Industrial development in the Italian regions, 1861-1913: new evidence

Carlo Ciccarelli
University of Rome "Tor Vergata"

Francesca Di Iorio
University of Naples Federico II

Stefano Fachin
"Sapienza" University of Rome

Abstract

This paper studies the growth of manufacturing industrial value added in the Italian regions from 1861 to 1913 using a recently released dataset at annual frequency and disaggregation in ten industries. Estimation of an approximate factor model shows long-run growth is essentially explained by two non-stationary factors, an essentially monotonous trend and a very long cycle. The loadings of the trend factor are highest in the North-West and lowest in the South. The residuals of the factor model are then studied using spatial autoregressive panel models, which suggest that spatial spillovers were significant and essentially similar in all industries.

Keywords: Italy, industrial 19th century growth, approximate factor model, spatial error model, bootstrap.

JEL codes: C38, C31, N13, N63, N93

May 2015

1 Introduction[1]

The second half of the 19th century is conventionally consider by economists (*e.g.*, Kuznets, 1971, Maddison, 1995, Pritchett, 1997) as a nodal point for advanced capitalist countries. For instance, the year 1870 is used by the *Cambridge Economic History* to distinguish two phases (1700-1870, and 1870 to present day) of European economic development. Study of this period at the national level are abundant, with a classical example given by Gerschenkron's (1962) analysis of patterns of industrial growth and development in backward countries. However, these studies do not typically account for economic heterogeneity at the regional and industry level. This might be a serious omission: it certainly is, for instance, in the case of the Austrian-Hungarian Empire (where the industry was mostly concentrated in the western regions, while in the eastern ones agriculture prevailed), and of Italy (where the differences between North and South were large under essentially all respects). It is then no surprise that historical studies at the regional level represent an important part of the economic history literature, with statistical reconstructions of regional GDP of the past representing an active field of research (for a recent example, see, *e.g.*, Geary and Stark, 2015). However, this stream of literature suffers from two limitations. First, typically only estimates for selected benchmark years (usually at ten-years intervals, as, *e.g.*, in Henning *et al.,* 2011, A'Hearn and Venables, 2013, Felice, 2011) are reported, so that detailed studies of the time series properties of the series are not possible. Second, by focusing on GDP or its main sectoral components (agriculture, industry, and services) it may fail to detect the origin of economic and structural changes characterizing the transition to modern economic growth[2]. The bottom line is that in order to reach a satisfactory understanding of the processes which drove the 19th century development of the industrial economies we would ideally need annual time series of data disaggregated at both the spatial and industrial level. Of course, the problem is that data with this level of detail are generally not available for the 19th century. An exception is Italy, for which, as a result of a long-term project sponsored by the Bank of Italy, Ciccarelli and Fenoaltea (2009, 2014, forthcoming) constructed annual time series of value added at 1911 prices for 12 manufacturing industries and 16 regions for the period 1861-1913. During these five decades Italian manufacturing value added grew significantly in real terms. Taking as reference periods the two census years 1871 and 1911, the average annual rate of growth has been 2.6 percent; as a consequence, 1911 value added was nearly three times that of forty years before. Contributions to aggregate growth varied widely, with the traditional industries producing consumption goods generally growing more slowly than those essentially tied to the production of investment and intermediate goods. The extremes of the ranking are on one side the stagnating Leather industry, with an average growth rate of just 0.3 percent, on the other the Metalmaking industry, which grew annualy at more than a 10 percent rate. Wide heterogeneity is also present among regions, with manufacturing valued added in Liguria (in

[1] Correspondance to carlo.ciccarelli@uniroma2.it, stefano.fachin@uniroma1.it. Research supported by MIUR PRIN grant 2010J3LZEN "Forecasting economic and financial time series: understanding the complexity and modelling structural change". We would like to thank Marco Lippi for comments and suggestions. The usual disclaimers apply. All computations have been carried out using programs written in Hansl, the programming language of the free software package Gretl (http://gretl.sourceforge.net/).

