of the river variable. These effects are illustrated for the sake of completeness in Figure A3.

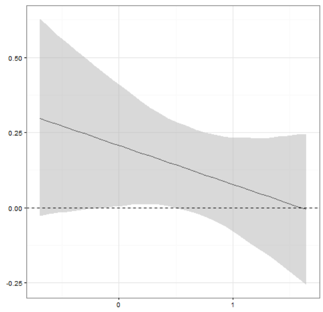
Form Figure 7 we learn that Alpine provinces had a comparative advantage in the location of industrial activity in the case of two Machinery, Cotton, Silk, Blacksmith, and Non-metallic mineral products. On the other hand, Apennine provinces had a comparative advantage in the location of industrial activity in the case of Metalmaking, Chemicals, Shipbuilding, Foodstuffs, Clothing and Wood. The estimated effect are instead not significant for Paper, Wool, Other natural fibers, and Leather. One further notice that in most low "K/L" sectors (such as for instance, Foodstuff, Clothing, and Wood) for which Apennine provinces have a competitive advantage, the effect is also increasing with the river variable.



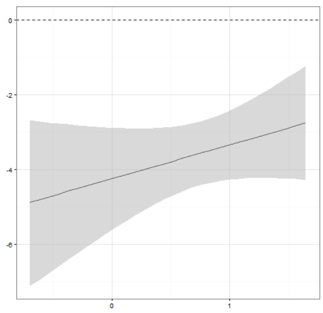
(a) Metalmaking



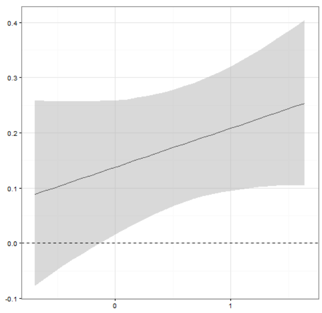
(b) Chemicals



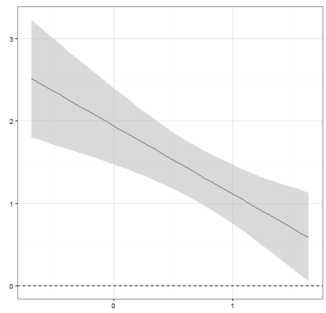
(c) Paper



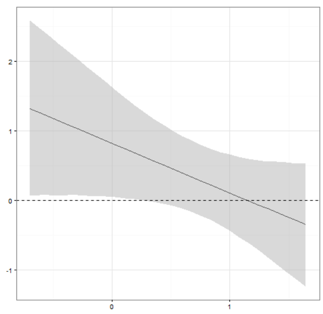
(d) Shipbuilding



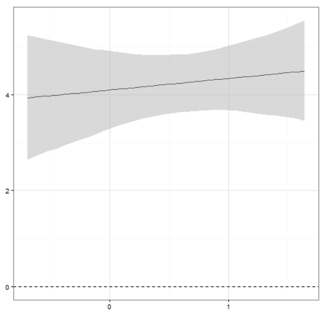
(e) Machinery



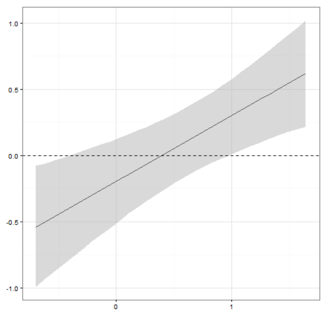
(f) Cotton



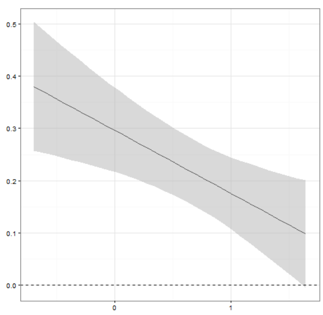
(g) Wool



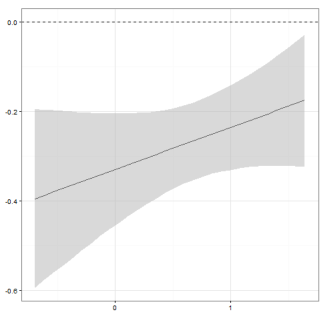
(h) Silk



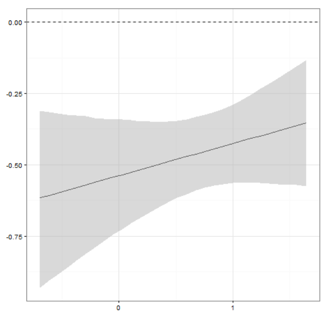
(i) Other nat. fibers



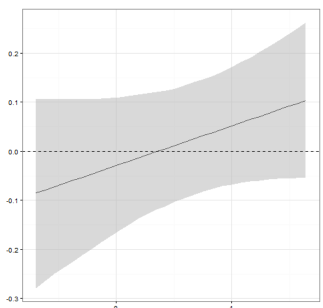
(j) Blacksmith



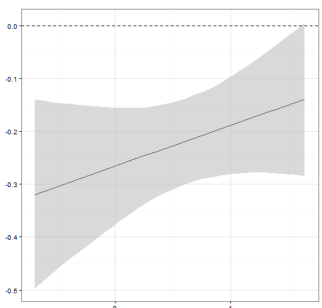
(k) Foodstuffs



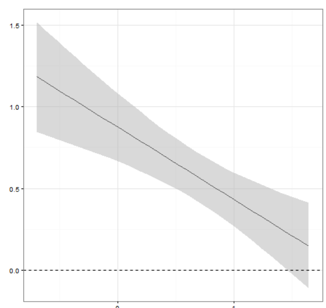
(l) Clothing



(m) Leather



(n) Wood



(o) Non-metallic min. prod.

FIGURE A3: Estimated coefficient of Alpine by ln(River) with simulated 95% confidence intervals.

References

A'Hearn, B. and A. Venables (2013): "Regional disparities: internal geography and external trade," in *The Oxford Handbook of the Italian Economy since Unification*, ed. by G. Toniolo, Oxford: Oxford University Press, 599–630.

Alonso-Villar, O., J.-M. Chamorro-Rivas, and X. González-Cerdeira (2004): "Agglomeration economies in manufacturing industries: the case of Spain," *Applied Economics*, 36, 2103–2116.

