Infrastructure and regional growth: railways and the productivity gap in Italy after Unification

Nicola Pontarollo, JRC - Joint Research Center,
Via E. Fermi, 2749, 21027 Ispra (VA), Italy.
Email: nicola.pontarollo@ec.europa.eu

Roberto Ricciuti, University of Verona and CESifo
Department of Economics, Via Cantarane 24, 37129 Verona – Italy.
Email: roberto.ricciuti@univr.it.

Abstract

The political unification of Italy in 1861 led to the establishment of a single market with a single currency. Market integration was the economic outcome of this process. At the same time, the Kingdom of Italy started a large infrastructure project to spread railways all over the country. Using tools from spatial econometrics, we find that railways played a positive effect on productivity, but this effect was stronger in the areas in which railways were already built. This effect is in line with New Economic Geography according to which infrastructure lead to a widening of territorial disparities.

Keywords: Railways, productivity, economic growth, spatial econometrics.

JEL Codes: R12, L92, C23, O18, N93.

1. Introduction

The political unification of Italy in 1861 led to the establishment of a single market, by removing the trade barriers across the pre-existing states, with a single currency. Market integration was a corollary of this process, and more productive Northern industries had the opportunity to reach a larger market, further reducing their average costs and boosting their productivity. Low-productivity Southern firms were crowded-out. At the same time, the Kingdom of Italy started a large infrastructure project based on railways.

Public capital, in general, and infrastructure, in particular, have been regarded as "'unpaid factor(s) of production' which directly encourage increased output; 'augmenting factors' which enhance the general productivity of private capital and labor inputs; and in a more dynamic sense incentives for firm and household (re)location and long term economic growth" (Lewis 1998: 142).

In this paper we analyze the interplay between the market forces that lead to concentration and this policy effort. In principle railways may reinforce the concentration process since they reduce transportation costs, making Northern goods cheaper and displacing Southern ones. However, the transportation cost argument may also apply to Southern firms, which could at least partially offset their productivity-disadvantage. We use economic history as a testbed for this analysis, taking advantage of the extremely low level of infrastructures at the time. This should provide a clearer evidence with respect to contemporary analyses, in which the construction of new infrastructure is only marginally incremental with respect to the existing stock.

Our main finding is that railways play a crucial role with respect to productivity. This effect is in line with New Economic Geography according to which infrastructure lead to a widening of territorial disparities: by providing central and peripheral regions with a similar degree of accessibility, lagging provinces result to be disadvantaged, as their firms are in a weaker position to compete than firms in the core (Puga, 2002). This is much more evident in post-Unification Italy where the fundamental production factors were the immovable natural resources.

We extensively use spatial econometrics techniques to assess the importance of relative location in space. These techniques allow us to analyze the main spatio-temporal dynamics of the selected variables. We first use an exploratory spatial data analysis (ESDA) that allows us to disentangle spatial evolution of productivity and railways. Then we examine the relationship between transport infrastructure and productivity with a spatial panel spatial filtering approach. This represents an advance with respect to classical technique because it is able to deal with spatially autocorrelated variables and it also accounts for omitted time-invariant variables explicitly considering their spatial dimension.

The use of spatial econometrics to understand the relationship between infrastructure improvements and economic performance is growing. For the US, Atack and Margo (2011) using GIS data, estimate that at least a quarter (and possibly two-thirds or more) of the increase in cultivable land in 1850s can be linked directly to the coming of the railroad to the Midwest. Farmers responded to the reduction in transportation costs, which raised agricultural revenue productivity, by expanding the area under cultivation. One-half or more of the growth in urbanization in the Midwest in the late antebellum period is attributed to the spread of the rail network (Atack et al. 2013). Donaldson and Hornbeck (2016) argue that as the railroad network expanded from 1870 to 1890, changes in market access were capitalized into county agricultural land values with an estimated elasticity of 1.1, and removing all railroads in 1890 lead to an estimated decrease of the total value of US agricultural land by 64%. Surveying studies for England, Bogart (2014) maintain that transport improvements reduced freight charges by 95 percent in real terms from 1700 to 1870, with a growth in annual TFP of more than 2 percent.

The paper is organized as follows: Section 2 reviews the history of railways in Italy, while Section 3 discusses some issues related with the North-South divide. In Section 4 we exploit the geographical dimension of our data in order to identify some patterns. Section 5 introduces the modeling techniques, whose results are discussed in Section 6. Section 7 provide some robustness checks. Section 8 concludes.

2. A brief history of railways in Italy

Schram (1997) distinguishes four phases in the development of railways in Italy. The first one took place from 1839 (when the first line was built) to 1865, foreign private companies built the main lines under concession from the pre-unitary states (figure 1). Piedmont was an exception, since the railways was built and managed by the state from 1850s. The concessionary regime was different across states and also within the same state, leading to confusion and disparities across the companies. After Unification in 1861, a spur of investments took place (figure 2), with foreign investors willing to increase their involvement, given the plans of the new state, and its higher merit of credit. It should be noted that more than half of the state's spending in the 1860s and 1870s was on railways.

The second phase started in 1965 with the Railways Act that reorganized railways companies among five franchisees (*Upper Italy Railway Company* that run the service in the North, the *Meridionali Railway Company*, which managed the line between Tuscany and the Adriatic South, the *Romanae Railway Company* in central Italy, *Reale Sarda Railway Company* in Sardinia and the *Victor Emmanuel Railway Company* in Sicily). Moreover, the Piedmontese State Railways were privatized because of the financial needs of the Kingdom, and the Upper Italy Railway Company was formed. The aim of the Act was to attract capital in order to expand the rail network by offering a high rate of return.

However, the returns were often negative and companies were bailed out by the state. By 1878 – which marks the beginning of the third phase - the government was in charge of most of the tracks and operating companies. In 1885 a second Railway Act passed, opting for a mixed system in which the tracks were state-owned, whereas the three operating companies (the *Mediterranean Railway Company* that operated the western network, the *Adriatic Railway Company* that was in charge of the eastern network, and the *Sicilian Railway Company* that ran the railways in the South and in the main islands) belonged to the private sector. The last phase began with the nationalization in 1905 that was needed because of the poor performance of *Mediterranean* and the *Adriatic* railway companies.[1]

[Figures 1, 2 and 3 about here]

The early literature on railways was concerned with the consequences of this infrastructure on growth. Romeo (1959) interpreted the resources devoted by the state to the construction of railways (and other public services) as a mechanism that, through taxation and government spending, redirected wealth from the relatively rich agriculture to the infant manufacturing sector. Investing in railways was a form of capital accumulation, conducive to the take-off of the Italian economy in the following decades. The Marxist interpretation was given by Sereni (1966) who claimed that railways fulfilled the aim of the bourgeoisie to create a unified market, and its cost was borne by the working class. Both the liberal and the left-wing approaches placed a lot of emphasis on the role of infrastructure, but they differed on the evaluation burden of this investment.

---

[1] This phase ended at the late 1990s, when some EU-promoted liberalizations started, mainly in the cargo sector, and a duopoly on high-speed lines was established in 2010.

Gerschenkron (1962) was the first scholar to call into doubts the positive effects of railways on economic growth. He maintained that the railway network was built too early with respect to the industrialization of the country. During the first wave of investments in the 1880s, it needed imports from abroad since there was not enough domestic supply of investment goods, therefore there was no supply to meet this demand. At the time of the second wave of investments after nationalization, it was too late to contribute to industrialization. Railways failed to unify the market because there was not enough national demand, and their cost was too high with respect to the benefits. Had the timing of investments in railways been different, the growth rate in the Giolittian age would have been higher.

Fenoaltea (1983, 2011) shares the idea that railways did not bring a unified market, but for completely different reasons. First, lines were subsidized by the state according to their actual length and not the air-distance between two cities, incentivizing long and tortuous railways that in the end were ineffective in cutting down journey times. Second, the fare structure was too costly for long-distance journeys. As a results, the railway network was used much below capacity, which in turn led to negative returns, underinvestment by the private companies and then nationalization. State-ownership was also inefficient because it expanded the personnel in absence of an increase in traffic.

3. The North-South divide

There is an established literature on the economic dualism between the North and the South in Italy that spans the last sixty years. Only recently regional estimates of industrial production and GDP have been produced. Felice (2011, 2013) has constructed regional disaggregations of the new national-product estimates for a number of benchmark years. A long-term research project culminated with the publication of a large number of regional series by Ciccarelli and Fenoaltea (2009, 2014).