[2] An example, borrowed from McCloskey (1991, pp. 99-100), illustrates. In a two-sector economy the traditional sector (agriculture) grows constantly at the 1% annual rate, with an initial share of GDP of 90%. The modern sector (industry) grows constantly at the 4% annual rate, with an initial share of GDP of 10%. The growth rate of GDP, initially very close to 1%, will asymptotically reach the 4% rate following an S-shaped path seemingly subject to structural breaks. Obviously, this just a composition effect, as the two sectors have a constant growth rate and no structural change is actually taking place in the economy.

the North-West; a map is reported in the Appendix) expanding three times as fast as Basilicata, in the South (respectively, 4.6 and 1.5 percent). Our aim is thus investigating the development of the Italian manufacturing industry in the in the second half of the 19th century going into industry and regional detail, a task which could be not undertaken with standard datasets including data at national level or for main GDP components. Unfortunately, the lack of similar estimates for variables like capital and labour will prevent us from providing a structural answer. However, using the tools provided by the approximate factor models literature we shall be able to extract the latent trends present in the dataset: in other terms, we shall identify a small number of patterns able to explain the evolution of the various industries in the different regions. To understand the impact of interaction between regions we shall then estimate spatial autoregressive models. We now (section 2) proceed to a descripitive analysis of the data, then describe the modelling set-up (section 3) and the estimates obtained (section 4). Section 5 concludes, while some details on the dataset are reported in the Appendix.

2 Industrial growth in the Italian regions, 1861-1913

The descriptive analysis of the data is in our case not a simple task, as with only one exception[3] our dataset covers N =10 manufacturing industries[4] in R =16 regions, so that we have 159 industry/region combinations. As a first step, Table 1 and Fig. 1 respectively report average compound growth rates of these series and their distribution. The latter is almost symmetrical, with the mean only slightly higher than the median (respectively, 2.6 and 2.2 percent) and a very small number (five) of negative values identifying shrinking industry/region pairs. The high dispersion (minimum -2.6, maximum 8.4, interquartile range 2.2) suggests that even taking such a simplified view there is a wide heterogeneity to be explained. This is of course bound to be much greater if the entire trajectories are taken into account. We can go into some detail examining the marginal row and columns of Table 1, which report simple averages[5] for each region and industry.

Starting with the regional values, both the leading role of the North-West (the first three regions, Piedmont, Liguria, Lombardy[6]) and the formation of the North-South divide are clearly visible. The North-West average, 3.4, is much higher than both that of the North-East (Emilia, Venetia) and of the Centre (from Tuscany to Latium), respectively 2.6 and 2.7 percent. These, in turn, are clearly higher than that of the South, 2.1 percent.

Taking the industry perspective we can see that the industries with the highest growth rates (Metalmaking, Non metallic minerals, Chemicals and Paper, with growth rates between 3.5 and 4.6) produce all intermediate and investment goods. All the industries at the other extreme of the ranking (Food, Textiles, Wood, with rates between 1.2 and 1.5, two or three times as smaller) produce consumption goods. The three remaining industries, all with average growth rates around 2 percent, produce in two cases consumption goods (Clothing, Leather) and in the other investment goods (Engineering).

[3] Metalmaking, virtually absent in Basilicata, a small southern region.

[4] We excluded from the analysis Tobacco and Sundry industries. The former was a very small industry (on the average over the period less than 1% of total manufacturing value added) run as a state monopoly, hence with peculiar localisation and growth patterns. The latter was even smaller, on the average about 0.5%, and far too an heterogenous aggregate to be of any interest.

[5] Giving equal weight to all values, these reflect the underlying heterogeneity better than average growth rates of the regional and national industry aggregates.

[6] About a century later, since the 1950's, the area with a triangular shape with vertices in the main cities of these three regions (Turin, Genoa and Milan) will be commonly referred to as "the industrial triangle".