Alter, G. and G. Clark (2010): "The demographic transition and human capital," in *The Cambridge Economic History of Modern Europe: Volume 1, 1700–1870*, ed. by S. Broadberry and K. H. O'Rourke, Cambridge: Cambridge University Press, 43–69.

Anselin, L. (2013): *Spatial econometrics: methods and models*, vol. 4, Springer Science & Business Media.

Arbia, G., R. Basile, and M. Mantuano (2005): "Does Spatial Concentration Foster Economic Growth? Empirical Evidence from EU Regions," *Trivez, FJ, Mur, J., Angulo, A. Kaabia, M e Catalan, B.(eds.), Contributions in Spatial Econometrics, Zaragoza.*

Arbia, G., L. Dedominicis, and H. De Groot (2013): "The spatial distribution of economic activities in Italy. A sectorally detailed spatial data analysis using local labour market areas," *Regional Studies*, 47, 405–418.

Bardini, C. (1998): *Senza carbone nell'età del vapore: gli inizi dell'industrializzazione italiana*, Milan: Mondadori.

Basile, R., M. Durbán, R. Mínguez, J. M. Montero, and J. Mur (2014): "Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities," *Journal of Economic Dynamics and Control*, 48, 229–245.

Basile, R. and M. Mantuano (2008): "La concentrazione geografica dell'industria in Italia: 1971-2001," *Scienze regionali.Italian Journal of Regional Science*, 7, 5–28.

Blundell, R. and J. Powell (2003): *Advances in Economics and Econometrics Theory and Application*, Cambridge University Press, chap. Endogeneity in Nonparametric and Semiparametric Regression Models.

Brambor, T., W. R. Clark, and M. Golder (2006): "Understanding interaction models: Improving empirical analyses," *Political analysis*, 14, 63–82.

Brian A'Hearn, G. V. (2015): *Measuring Wellbeing*, Oxford University Press, chap. Education, pp. 175–214.

Broadberry, S., G. Federico, and A. Klein (2010): *The Cambridge Economic History of Modern Europe: Volume 2, 1870 to the Present*, Cambridge University Press, chap. Sectoral Developments, 1870-1914.

Broadberry, S. N. and N. F. Crafts (1990): "Explaining anglo-american productivity differences in the mid-twentieth century," *Oxford Bulletin of Economics and Statistics*, 52, 375–402.