Ciccarelli and Fenoaltea (2013) take a step forward and provide measures of industrial production for the 69 Italian provinces for the same years. This disaggregation over space was obtained allocating regional Value Added estimates to provinces using census labor-force data. The picture they provide is faceted. In the aftermath of Unification the leaders were in former capital cities, where there was a strong base of artisans, and a small part of the South, where selected provinces reaped the gains from the freer foreign trade (the extension to the whole country of the free-trade policy of the Piedmont's Kingdom), and infrastructure investment. But this was a short-term effect: over the later nineteenth century, when the movement from craftsmanship to industry became stronger, industry concentrated into the 'industrial triangle'. The early twentieth century brought both industrial diffusion and concentration. The latter to the center/north-east, where it was tied to the production of perishables on recently improved land, the latter within the north-western triangle itself, into its major cities. This movement was brought by progress in energy transmission, which made production of goods less tied to the waterfalls from which electricity was produced.

Figure 4 depicts the quintiles of the share of manufacturing Gross Value Added (GVA) at 1911 prices on male population over 15 (Ciccarelli and Fenoaltea, 2013) in the first and the last years considered in our analysis (and described in Section 4).

[Figure 4 about here]

We can see the two movements mentioned above: concentration and limited diffusion. Provinces in the first quintile in 1871 remain there (with the exception of Venice), and are joined by a few contiguous provinces. At the same time, provinces in the North/North-East tend to become darker, therefore upgrading in the distribution. However, provinces in the South go down in the development ladder in relative terms, as shown by their lighter colors.

According to Ciccarelli and Fenoaltea, railways play some role in this pattern. For example, in the North-West there are two opposite cases. Cuneo, in south-west Piedmont, exploited his position on the road from Turin to Nice until the latter was ceded to France because of the Turin Treaty in 1860. Then the rail line was built from Genoa to Turin through Alessandria, sidelining Cuneo. Genoa was one of the main beneficiaries of railway construction: the inland railway was used to ship the goods produced in the upper Po valley through its port, avoiding transportation along the lower Po valley to the Adriatic Sea. Also these provinces were cut from trade done by river navigation, which up to the introduction of railways was the cheapest transportation mean.

4. Data and their geographical structure

The data used in this empirical analysis refer to 69 provinces and 4 benchmark years (1871, 1881, 1901 and 1911).[2] In addition to the data on manufacturing GVA and male population over 15, also area is from Ciccarelli and Fenoaltea (2013), km of railways are from Ministero delle Comunicazioni, Ferrovie dello Stato (1927), km of roads[3] from L'Italia Economica (1873) and Annuario Statistico Italiano (1892, 1900 and 1912), and literacy rate from census (1871, 1881, 1901 and 1911) data (Ministero di agricoltura, industria e commercio, 1875, 1883, 1904, 1914).

While data on roads are disaggregated by province, data on the Italian railroad network are not. In the Ministero delle Comunicazioni, Ferrovie dello Stato (1927) only the year in which each rail line has been opened and the departure and destination locations are reported. In order to calculate the km of railroad network by province, we reconstructed the route of each

---

[2] In 1891 there was no census.

[3] Due to availability problems, we have data for roads on 1872, 1880, 1897 and 1910 which are used for the most proximate benchmark year.

line. Due to the importance of intra-regional and intra-regional trade in subnational development (Martin and Rogers, 1995), we decided to include in the analysis secondary lines with the same gauge of main lines, thus excluding extra urban tramway lines.

Manufacturing GVA/male population over 15 has been used as a proxy of labor productivity because we followed the framework proposed by Ciccarelli and Fenoaltea (2013) who maintain that the share of the male population over age 15 is equivalent to the male population of working age, or, to a first approximation, the male labor force. Although the construction sector was sizable, we decided to concentrate only on manufacturing, because construction was extremely cyclical and in an empirical analysis using only a few benchmark years these swings could strongly affect the results.

Figure 5 depicts the quintile maps of the mentioned variables for each year. Railways appear rather concentrated in the north-west and in the center-north of Italy in 1871, while, in 1911, they seem a little more sparse. Anyway, a clear spatial pattern is not evident neither in the first nor in the last year. Some areas that in 1871 were characterized by a low density of this infrastructure kept their advantage, while others caught up from the initial lagging situation.

Along railways, there was also an investment in roads. The law on public works (March 20, 1865, n. 2248) classified the roads in four categories according to their importance: national, provincial, municipal and local. In figure 5 we observe that in both periods mandatory municipal roads were mainly concentrated in the North. The provincial and national roads change their spatial patterns between the first and the second considered year. In 1871 they were much more concentrated in the center-north part of Italy, while in the last year the northern part was characterized by a lower concentration, probably because the development of the railway was so strong and widespread to make not necessary the construction of such type of roads.

[Figure 5 about here]

The spatial dimension of the selected variables, clearly shown in figures 4 and 5, can be investigated through a specific statistics called Moran's I (*MI*). The Moran's I has been widely used in the literature to describe economic phenomena whose distribution is not random in space (Le Gallo and Ertur, 2003 and Gregory and Patuelli, 2015).

The Moran's I, which describes the degree of clustering in spatial data, is defined as:

$$MI = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}\left(y_i - \bar{y}_i\right)\left(y_j - \bar{y}_j\right)}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (1)$$

where *i* and *j* refer to different spatial units (i.e., cell centroids) of which there are *n*, *y* is the data value in each and *w* the element of the line *i* and row *j* of the spatial weights matrix **W** of $n \times n$ size. The calculated Moran's I for global autocorrelation varies between -1 and 1. A positive significant coefficient points to positive spatial autocorrelation, i.e. similar values cluster together in a map. The reverse represents regimes of negative association, i.e. dissimilar values cluster together in a map.

The choice of spatial weights matrix **W** is based on potential accessibility (ESPON, 2007), based on the assumption that the attraction of a destination increases with size, and declines with distance, travel time or cost. Potential accessibility of province *i* with respect to *j*, $w_{i,j}$, is a construct of two functions combined multiplicatively, i.e. the total of the activities reachable at *j* weighted by the ease of getting from *i* to *j* (Wegener et al., 2002):

$$w_{ij} = A_j \exp\left(-\beta d_{ij}\right) \qquad (2)$$

where A_j is the provincial population to be reached in province *j*, and $d_{i,j}$ is the distance between province *i* and *j*. The interpretation is that the more province *j* is attractive, and the more is accessible from province *i*, the greater is the accessibility of province *j* with respect to *i*. In other terms, province *j* is much closer to province *i* in term of market potential. For this weighting, the parameter β has been set to 0.02, as customary when we consider national trips (Andersson and Karlsson, 2007).

We also imposed that each province must have at least one neighbor and set a cut-off distance of 110 km to avoid to include neighbors with negligible weights. The idea behind this choice, largely accepted in spatial econometrics, is to construct a spatial weights matrix that accounts for the important linkages among regions, avoiding the inclusion of too many neighbors.[4]

In table 1 we report the descriptive statistics and the Moran's I of the selected variables that we will use in our econometric model by year. The GVA per working age male person is strongly spatially autocorrelated and it tends to be much more clustered over the time. This clearly shows an increasing polarization of the industry, a result found also by Ciccarelli and Fenoaltea (2013) using a different methodology and observed also in the increasing gap between the more and less productive provinces. A similar but stronger concentration pattern is followed by literacy. The railways are rather positively autocorrelated only until 1871, with a p-value of 0.035, and in the following periods they are much more randomly distributed in space, highlighting the role of the central Government in bridging the gap among provinces with respect to this particular kind of infrastructure. The average km of railways over square kilometers triplicates between 1871 and 1911, but there are still provinces without rail accessibility. Also population density is not clustered and does not follow a defined spatial pattern. There are permanent differences between provinces with zones of sparse population and others with strong polarization that tend to attract people from other areas.

---

[4] The spatial weights matrix is then, as customary, standardized by row.

Finally, the municipal roads density (figure 5) is characterized by a strong spatial pattern, confirmed by the Moran's I for all examined years. However, the construction of new infrastructure tends to mitigate this situation. Municipal roads in 1871 were strongly clustered in the North of the peninsula (Moran's I = 0.80, p-value < 0.001) but the situation improved over time with a spatial autocorrelation index equal to 0.36 (p-value < 0.001) in 1911 pointing to a more homogeneous distribution. The density of municipal roads is stronger than national/provincial roads over the whole period, but it is noteworthy that there are provinces where there are no municipal roads, which means that there are municipalities without adequate connections with capital cities or with the main communication network. National and provincial roads are more equally distributed across provinces as shown by the not significant Moran's I. The exception is year 1901 and it is due to three outliers (Avellino, Chieti and Naples).

[Table 1 about here]

The spatial relation described by the Moran's I can be shown in the Moran scatterplot which relates a selected variable with its spatially lagged values (figure 5).[5]

Most provinces are either in the first (top right) or third quadrant (bottom left), which means that there is a clusterization of provinces with similar values of GVA/male population, respectively low and high, in space. The variations between the first and last years, as well as

---

[5] The scatterplot is divided into 4 quadrants (anticlockwise from top right): in the first and third (high-high, HH, and low-low, LL, respectively) a province that exhibits a high (low) value of the variable is surrounded by provinces with a high (low) value of the variable as well. In the second and fourth (low-high: LH and high-low: HL, respectively) a province with a low (high) value of the variable is surrounded by provinces with a low (high) value of the variable.

the higher *MI* in year 1911 and the increasing concentration of points in third quadrant of Moran's scatterplot confirm the divergence between North and South of Italy. Following Rey (2001), the transition dynamics between the four different types of spatial association outlined above is reported in table 2. The main diagonal shows the number of provinces (the percentage by row is in brackets) that do not move from their original quadrant. The higher persistence is in the first and third quadrants with around 63% and 96% of provinces, respectively, that did not change their original cluster.