Table 1

Average Value Added growth rates, 1861-1913

	Food	Text	Cloth	Leath	Wood	Metal	Eng	Non Met Min Prods	Chem	Paper	*Average*
Piedmont	1.33	**3.19**	**2.55**	1.69	1.37	**4.97**	**3.50**	**3.80**	**4.47**	**4.94**	*3.18*
Liguria	**2.35**	**2.79**	**2.74**	**2.09**	**1.90**	**6.20**	**4.20**	4.92	**4.89**	3.75	*3.58*
Lombardy	**1.51**	**3.48**	**2.35**	**1.95**	**1.92**	**4.80**	**3.89**	4.79	**4.97**	**5.12**	*3.48*
Venetia	1.20	**3.26**	1.41	1.58	1.56	**3.59**	1.97	**3.70**	**3.71**	3.92	*2.59*
Emilia	**2.03**	-0.28	**2.49**	1.84	**1.64**	**3.51**	**2.56**	**3.65**	**3.74**	4.47	*2.57*
Tuscany	1.18	**2.43**	1.75	1.88	**1.62**	**5.94**	**2.58**	3.26	**4.56**	**4.65**	*2.99*
Marches	0.96	0.32	1.61	1.80	1.33	2.86	1.51	**4.11**	**3.59**	**5.12**	*2.32*
Umbria	0.83	**2.62**	1.41	1.86	1.08	**8.36**	2.06	**4.39**	**7.34**	**4.85**	*3.48*
Latium	1.14	-0.9	**2.68**	**1.96**	1.28	1.04	**2.74**	2.46	3.26	**5.78**	*2.14*
Abruzzi	0.78	1.35	1.33	1.86	1.48	**7.40**	0.66	2.44	3.42	**4.94**	*2.57*
Campania	1.41	-0.61	1.60	**2.19**	1.44	**4.42**	**2.28**	1.81	2.67	3.97	*2.12*
Apulia	1.50	0.94	2.15	**2.26**	**2.20**	**2.89**	**2.24**	3.07	2.87	**5.88**	*2.60*
Basilicata	0.34	-0.08	0.96	1.62	0.92	-	0.16	1.79	0.20	3.30	*1.02*
Calabria	1.12	0.59	1.48	**2.05**	**1.77**	-2.63	0.67	**5.18**	1.74	4.34	*1.63*
Sicily	**1.41**	-2.15	**2.17**	**2.19**	**1.69**	**2.46**	1.93	**4.16**	2.41	4.41	*2.07*
Sardinia	**2.22**	**2.70**	**2.40**	**2.18**	**2.20**	**4.08**	1.73	3.09	1.53	4.17	*2.63*
Average	*1.33*	*1.23*	*1.94*	*1.92*	*1.59*	*3.99*	*2.17*	*3.54*	*3.46*	*4.60*	2.57

Abbreviations: Text, Textiles; Cloth, Clothing; Leath, Leather; Metal, Metalmaking; Eng, Engineering; Non Met Min Prods, Non Metallic Mineral Products; Chem, Chemicals.
In bold case: values greater than the national industry average (bottom row).
Source: see Appendix.

We now try to gain some understanding of the entire trajectories examining those followed by the aggregates over regions ($Y_{rt} = \sum_{i=1}^{N} Y_{irt}$, Fig.2; since for our purpose there is no need to enter into a detailed discussion of each region, in this figure individual labels are not provided to keep the presentation compact) and industries ($Y_{it} = \sum_{r=1}^{R} Y_{irt}$, Fig. 3 and 4). The key message conveyed by both plots is of a general positive trend with widely variable slopes, and in the case of the industry time series, some cyclical patterns also (mostly found in investment and intermediate goods industries).