Broadberry, S. N. and R. Fremdling (1990): "Comparative Productivity in British and German Industry 1907–37," *Oxford Bulletin of Economics and Statistics*, 52, 403–421.

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

Ciccarelli, C. and S. Fachin (forthcoming): "Regional growth with spatial dependence: a case study on early Italian industrialization," *Papers in Regional Sciences.*

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.

——— (2014a): "The Early Growth of the Engineering Industry in Italy's Regions," Tech. rep., Collegio Carlo Alberto.

——— (2014b): *La produzione industriale delle regioni d'Italia, 1861-1913: una ricostruzione quantitativa. 2. Le industrie estrattivo-manifatturiere*, Rome:Bank of Italy.

Ciccarelli, C. and T. Proietti (2013): "Patterns of industrial specialisation in post-Unification Italy," *Scandinavian Economic History Review*, 61, 259–286.

Ciccarelli, C. and J. L. Weisdorf (2016): "The Effect of the Italian Unification on the Comparative Regional Development in Literacy, 1821-1911," *WP CEIS no. 392, University of Rome Tor Vergata.*

Cipolla, C. (1969): *Literacy and Development in the West*, Pelican books, Harmondsworth: Penguin.

Combes, P.-P. and L. Gobillon (2015): *Handbook of Regional and Urban Economics SET*, Elsevier, chap. The Empirics of Agglomeration Economies.

Combes, P.-P., T. Mayer, and J.-F. Thisse (2008): *Economic geography: The integration of regions and nations*, Princeton University Press.

Crafts, N. (2005): "Market potential in British regions, 1871–1931," *Regional Studies*, 39, 1159–1166.

Crafts, N. and A. Mulatu (2005): "What explains the location of industry in Britain, 1871–1931?" *Journal of Economic Geography*, 5, 499–518.

——— (2006): "How did the location of industry respond to falling transport costs in Britain before World War I?" *Journal of Economic History*, 66, 575–607.

de Jong, H. (2003): *Catching up twice: The nature of Dutch industrial growth during the 20th century in a comparative perspective*, Jahrbuch für Wirtschaftsgeschichte. Beihefte, De Gruyter.

Devereux, M. P., R. Griffith, and H. Simpson (2004): "The geographic distribution of production activity in the UK," *Regional Science and Urban Economics*, 34, 533–564.

Disdier, A.-C. and K. Head (2008): "The puzzling persistence of the distance effect on bilateral trade," *The Review of Economics and statistics*, 90, 37–48.

Dumais, G., G. Ellison, and E. L. Glaeser (2002): "Geographic concentration as a dynamic process," *Review of Economics and Statistics*, 84, 193–204.

Ellison, G. and E. L. Glaeser (1999): "The geographic concentration of industry: does natural advantage explain agglomeration?" *American Economic Review*, 311–316.

Federico, G. (2006): "Industrial structure (1911–2001)," in *Evolution of Italian Enterprises in the 20th century*, Springer, 15–47.

Federico, G., S. Natoli, G. Tattara, and M. Vasta (2011): *Il commercio estero italiano: 1862-1950*, Collana storica della Banca d'Italia / Serie "Statistiche storiche": Collana storica della Banca d'Italia, Laterza.

Federico, G. and A. Tena-Junguito (2014): "The ripples of the industrial revolution: exports, economic growth, and regional integration in Italy in the early nineteenth century," *European Review of Economic History*, 18, pp. 349–369.

Felice, E. (2013): "Regional income inequality in Italy in the long run (1871–2001). Patterns and determinants," *Universitat Autï£¡noma de Barcelona, UHE working paper.*

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

——— (2004): "Textile production in Italy's regions, 1861-1913," *Rivista di storia economica*, 20, 145–174.

——— (2011): *The reinterpretation of Italian economic history: from unification to the Great War*, Cambridge University Press.

——— (2014): "The measurement of production: lessons from the engineering industry in Italy, 1911," Carlo Alberto Notebooks, No. 373, http://www.carloalberto.org/assets/working-papers/no.373.pdf.

——— (2016): "The Measurement of Production: Lessons from the Engineering Industry in Italy, 1911," *Research in Economic History*, 32, 73–145.

Fernihough, A. and K. O'Rourke (2014): "Coal and the European industrial revolution," University of Oxford, Discussion papers in economic and social history, No. 124, http://www.economics.ox.ac.uk/materials/papers/13183/Coal%20-%20O'Rourke

Fujita, M., P. Krugman, and A. Venables (1999): *The spatial economics: Cities, regions and international trade*, Cambridge, MA: MIT Press.