[Figure 6 about here]

The provinces in the second and fourth quadrants are residual and generally show a transition to a cluster of low productive provinces. The shift has been sizable since in 1871 there were 28 provinces in third quadrant (low-low) and 35 in 1911. Table 2 allows us to analyze whether the transition involves only the province or also its neighbors. The most important variations concern the first and second quadrants, the first and second rows, respectively. Regarding the first quadrant, in 3 cases the provinces lowered their GVA/male population with respect to their neighbors and in other 3 cases both the province and the neighbors lowered their GVA/male population. In the second quadrant, in 3 cases the neighbors decreased their levels of GVA/male population at the level of the considered province, while only one province increased its productivity per male person at the levels of the surrounding provinces. Therefore, declining provinces tend to lose ground together with their neighbors.

[Table 2 about here]

5. Model and estimation technique

In this section we explicitly focus on the relation between infrastructure and provincial productivity, taking into account the time dimension of our data.

According to classical location theory, transport infrastructure endowment and investment lead to high returns. In this extent greater accessibility and lower transportation costs facilitate trade and lead to a reduction in the prize of traded goods, by allowing different territories to maximize their comparative advantage. The baseline model, similarly to Holtz-Eakin and Schwartz (1995), takes the form of an infrastructure-augmented, production function over time where we want to single out the correlation of disaggregated infrastructure stock and provincial productivity levels:

$$Y_t = K_t^{\alpha} G_t^{\beta} \left(\psi_t L_t\right)^{1-\alpha-\beta} \tag{3}$$

where Y_t is the level of output, K_t is the private capital, G_t is the public capital, L_t is the labor force and ψ_t is an index of technical efficiency and t the time periods. We consider that both ψ_t and L_t grow at constant rates. Taking the logs and dividing all variables by the effective quantity of labor we have the following equation:

$$\ln y_{et} = \alpha \ln k_{et} + \beta \ln g_{et} \tag{4}$$

where the subscript e denotes quantity per effective labor unit and t time. In the model, public capital is represented by infrastructure and it is assumed to be complementary to labor and capital (Moreno and Lopez-Bazo, 2007). According to Barro (1990) when public infrastructure is an input in the production function, an increase in public infrastructure raises the marginal product of private capital, which leads to an increase in capital accumulation and then the speed the convergence. Another position comes from the New Economic Geography. Krugman (1991) and Krugman and Venables (1995) highlight that economic integration may lead to a "core-periphery" pattern due to the reduction in transportation costs. At this regard, the public investments in infrastructure have a crucial role in exacerbating or mitigating the spatial concentration of increasing returns to scale industries in the "core" and the concentration of constant returns to scale industries in the "periphery". Following Martin and Rogers (1995), public infrastructure facilitates transactions inside the region (intra-regional trade attracting firms and contributing to convergence). If the infrastructure is financed by transfers, it facilitates inter-regional trade, rather than intra-regional trade, and regional policy can have a detrimental effect for the poor region.

In our empirical model we have quite detailed data on infrastructure, but we do not have data on capital. Then, in order to avoid model misspecification, we include some additional control variables and we adopt a technique able to deal with the problems of omitted variables and spatial dependence.

The empirical model is:

$$\begin{aligned}
&\ln\left(GVA/MPOP_{i,t}\right)=\delta+\gamma_{1,t}\ln RAIL_{i,t}+\gamma_{2,t}\ln MUN\_ROADS_{i,t}+\gamma_{3,t}\ln NAT\_ROADS_{i,t}+\gamma_{4,t}\ln AGG_{i,t}+\\
&\gamma_{5,t}\ln DIST\_PORTS_{i,t}+\gamma_{6,t}\ln RAIL_{i,t}\times\ln MUN\_ROADS_{i,t}+\gamma_{7,t}\ln RAIL_{i,t}\times\ln NAT\_ROADS_{i,t}+\\
&\gamma_{8,t}\ln RAIL_{i,t}\times D_{MORAN\_RAIL}+\gamma_{9,t}D_{MORAN\_RAIL}+\gamma_{10,t}D_{SOUTH}+\gamma_{11,t}\ln LITERACY+\gamma_{12,t}\ln LOWLAND+\\
&\gamma_{13,t}\ln RAIL_{i,t}\times\ln LOWLAND+\varepsilon_{i,t}
\end{aligned} \tag{5}$$

where the subscript i and t denote, respectively, the province and the year and ε the idiosyncratic error term. The variables are in logs and are the same described in the previous section.[6] D_{MORAN_RAIL} represents a dummy that takes the value 1 for the provinces which lies in the first quadrant of the Moran Scatterplot with respect to their railways endowment until 1871, i.e. the cluster of provinces that was the earliest adopter of this type of infrastructure.[7] As the

---

[6] The problem of some regions with zero kilometers of roads and/or railways has been overcome by adding a 1 to these variables.

[7] They include: Turin, Alessandria, Novara, Genoa, Pavia, Milan, Cremona, Bergamo, Leghorn, Pisa, Siena, Grosseto, Florence, Bologna, Ravenna, and Forlì.

distribution of rail lines has been polarized only in the first period, the variable allows to account for the possible advantage of being the first adopters of a new technology. The dummy meets two conditions: the first is that the railways were above the national average, and the second is that the railways were above the average also in the neighbor provinces. D_{SOUTH} is a categorical variable that refers to the provinces of Southern Italy.[8] Data on percentage of people able to read and write (*LITERACY*) have been added to account for human capital, the percentage of provincial lowland territory (provincial data are reconstructed from ISTAT, 2009) is a proxy for the relatively easier task of building a railway on a flat area and *DIST_PORTS* is distance, computed as the km, of provincial centroids from the nearest main port: Trieste, Venice, Livorno, Naples, Messina, Palermo. Finally, we included three interaction terms to verify the relation between different types of infrastructure, the interaction terms between railways and D_{MORAN_RAIL} to test if the early railways endowment has an effect on improving the effectiveness of subsequent railways construction, an interaction term to assess the joint effects of the railways with road infrastructure and an interaction term to capture whether orography affected the effectiveness of railway infrastructure.

In the context of our analysis, we think that we do not face an endogeneity problem. This is well explained by Russel (1985: 42) who made clear that "the Italian 'railway boom' was not the driving stimulus to industrial development that it was elsewhere" and he adds that "the process of railway building was not closely related to the progress of industry in time or space".

The strong spatial autocorrelation detected in the previous section makes it necessary to consider this issue to avoid biased estimations in regression analysis. When dealing with data at regional or subnational level, however, as pointed by McMillen (2003), model

---

[8] Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia and Sardinia belong to this area.

misspecification, due for example to incorrect functional form and omitted relevant variables, can lead researchers to wrongly fit a spatial autoregressive model to the data. Among the omitted variables there is a number of factors either unobservable or not properly measured related, for example, to culture, social capital and institutional characteristics that can affect regional performance. These factors, which are often not randomly distributed in space, can be modelled through the so-called spatial filters which decompose Moran's I to extract eigenvectors to be employed as additional explanatory variables that surrogate spatially distributed region-specific unobservable information usually incorporated in the fixed effects parameters. The advantage of this technique, with respect to spatial lag and error models, is that it allows to reduce the stochastic noise normally found in the residuals of standard statistical methods (Patuelli *et al.*, 2011), managing the omitted variables issue, avoiding to specify an incorrect functional form.

Our spatial model is based on the spatial contiguity matrix **W** and the associated Moran'I. If we rewrite equation (1) in matricial form we have:

$$MI = \frac{n}{1^t \mathbf{W} 1} \frac{\mathbf{Y}^t \mathbf{MWMY}}{\mathbf{Y}^t \mathbf{MY}} \tag{6}$$

where $\mathbf{M} = (\mathbf{I} - \mathbf{1}\mathbf{1}^t)/n$ is the matrix in which **I** is the identity matrix of size n-by-n, **1** is a vector of one dimension n-by-1 and the apex t points the transposed matrix. The peculiarity of the **M** matrix is that it centers the vector of data value **Y**. Tiefelsdorf and Boots (1995) demonstrate that each of the n eigenvalues of expression **MWM** is a value of the Moran's I, once it is multiplied by the left-hand term of expression (6), namely $n/\mathbf{1}^t\mathbf{W}\mathbf{1}$. This allows the extraction from the n-by-n matrix of uncorrelated numerical orthogonal components (Tiefelsdorf and Boots, 1995). This nonparametric approach has the aim of managing the presence of spatial autocorrelation by introducing a set of variables, the eigenvectors, able to catch the latent spatial association of georeferenced variables (Getis and Griffith, 2002). These *n* eigenvectors describe the full range of possible orthogonal, unrelated spatial patterns and can be interpreted as a summary map of variables that describe the nature (positive or negative) and the level (low, moderate, high) of spatial autocorrelation. Selected eigenvectors can be used also as predictors instead of not explicitly considered variables (Fischer and Griffith, 2008) and, since they are both orthogonal and uncorrelated, a stepwise linear regression can be used to achieve this end.