The plots clearly support *a-priori* expectations of non-stationarity. Following routine practice, we tested this hypothesis using the ADF-GLS test by Elliot, Rothenberg and Stock (1996), allowing for a linear deterministic trend. The results are summarised in Fig. 5. As it is immediately seen, using customary significance values almost all statistics fall in the non-rejection region. More precisely, no statistic falls below the 1% critical value (the smallest statistic is -3.32, much higher than the 1% critical value, -3.77), and only two and five (1.3% and 3.1% of the total) respectively exceed the 5% and 10% critical values. In the light of this overwhelming evidence a proper panel unit root procedure may be considered not necessary to conclude the all the series of our dataset are non-stationary.

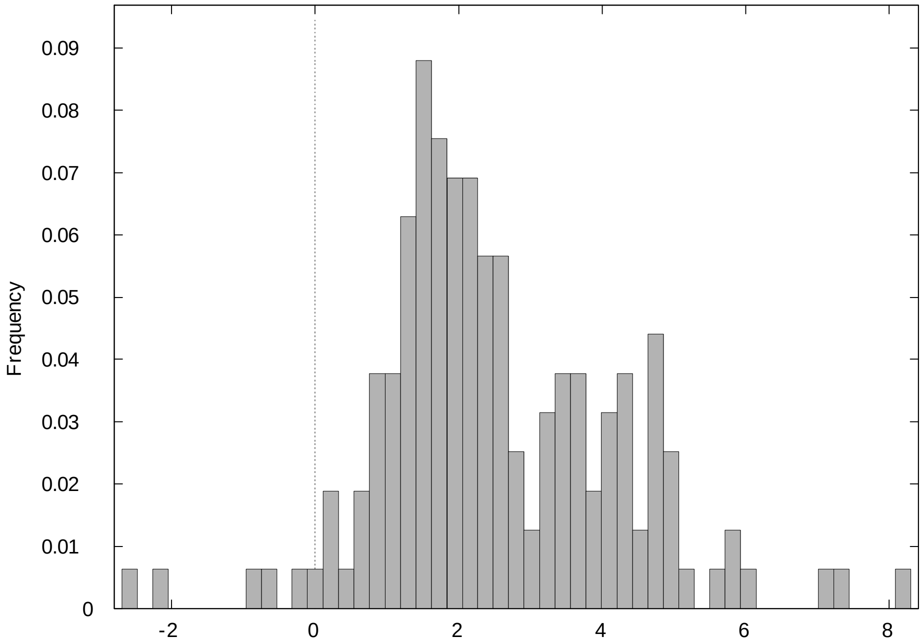

Figure 1: Distribution over industries and regions of average compound rates of growth, 1861-1913. *Source*: see Appendix.

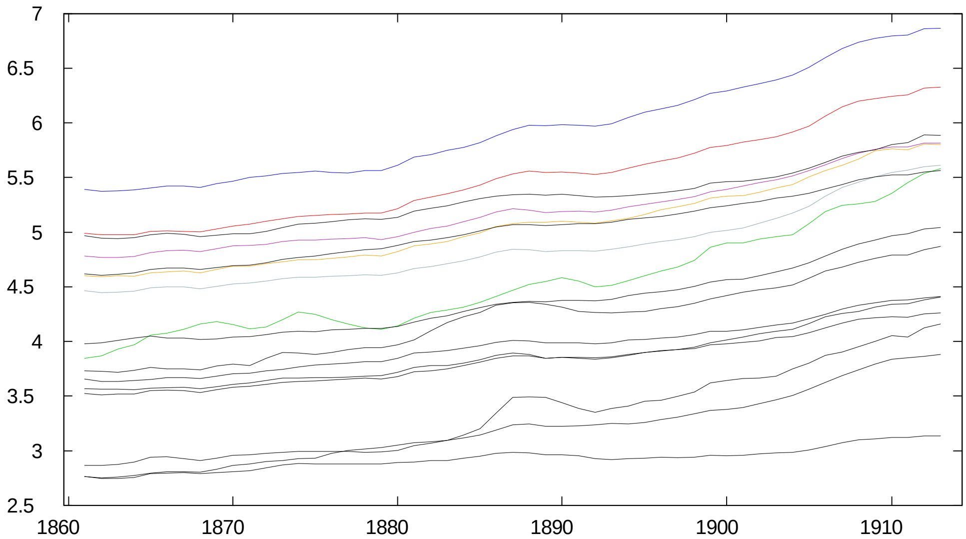

Figure 2: Time series of regional aggregates (logs). *Source*: see Appendix.