Gelman, A. and J. Hill (2006): *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press.

Guillain, R. and J. Le Gallo (2010): "Agglomeration and dispersion of economic activities in and around Paris: an exploratory spatial data analysis," *Environment and planning. B, Planning & design*, 37, 961–981.

Harris, C. D. (1954): "The Market as a Factor in the Localization of Industry in the United States," *Annals of the Association of American Geographers*, 44, 315–348.

Head, K. and T. Mayer (2006): "Regional wage and employment responses to market potential in the EU," *Regional Science and Urban Economics*, 36, 573–594.

——— (2011): "Gravity, market potential and economic development," *Journal of Economic Geography*, 11, 281–294.

IGM (2007): *Classificazione dei corsi d'acqua d'Italia*, Florence: Istituto Geografico Militare.

Klein, A. and N. Crafts (2012a): "Making sense of the manufacturing belt: determinants of US industrial location, 1880–1920," *Journal of Economic Geography*, 12, 775–807.

——— (2012b): "Making sense of the manufacturing belt: determinants of US industrial location, 1880–1920," *Journal of Economic Geography*, 12, 775–807.

Krugman, P. R. (1991): *Geography and trade*, MIT press.

Lee, D.-J. and M. Durbán (2011): "P-spline ANOVA-type interaction models for spatio-temporal smoothing," *Statistical Modelling*, 11, 49–69.

Martinez-Galarraga, J. (2012a): "The determinants of industrial location in Spain, 1856–1929," *Explorations in Economic History*, 49, 255–275.

——— (2012b): "The determinants of industrial location in Spain, 1856–1929," *Explorations in Economic History*, 49, 255–275.

Martínez-Galarraga, J. (2014): "Market potential estimates in history: a survey of methods and an application to Spain, 1867-1930," *EHES Working Papers in Economic History*, 51.

McMillen, D. P. (2012): "Perspectives on spatial econometrics: Linear smoothing with structured models," *Journal of Regional Science*, 52, 192–209.

Midelfart-Knarvik, K. H. and H. G. Overman (2002): "Delocation and European integration: is structural spending justified?" *Economic policy*, 17, 321–359.

Midelfart-Knarvik, K. H., H. G. Overman, S. J. Redding, and A. J. Venables (2000a): "The location of European industry," *European Economy. Economic Papers*, no. 142, 1–67.

Midelfart-Knarvik, K. H., H. G. Overman, and A. Venables (2000b): "Comparative advantage and economic geography: estimating the determinants of industrial location in the EU," *Centre for Economic Policy Research. Discussion papers, no. 2618.*

Moran, P. A. (1950): "Notes on continuous stochastic phenomena," *Biometrika*, 17–23.

Redding, S. and A. J. Venables (2004): "Economic geography and international inequality," *Journal of international Economics*, 62, 53–82.

Rosés, J. R. (2003): "Why isn't the whole of Spain industrialized? New economic geography and early industrialization, 1797–1910," *Journal of Economic History*, 63, 995–1022.

Rostas, L. (1948): "International comparisons of productivity," *Int'l Lab. Rev.*, 58, 283.

Ruppert, D., M. Wand, and R. Carroll (2003): *Semiparametric Regression*, Cambridge: Cambridge University Press.

Sundar, V. J., R. Ramesh, P. S. Rao, P. Saravanan, B. Sridharnath, and C. Muralidharan (2001): "Water management in leather industry," *Journal of Scientific and Industrial Research*, 60, 443–450.

Tiefelsdorf, M. and D. A. Griffith (2007): "Semiparametric filtering of spatial autocorrelation: the eigenvector approach," *Environment and Planning A*, 39, 1193–1221.

Tirado, D. A., E. Paluzie, and J. Pons (2002): "Economic integration and industrial location: the case of Spain before World War I," *Journal of Economic Geography*, 2, 343–363.

Vigo, G. (1971): *Istruzione e sviluppo economico in Italia nel secolo XIX*, ILTE.

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.

Wood, S. (2006): *Generalized additive models: an introduction with R*, CRC press.

Wooldridge, J. M. (2010): "Econometric Analysis of Cross Section and Panel Data, Vol. 1 of MIT Press Books," .

Zamagni, V. (1993): *The economic history of Italy 1860-1990*, Oxford: Clarendon Press.

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