In case of panel analysis with spatial filters, we follow the stages proposed by Patuelli *et al*. (2011) in which a stepwise regression is performed for each year in order to select the significant eigenvectors. Then, for the subsequent panel analysis, only eigenvectors that are common for each year are accounted.

The panel model with spatially structured random effects is able to capture dependence obtained throughout space in the whole period via a Mixed Generalized Linear Model with an intercept that varies in space according to a normal distribution. The advantage of using this model relies in the exact identification of the time specific effects via spatial eigenvectors, which are able also to take into account the spatial dimension of the omitted variables. Then, we will avoid the degree of freedom problems, typical of fixed effect framework and we will explicitly account for the spatial dimension of time invariant variables, without the need to surrogate this dimension with the use of specific dummies.

6. Results

The estimated baseline model is in table 3. The coefficient of railways is positive and highly significant across all specifications. The result is in line with Aschauer's (1989) idea that differences in the stocks of public infrastructure and private capital could provide an important explanation for differences in levels of output. Other things being equal, railways increase productivity throughout the country, with no significant differences between the North and the South. In additional regressions we find that the South dummy is not significantly different from zero, therefore excluding the presence of a dichotomy *per se* due to a pure geographical criterion. Lowland is not significant too, as well as the interaction term, showing that the construction of Italian railways was not facilitated by this type of territory. Human capital, instead, is strongly positive and significant.

[Table 3 about here]

Also the dummy indicating the provinces characterized by the higher initial endowment of railways is not significant, showing that they did not get a permanent advantage from the early construction of this transport infrastructure. This is due to the type of industrial production, which was much related to the exploitation of natural (mainly water) resources and then located near them. Water was the fundamental in the textile industry (Federico, 2005) and the main resource used in electricity production (Bezza, 1986). Water was unevenly distributed across Italy, with the North enjoying an abundant and stable supply, not available in the South. Railways, then, served as a link between firms – which located close to water and electricity plants - and the main markets. The early presence of this infrastructure reduced travel time making it possible to increase the reachable potential market, but it did not produce a comparative and lasting advantage by itself. An explanation of the absence of permanent advantage for those regions with an early presence of railways can be found in the lack of complementary infrastructure, mainly roads whose situation was particularly bad, or in the underutilization of the railways for the lack of connections, that did not generate an appreciable advantage in the territories where this infrastructure came first.

However, the lack of difference between the North and the South stops here. The interaction term between the mentioned dummy and the constructed km of railways allows us to check whether the clusters of provinces that benefitted first from this infrastructure had a comparative advantage from the enlargement of the network. Results in models (3), (5) and (6) show that this variable is positive and significant: the addition of a new railway in early infrastructured provinces leads to an additional advantage in productivity of these provinces, contributing to the divergent spatial patterns clearly shown in figure 5 and table 2. To this extent, it is interesting to observe that, of the 13 provinces belonging to first quadrant (high-high) of the Moran Scatterplot in 1911 (figure 6), 11 belong also to the group with early railways construction. From this result we can state that under equal conditions with respect to natural resources, provinces with early railways are more productive because they exploit the whole potential of new railways construction that have a direct multiplier effect within each provinces. On the other hand, we have that the presence of a widespread network tends to concentrate its positive effects in a limited number of provinces without spreading its potential positive effects to the neighbors.

Was there enough trade between the North and the South to motivate this developments? Fenoaltea (1983) claimed that there was little scope for trade between northern and southern Italy because of the modest complementarity between the goods produced. However, as Schram (1998) documented, imports by rail from southern Italy grew from 67,340 tons in 1867 (a mere 9% of overall import) to 107,536 tons in 1870, 225,468 tons in 1875, 327,886 tons in 1880 and 346,423 tons in 1884 (12% of the total). At the same time, export by rail to southern Italy was 121,013 tons in 1867 (34% of the total), 89,927 tons in 1870, 144,149 tons in 1875, 175,550 tons in 1880 and 202,258 tons in 1884 (24% of overall exports). Economic integration between the two areas strongly increased in absolute terms, with the North importing more than the South, but concentrating in raw and intermediate materials that were further processed in northern factories. Therefore the value adding process mostly took place in this area.[9] Moreover, 60 percent of the traffic was on the northern network, and also within this area we observe some sluggishness in the distribution of productivity over time (figure 4i and 4l). Finally, since northern provinces were connected by rail with more neighboring countries than the South, this enhanced the productivity of the North with respect to those of the South, increasing the gap between the two areas.

We also find some interesting interactions between railways and other transportation infrastructures. First, we find that national and provincial roads, on the one hand, and municipal roads, on the other hand, although less significant, are positive. Second, the interaction term between railways and municipal roads is positive but little significant, claiming that it is possible that these two types of infrastructure tend to reciprocally strengthen their effect on productivity. Conversely, the interaction term of railways with national/provincial roads is negative. For the interpretation of the partial effect we refer, for example, to model (6) which highlights that the most important variables in explaining productivity are provincial/national and municipal roads, railways, the relevant interaction terms and the interaction term between railways and the dummy concerning their early construction.

The partial derivative of GVA/male person with respect to municipal roads is:

$$\frac{\partial \log(GVA/MPOP)}{\partial \log(1+RAIL)} = 5.424 - 19.690 \log(1+NAT\_ROADS) \tag{7}$$

---

[9] Openness to trade should not be very high in order to call an area integrated. In 2013 the US exported 13.5% of GDP (World Bank Economic Indicators), yet it is extremely integrated in world trade, with the Silicon Valley setting the pace of technological innovation to the rest of the world, and low-skilled workers suffering from cheap imports from China (Freeman, 1995). The EU is another example (De Grauwe, 2014): in 2007 large countries such as Germany, France and Italy shipped to other EU members 25.9%, 14.0% and 13.9% of their overall exports, respectively. Smaller countries have much higher percentages, but all of them are largely integrated in the European single market.

If we want to know what is the level of communal roads that makes positive the impact of railways we have to make the previous equation equal to zero:

$$\begin{aligned}&5.424 - 19.690 \log(1 + NAT\_ROADS) \geq 0 \Leftrightarrow 0.275 \geq \log(1 + NAT\_ROADS)\\&\Leftrightarrow \exp(0.275) \geq 1 + NAT\_ROADS \Leftrightarrow 0.316 \geq NAT\_ROADS\end{aligned} \tag{8}$$

The density of national and provincial roads needs to be less than 0.316 km per squared km in order to have a positive impact on railways. Until 1901 only one province was above this threshold, and only 6 provinces in 1911. There is probably a substitution effect: national roads were cheaper because it was not required the payment of a ticket to access, but they were slower. A higher density of these roads lowers the use of railways and their positive impact on productivity.

In contrast with the previous results on roads, the distance from the nearest main port is negative as expected, but the statistical significance is very limited. The explanation can be related to the phases of Italian industrialization: initially, around 1870, production was essentially devoted to local consumption, and then ports were not of great importance for trade. Subsequently, as long as production increased, the problem of distance between markets and from ports has been overcome by improvements of transport infrastructure, mainly railways. Agglomeration is significant and positive, which is related with the importance of the domestic market as well as the presence of a larger labor market.

Finally, in each regression we have different sets of selected eigenvectors (out of the 39 candidates). This is due to the combination of variables selected for each model, and their capacity to fit the data, that implies a different combination of additional regressors (the eigenvectors) that surrogate the missing explanatory independent variables. Ideally, the eigenvectors can be associated with a geographical scale. In our case we have differentiated eigenvectors for each estimate and we do not have a clear prevalence of a determined set of eigenvectors. This indicates that the explanatory variables do not accounted in the analysis have no clear geographical scale. The (unconsidered) factors that affect productivity, then, do not have an explicit relation to the geography but have a stronger relation with the endogenous provincial characteristics. This is made clear in Appendix 1 where we plot the eigenvectors of Model (6) taken as reference.

7. Robustness checks

This section assesses the robustness of the results for the effect of public capital presented above. We check the sensitivity of the estimates to alternative definitions of the spatial weights matrix. Our analysis, so far, considered as a contiguity criterion a mix of geographical distance and economic characteristics. The possibility other types of proximities is investigated in order to check whether the results obtained are conditional to the use of the contiguity.