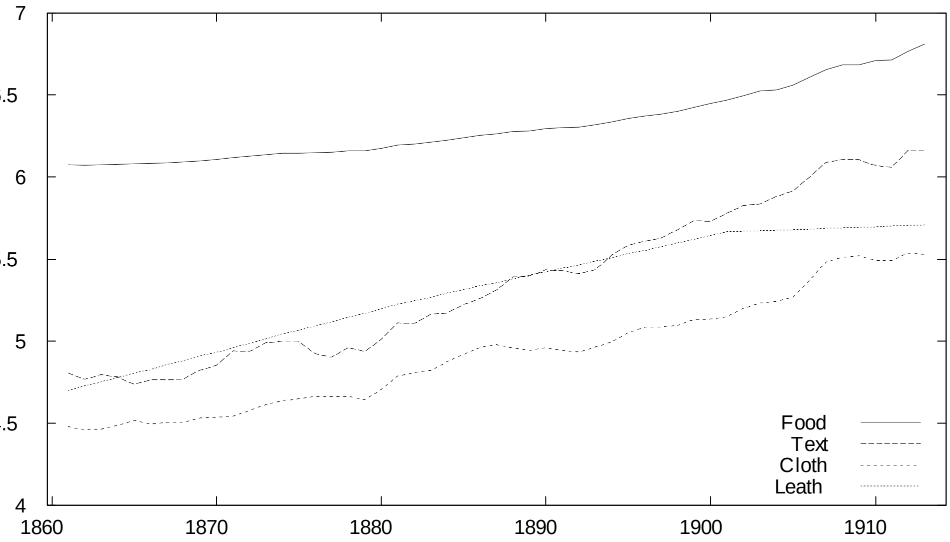

Figure 3: Time series of national industry aggregates, consumption goods (logs). *Source*: see Appendix.

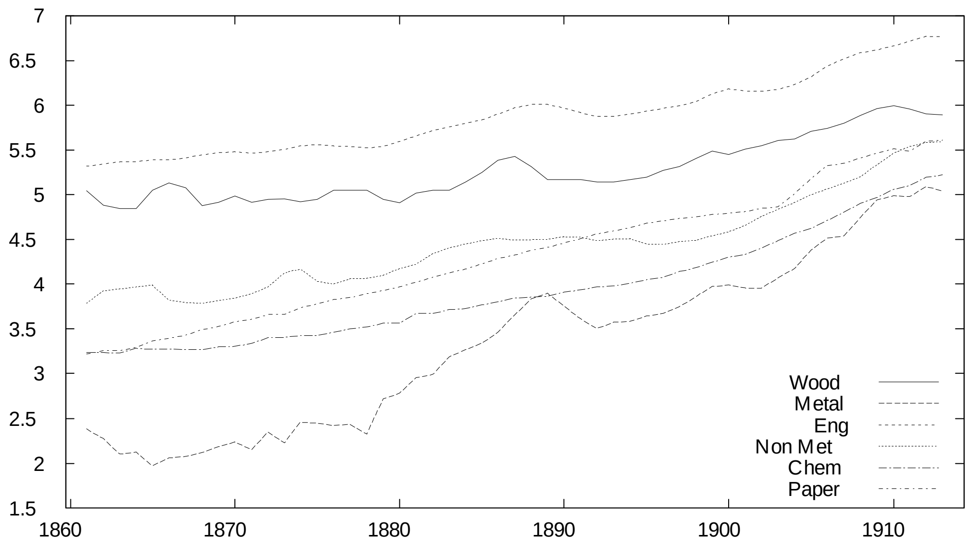

Figure 4: Time series of national industry aggregates, intermediate and investment goods (logs). *Source*: see Appendix.