The first alternative definition is purely aspatial and is based on the idea that the more similar the economies of two regions are, the greater their weights. Hence, following Moreno and Lopez-Bazo (2007) we will use population density as a rough proxy for agglomeration economies. The weights of the **W** matrix are constructed as follows:

$$w_{ij} = \frac{1}{\left| POP\_DENS_i - POP\_DENS_j \right|} \qquad (9)$$

The resulting weights matrix accounts only for similarity in population density between each pair of regions, irrespective of their proximity. The idea behind this specification is that more similar regions stiffly compete for mobile factors of production.

The results in table 4 show that the signs and significance of the parameters is very close to previous estimation. The main difference is that the parameter associated to the distance from the main ports is not significant in any model.

The second specification of matrix **W** is based on geographical criterion. We define two regions as neighbor if they share their respective boundaries for at least a point. In this case we consider the interrelation only caused by geographical proximity. Table 5 confirms our previous results in terms of significant variables and signs.

[Table 4 and 5 about here]

8. Conclusions

This paper provided the first empirical evidence based on spatial econometrics on the effects of railway construction in post-Unification Italy.

The spatial concentration pattern of the variables over time is not stable. Railways are little clustered in 1871, but they are not in the following periods, highlighting that they are widespread in the peninsula. Provincial/national and municipal roads follow a similar pattern. In contrast, the concentration of industrial productivity in the northern regions increase dramatically with 19 provinces that belong to the cluster high-high in 1871 and only 13 in 1911. Of these 13 provinces, 11 belong to the cluster with high level of rail infrastructure in 1871, gaining competitiveness by the additional rail construction. This is consistent with the NEG which claims that the development of transport infrastructure, by increasing the accessibility of weaker regions, “not only gives firms in less developed regions better access to inputs and markets of more developed regions [. . .] but it also makes it easier for firms in richer regions to supply poorer regions at a distance, and can thus harm the industrialization prospects of less developed areas” (Puga 2002: 396). Our results add some empirical evidence to Martin and Rogers (1995), highlighting that public infrastructure facilitates transactions inside the cluster of regions that built railways as first at the expenses of provinces that did not experience this early infrastructure endowment.

The situation in Italy immediately after the unification was not so strongly polarized in terms of productivity, but it was in terms of infrastructure and natural resources endowments. The effort of the Kingdom to provide a balanced infrastructure level increased industrial productivity across the board but the combined advantage of water availability and of early construction of railways made that some provinces more developed than others. Therefore, the small initial gap in industrial productivity persisted and mostly increased over time.

References

Aschauer D.A. (1989). Is public expenditure productive?, *Journal of Monetary Economics*, 23: 177-200.

Andersson M., Karlsson C. (2007). Knowledge in regional economic growth. The role of knowledge accessibility, *Industry and Innovation*, 14: 129-149.

Annuario Statistico Italiano (1892, 1900, 1912). Roma.

Atack J., Margo, R.A. (2011). The impact of access to rail transportation on agricultural improvement. The American Midwest as a test case, 1850–1860, *The Journal of Transportation and Land Use*, 4: 5–18.

Atack J., Bateman F., Haines M. and Margo R.A. (2010). Did railroads induce or follow economic growth? Urbanization and population growth in the American Midwest, 1850–1860, *Social Science History*, 34: 171-197.

Barro R., (1990). Government spending in a simple model of endogenous growth, *Journal of Political Economy*, 98: 103-125.

Bogart D. (2014). The transport revolution in industrializing Britain: A survey, in R. Floud, J. Humphries and P Johnson (eds.) *The Cambridge Economic History of Britain 1700 to 1870*, Cambridge University Press, 368-390.

Bezza B. (1986). *Energia e sviluppo. L'industria elettrica italiana e la Società Edison*. Einaudi, Torino

Ciccarelli C., Fenoaltea, S. (2009). *La produzione industriale delle regioni d'Italia, 1861-1913: una ricostruzione quantitativa. 1. Le industrie non manifatturiere*. Roma: Banca d'Italia,

Ciccarelli C., Fenoaltea, S. (2014). *La produzione industriale delle regioni d'Italia, 1861-1913: una ricostruzione quantitativa, 2. Le industrie estrattivo-manifatturiere*. Roma: Banca d'Italia.

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

Del Bo C., Florio M. (2012). Infrastructure and growth in a spatial framework: Evidence from the EU regions, *European Planning Studies*, 20: 1393-1414.

De Grauwe P. (2014). *Economics of Monetary Union*, Oxford University Press.

Donaldson D., Hornbeck R. (2016). Railroads and American economic growth: A “market access” approach, *Quarterly Journal of Economics*, forthcoming

ESPON (2007). Update of selected potential accessibility indicators (Final report). Dortmund, Spiekermann & Wegener.

Federico G. (2005). Seta, agricoltura e sviluppo economico in Italia, *Rivista di Storia Economica*, 21, 117-149.

Felice, E. (2011). Regional value added in Italy, 1891-2001, and the foundation of a long-term picture, *Economic History Review*, 64, 929-950.

Felice, E. (2013). *Perché il Sud è rimasto indietro*, Bologna, Il Mulino.

Fenoaltea S. (1983). Italy. In: P. O’Brien (ed.), *Railways and the Economic Development of Europe.* Oxford: Macmillan.

Fenoaltea S. (2011). *The Reinterpretation of Italian Economic History*, Cambridge MA.

Freeman R.B. (1995). Are your wages set in Beijing?, *Journal of Economic Perspectives*, 9, 15-32.

Gerschenkron A. (1962). *Economic backwardness in historical perspective*, Cambridge, Massachusetts: Belknap Press of Harvard University Press.

Getis A., Griffith D.A. (2002). Comparative spatial filtering in regression analysis, *Geographical Analysis*, 34: 130-140.

Gregory, T. Patuelli R. (2015). Demographic ageing and the polarization of regions. An exploratory space-time analysis, *Environment and Planning A*, 47: 1192-210.

Griffith D.A. (2003). *Spatial autocorrelation and spatial filtering: gaining understanding through theory and scientific visualization*, Springer, Berlin.

ISTAT (2009). Atlante Statistico dei Comuni, Roma.

Krugman P. (1991). Increasing returns and economic geography, *Journal of Political Economy*, 99: 483-499.

Krugman P., Venables A. (1995). Globalization and the inequality of nations, *Quarterly Journal of Economics*, 110: 857-880.

Holtz-Eakin D., Schwartz A.E. (1995). Infrastructure in a structural model of economic growth *Regional Science and Urban Economics*, 25: 131-151.

L'Italia Economica (1873). Roma

Le Gallo J., Ertur C. (2003). Exploratory spatial data analysis of the distribution of regional per capita GDP in Europe, 1980-1995, *Papers in Regional Science*, 82: 175-201.

Lewis B.D. (1998). The impact of public infrastructure on municipal economic development: Empirical results from Kenya, *Review of Urban and Regional Development Studies* 10: 142-155.

Martin P., Rogers A.R. (1995). Industrial location and public infrastructure, *Journal of International Economics*, 39: 335-351.

McMillen D.P., (2003). Spatial autocorrelation or model misspecification?, *International Regional Science Review*, 26: 208-217.

Ministero di agricoltura, industria e commercio (1875). *Statistica del Regno d'Italia. Censimento 31 dicembre 1871*, vol. 2, Roma.

Ministero di agricoltura, industria e commercio (1883). *Censimento della popolazione del Regno d'Italia al 31 dicembre 1881*, Roma.

Ministero di agricoltura, industria e commercio (1904). *Censimento della popolazione del Regno d'Italia al 10 febbraio 1901*, Roma.

Ministero di agricoltura, industria e commercio (1914). *Censimento degli opifici e delle imprese industriali al 10 giugno 1911*, Roma.

Ministero delle comunicazioni, Ferrovie dello Stato (1927). *Sviluppo delle ferrovie italiane dal 1839 al 31 dicembre 1926*, Roma.

Moreno R., López-Bazo E. (2007). Returns to local and transport infrastructure under regional spillovers, *International Regional Science Review*, 30, 1: 47-71.

Patuelli R., Griffith D., Tiefelsdorf M., Nijkamp P. (2011). Spatial filtering and eigenvector stability: space-time models for German unemployment data, *International Regional Science Review*, 34: 253-280.

Puga D. (2002). European regional policies in the light of recent location theories, *Journal of Economic Geography* 2: 373-406.

Romeo R. (1959). *Risorgimento e capitalismo*, Laterza, Bari.

Russel K. (1985). *The Industrial Geography of Italy*, Palgrave Macmillan.

Schram A. (1997). *Railways and the Formation of the Italian State in the Nineteenth Century*, Cambridge University Press.

Sereni E. (1966). *Capitalismo e mercato nazionale in Italia*, Roma, Editori riuniti.

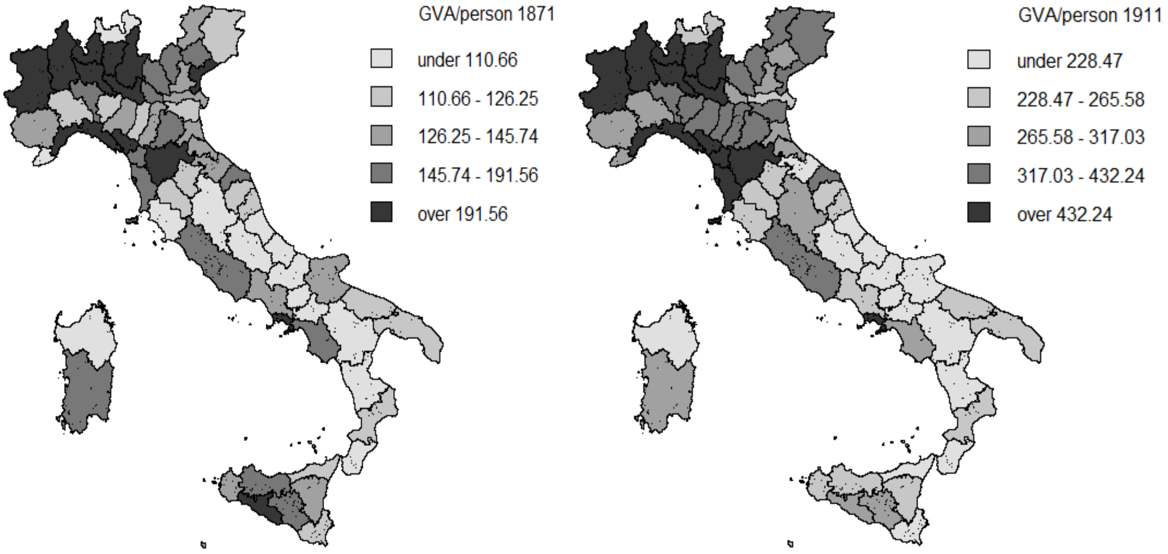
Tiefelsdorf M., Boots B., (1995). The exact distribution of Moran's I, *Environment and Planning A*, 27: 985-999.

Wegener M., Eskelinnen H., Fürst F., Schürmann C., Spiekermann K. (2002). Criteria for the spatial differentiation of the EU territory: Geographical position. Forschungen 102.2, Bonn: Bundesamt für Bauwesen und Raumordnung.

Figure 1 – Railways in 1861

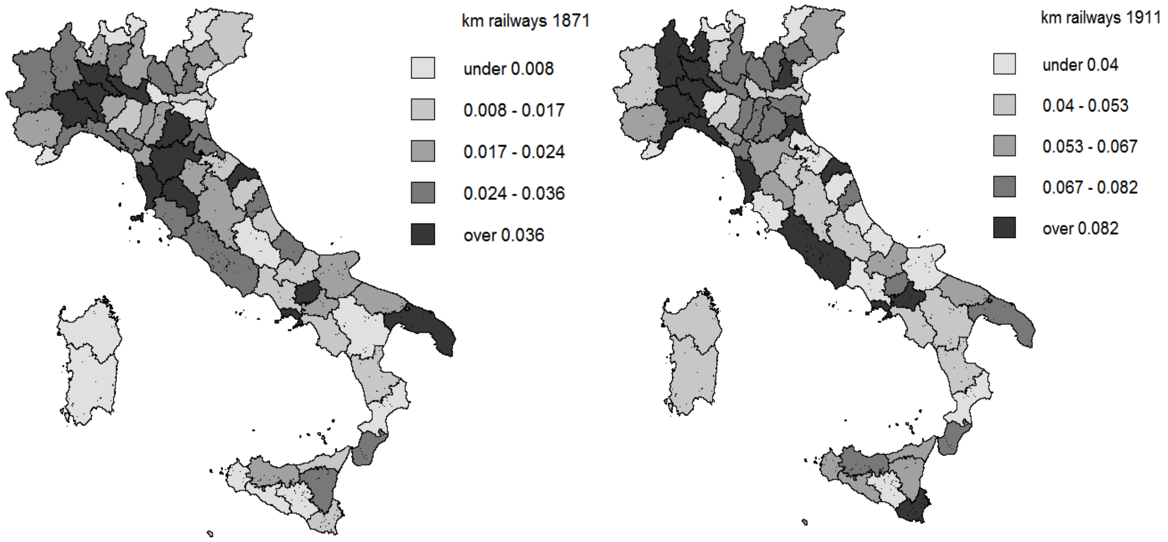
Figure 2 – Railways in 1870

Figure 3 – Railways in 1885



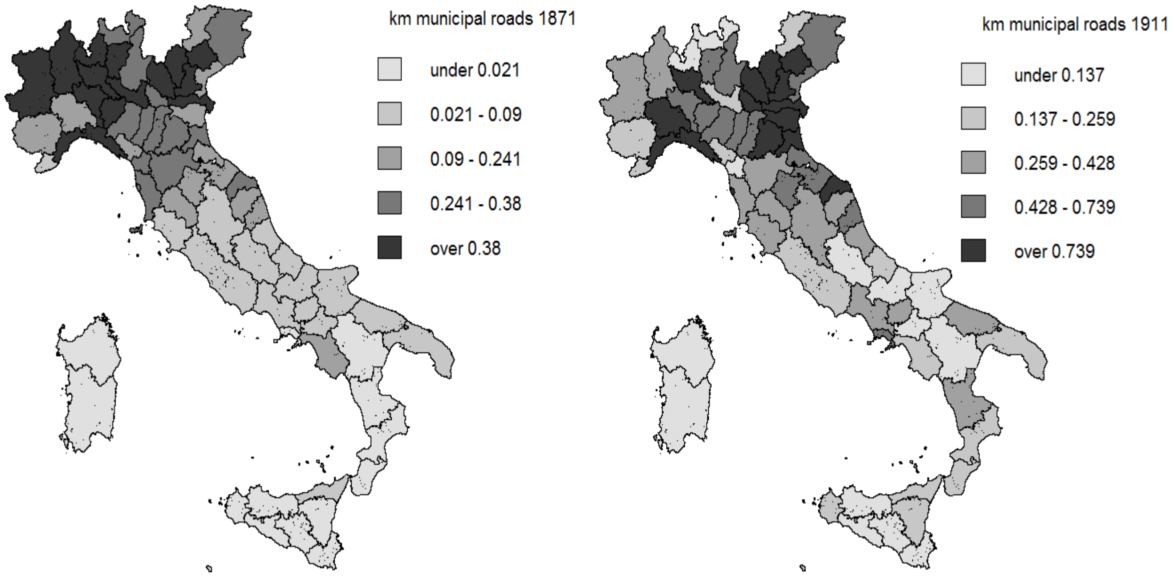
a) 1871 b) 1911

Figure 4 – Gross value added/male population in 1871 and 1911



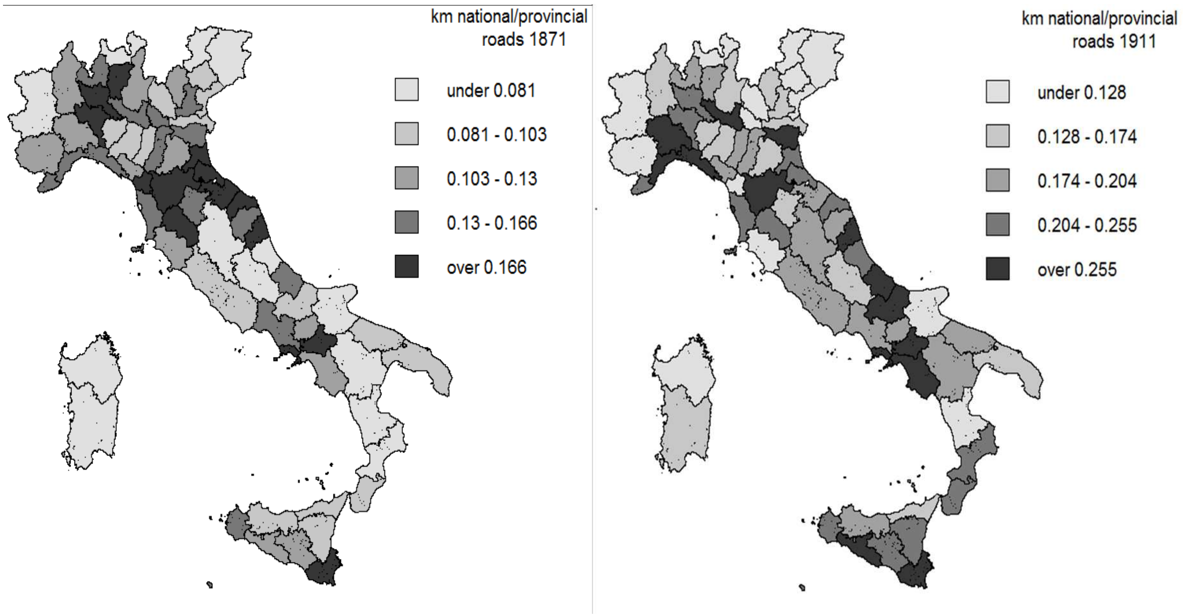
a) km railways/area 1871

b) km railways/area 1911

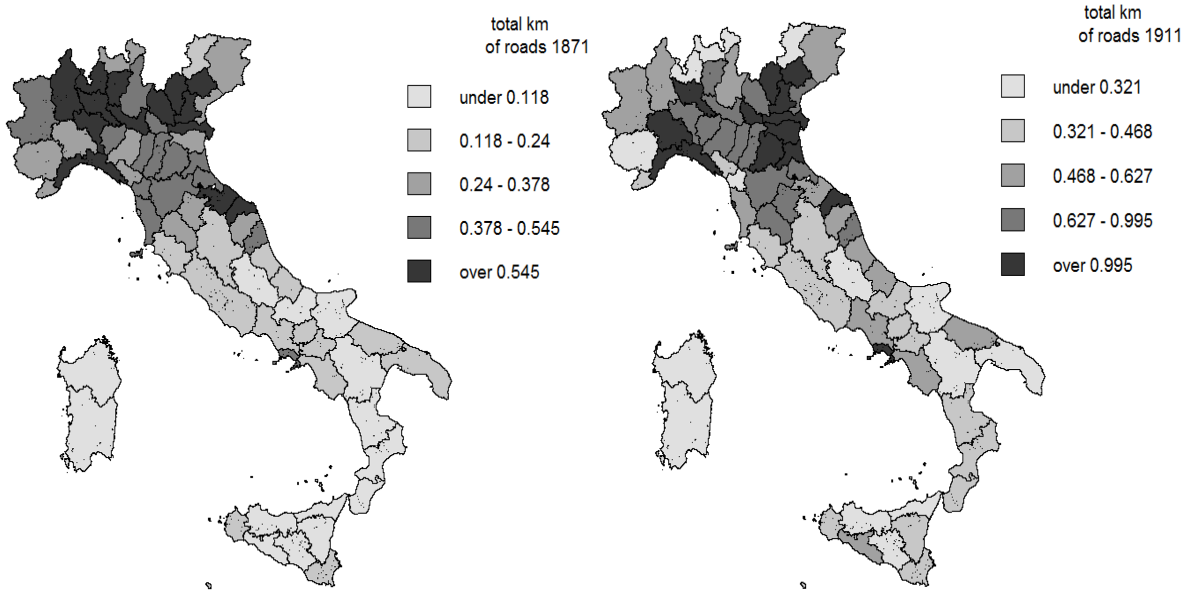


c) km communal roads/area 1871

d) km communal roads/area 1911



e) km provincial and national roads/area 1871 f) km provincial and national roads/area 1911



g) km total roads/area 1871 h) km total roads/area 1911

Figure 5 - Quintile map of infrastructural level in 1871 and 1911

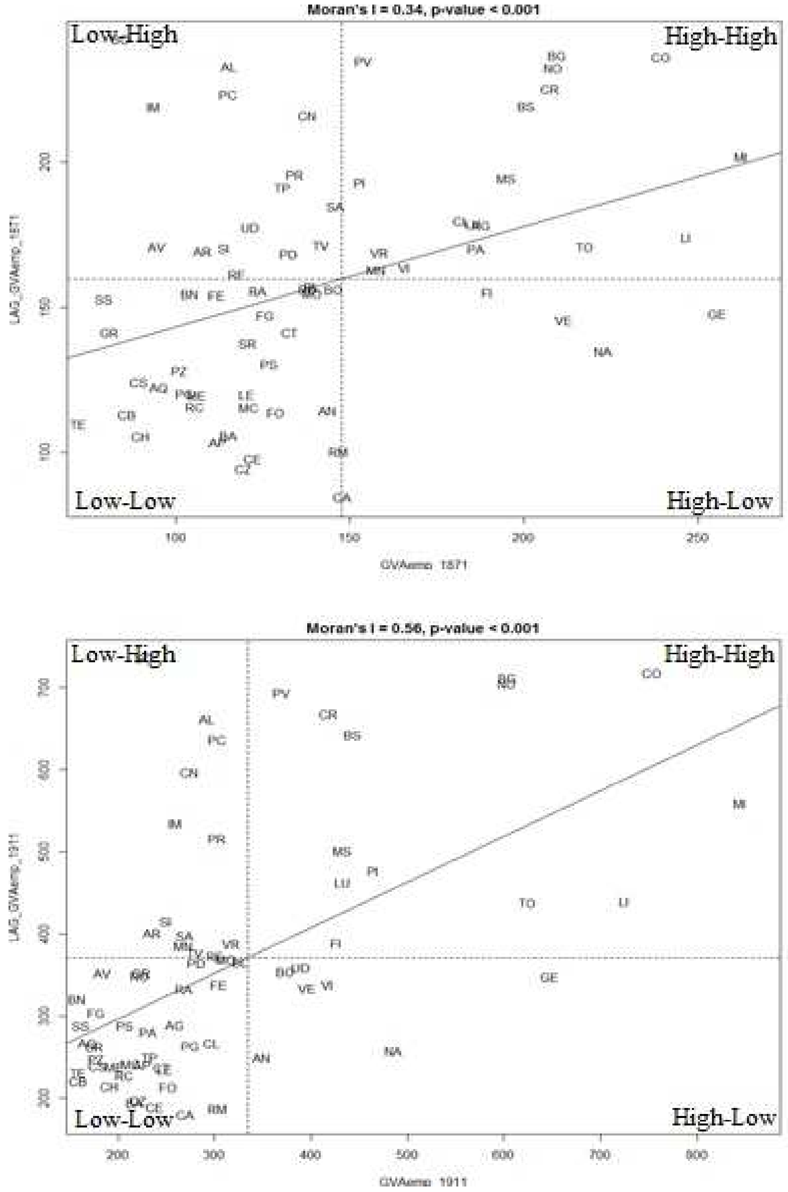


Figure 6: Moran's I scatterplot for GVA/male population in 1871 and 1911

Table 1: descriptive statistics

Variables	Mean	Std. dev.	Min.	Max.	Moran’s I
GVA/male pop. 1871	147.700	46.094	76.470	266.40	0.345 (<0.001)
GVA/male pop. 1881	167.800	56.021	86.360	322.400	0.313 (<0.001)
GVA/male pop. 1901	232.200	98.745	111.400	576.500	0.459 (<0.001)
GVA/male pop. 1911	334.700	150.664	174.300	858.900	0.556 (<0.001)
Literacy 1871	0.363	0.191	0.100	0.820	0.725 (<0.001)
Literacy 1881	0.457	0.216	0.140	0.900	0.731 (<0.001)
Literacy 1901	0.568	0.215	0.210	0.940	0.784 (<0.001)
Literacy 1911	0.663	0.200	0.310	0.960	0.788 (<0.001)
Railways 1839-1871	0.021	0.023	0.000	0.132	0.140 (0.035)
Railways 1839-1881	0.035	0.025	0.000	0.132	0.040 (0.265)
Railways 1839-1901	0.059	0.036	0.000	0.269	0.043 (0.230)
Railways 1839-1911	0.066	0.040	0.000	0.307	0.072 (0.129)
Agglomeration 1871	122.500	124.960	22.800	999.700	0.030 (0.234)
Agglomeration 1881	130.100	136.935	24.380	1103.000	0.023 (0.267)
Agglomeration 1901	147.300	157.410	28.860	1269.000	0.000 (0.400)
Agglomeration 1911	159.000	180.141	31.110	1444.000	-0.017 (0.515)
Municipal roads 1871	0.230	0.227	0.000	0.947	0.803 (<0.001)
Municipal roads 1881	0.203	0.202	0.002	0.955	0.569 (<0.001)
Municipal roads 1901	0.230	0.165	0.000	0.908	0.562 (<0.001)
Municipal roads 1911	0.327	0.337	0.006	1.481	0.356 (<0.001)
National/provincial roads 1871	0.127	0.067	0.043	0.455	0.070 (0.154)
National/provincial roads 1881	0.137	0.054	0.048	0.407	0.132 (0.041)
National/provincial roads 1901	0.172	0.055	0.068	0.426	0.338 (<0.001)
National/provincial roads 1911	0.224	0.210	0.047	1.690	0.019 (0.297)

p-values in parentheses.

Table 2: Moran Scatterplot transition probabilities for GVA/male population (1871-1911)

		1911				Provinces by quadrant in 1871
		High-high	Low-high	Low-low	High-low	
1871	High-high	12 (63.2)	3 (15.8)	3 (15.8)	1 (5.3)	19 (27.5)
	Low-high	0 (0.0)	10 (71.4)	3 (21.4)	1 (7.1)	14 (20.3)
	Low-low	0 (0.0)	0 (0.0)	27 (96.4)	1 (3.6)	28 (40.6)
	High-low	1 (12.5)	0 (0.0)	2 (25.0)	5 (62.5)	8 (11.6)
	Provinces by quadrant in 1911	13 (18.9)	13 (18.9)	35 (50.7)	8 (11.6)	

In brackets percentage by row

Table 3: panel data estimation (mkt potential matrix)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	4.840 ***	4.725 ***	5.166 ***	5.112 ***	4.396 ***	3.885 ***	3.492 ***	5.141 ***
	(0.114)	(0.127)	(0.112)	(0.113)	(0.068)	(0.325)	(0.316)	(0.109)
Railways	7.920 ***	8.287 ***	6.277 ***	7.150 ***	12.470 ***	5.424 ***	11.787 ***	6.286 ***
	(0.692)	(0.707)	(1.223)	(1.251)	(1.071)	(1.235)	(1.096)	(1.219)
National/provincial roads	0.755 **	0.671 **	1.801 ***	1.996 ***	2.459 ***	1.737 ***	2.582 ***	1.912 ***
	(0.237)	(0.236)	(0.387)	(0.388)	(0.425)	(0.372)	(0.416)	(0.3844)
Municipal roads	0.653 ***	0.724 ***	-0.057 **	-0.124	0.500 **	-0.147	0.359	-0.125
	(0.135)	(0.133)	(0.305)	(0.210)	0.214)	(0.200)	(0.220)	(0.208)
Distance from ports	-0.039	-0.021						
	(0.023)	(0.025)						
Dummy Moran		0.072	-0.229 *	0.003				
		(0.088)	(0.130)	(0.084)				
Dummy South							-0.081	0.015 *
							(0.088)	(0.082)
Agglomeration						0.272 ***	0.214 ***	
						(0.067)	(0.069)	
National/provincial roads × Railways			-20.635 ***	-22.592 ***	-28.823 ***	-19.690 ***	-30.736 ***	-21.878 ***
			(5.118)	(5.163)	(5.624)	(4.935)	(5.497)	(5.074)
Municipal roads × Railways			7.603 ***	8.704 ***	3.765	7.753 ***	5.443 *	9.208 ***
			(2.561)	(2.613)	(2.814)	(2.138)	(2.742)	(2.527)
Dummy Moran × Railways			4.871 ***		2.561 ***	2.138 *		
			(1.653)		(1.226)	(1.096)		
Alphabetization			0.454 ***	0.434 ***		0.435 ***		0.485 ***
			(0.056)	(0.054)		(0.053)		(0.057)
Lowland				-0.004		-0.041 **		0.005
				(0.016)		(0.018)		(0.014)
Lowland × Railways				0.110		0.065		
				(0.145)		(0.137)		
Common eigenv.	E1, E12, E17, E22, E32, E37	E2, E12, E21, E32, E33, E36	E1, E13, E16, E19, E34	E13, E16, E32, E34	E32	E17, E21, E36	E7, E12, E32	E19, E32
AIC	121.411	135.692	30.090	48.712	78.617	22.400	90.221	39.970

*Significant at 1%, ** significant at 5%, *** significant at 10%. Standard errors in parentheses.

Table 4: panel data estimation (density)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	4.732 ***	4.714 ***	5.184 ***	5.153 ***	4.400 ***	4.281 ***	3.575 ***	5.145 ***
	(0.114)	(0.124)	(0.107)	(0.109)	(0.068)	(0.323)	(0.316)	(0.107)
Railways	8.149 ***	8.082 ***	6.163 ***	6.803 ***	12.445 ***	6.016 ***	11.804 ***	6.330 ***
	(0.692)	(0.700)	(1.179)	(0.122)	(1.076)	(1.232)	(1.106)	(1.203)
National/provincial roads	0.58 **	0.612 **	1.808 ***	2.018 ***	2.518 ***	1.900 ***	2.518 ***	1.954 ***
	(0.233)	(0.235)	(0.379)	(0.381)	(0.443)	(0.390)	(0.434)	(0.380)
Municipal roads	0.799 ***	0.733 ***	-0.101	-0.221	0.483 **	-0.114	0.355	-0.170
	(0.128)	(0.132)	(0.203)	(0.210)	(0.215)	(0.204)	(0.218)	(0.207)
Distance from ports	-0.018	-0.015						
	(0.023)	(0.024)						
Dummy Moran		0.081	-0.255 **	-0.0003				
		(0.085)	(0.120)	(0.078)				
Dummy South							-0.084	0.158 **
							(0.856)	(0.078)
Agglomeration						0.186 ***	0.198 ***	
						(0.068)	(0.070)	
National/provincial roads × Railways			-19.931 ***	-22.140 ***	-30.008 **	-21.887 ***	-29.474 ***	-22.157 ***
			(5.026)	(5.042)	(6.072)	(5.339)	(5.931)	(5.015)
Municipal roads × Railways			7.551 ***	9.147 ***	4.175	7.989 ***	4.993 *	9.455 ***
			(2.551)	(2.602)	(2.890)	(2.635)	(2.805)	(2.513)
Dummy Moran × Railways			4.558 ***		2.378 *	1.745		
			(1.613)		(1.221)	(1.104)		
Literacy			0.453 ***	0.444 ***		0.429 ***		0.481 ***
			(0.052)	(0.052)		(0.052)		(0.056)
Lowland				0.004		0.003		0.014
				(0.015)		(0.017)		(0.014)
Lowland × Railways				0.112		0.068		
				(0.145)		(0.140)		
Common eigenvalues	E8, E17	E17, E23, E35	E1, E14, E17	E1, E7 E11, E14	E1, E7, E8, E17, E28	E5, E6, E8, E14, E17, E22, E28	E17, E23, E28	E11, E17
AIC	127.641	114.767	19.217	41.552	82.423	26.982	83.719	33.046

*Significant at 1%, ** significant at 5%, *** significant at 10%. Standard errors in parentheses.

Table 5: panel data estimation (queen)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intercept	4.709 ***	4.695 ***	5.153 ***	5.110 ***	4.402 ***	4.146 ***	3.704 ***	5.057 ***
	(0.445)	(0.125)	(0.107)	(0.116)	(0.067)	(0.332)	(0.282)	(0.119)
Railways	8.152 ***	8.12 ***	6.521 ***	7.144 ***	12.393 ***	5.324 ***	11.876 ***	7.021 ***
	(1.445)	(0.701)	(1.182)	(1.282)	(1.071)	(1.306)	(1.085)	(1.256)
National/provincial roads	0.644 **	0.653 **	1.851 ***	2.014 ***	2.441 ***	1.726 ***	2.632 ***	2.047 ***
	(2.445)	(0.235)	(0.380)	(0.392)	(0.423)	(0.388)	(0.408)	(0.393)
Municipal roads	0.747 ***	0.727 ***	-0.059	-0.125	0.501 **	-0.092	0.330	-0.119
	(3.445)	(0.132)	(0.203)	(0.209)	(0.213)	(0.204)	(0.217)	(0.207)
Distance from ports	-0.012	-0.012						
	(4.445)	(0.025)						
Dummy Moran		0.086	-0.304 **	0.017				
		(0.086)	(0.124)	(0.086)				
Dummy South							-0.075	0.167
							(0.077)	(0.118)
Agglomeration						0.220 ***	0.167 ***	
						(0.068)	(0.062)	
National/provincial roads × Railways			-20.866 ***	-23.210 ***	-28.374 ***	-19.961 ***	-31.231 ***	-23.607 ***
			(5.062)	(5.160)	(5.608)	(5.099)	(5.398)	(5.143)
Municipal roads × Railways			7.480 ***	9.103 ***	3.737	7.860 ***	5.734 **	9.376 ***
			(2.552)	(2.600)	(2.809)	(2.556)	(2.711)	(2.520)
Dummy Moran × Railways			4.690 ***		2.335 *	2.629 **		
			(1.612)		(1.211)	(1.221)		
Literacy			0.432 ***	0.438 ***		0.457 ***		0.439 ***
			(0.052)	(0.058)		(0.059)		(0.060)
Lowland				-0.003		-0.020		0.005
				(0.015)		(0.017)		(0.013)
Lowland × Railways				0.104		0.056		
				(0.145)		(0.138)		
Common eigenvalues	E33, E42	E33	E9, E16, E33, E42	E2, E12, E16, E33, E35, E42, E49	E33, E42	E2, E9, E12, E33, E35	E9, E33, E42, E45, E49	E2, E7, E9,E33, E35, E37, E42, E49
AIC	123.095	127.192	21.222	47.785	76.798	38.697	76.988	48.413

*Significant at 1%, ** significant at 5%, *** significant at 10%. Standard errors in parentheses.

Appendix 1: selected eigenvectors of model (6), table 3