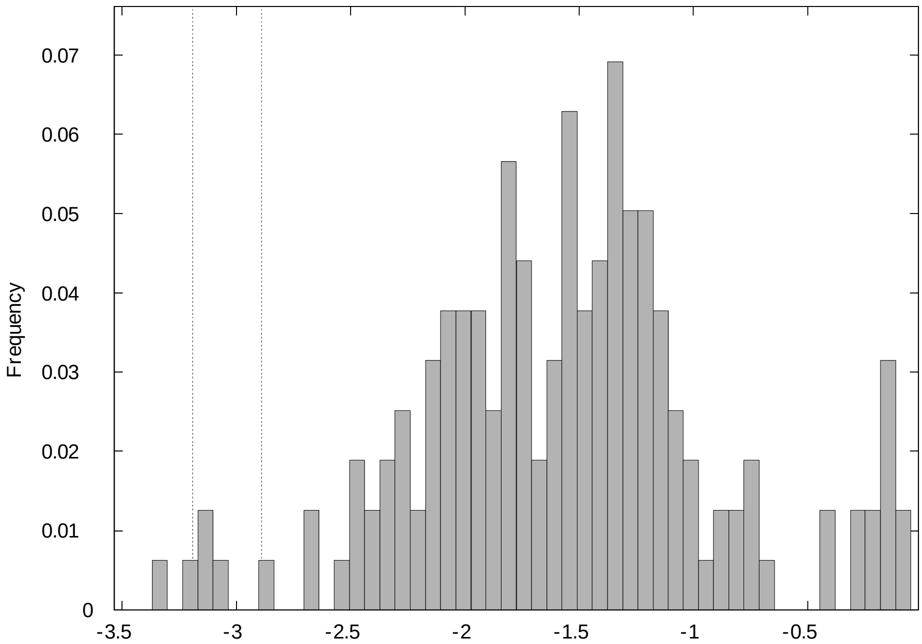

Figure 5: ADF-GLS unit root test with constant and trend (5% and 10% critical values: -3.19, -2.89, marked by the dotted lines).

3 Modelling set-up

Let $\mathbf{Y}$ be the $T \times (N \cdot R)$ data matrix obtained joining N matrices of dimension $T \times R$, so that the time series of the first industry in all regions are placed in the first R columns, and those of industry N in the last R columns:

$$\mathbf{Y} = \begin{bmatrix} y_{111} & \cdots & y_{1R1} & \cdots & y_{211} & \cdots & y_{2R1} & \cdots & y_{NR1} \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_{11T} & \cdots & y_{1RT} & \cdots & y_{21T} & \cdots & y_{2RT} & \cdots & y_{NRT} \end{bmatrix}$$

The comovements of these $N \cdot R$ series are naturally modelled resorting to a factor model. To this end, let $\mathbf{F}$ be a $T \times k$ matrix of unobservable non-stationary common factors, $\mathbf{\Lambda}$ a $k \times (R \cdot N)$ matrix of factor loadings, and $\boldsymbol{\varepsilon}$ a $T \times (R \cdot N)$ error matrix. Then the data can be represented by the approximate factor model (Chamberlain and Rothschild, 1983)

$$\mathbf{Y} = \mathbf{F}\mathbf{\Lambda} + \boldsymbol{\varepsilon} \tag{1}$$

where the errors are allowed to be weakly dependent. Usually in this stream of literature no attention is paid to the structure of the errors, literally seen as "residuals". However, in our set-up a crucial part of the information is actually included in the errors. Given the structure of the dataset we can in fact expect them to be correlated in three directions: over time, across industries and between regions. Since we are interested in long-run developments, weak dependence over time is not particularly interesting and will not be examined further. To gain an understanding of the other two type of correlations, over industries and regions, let $\mathbf{W}_{ij}$ be a $R \times R$ weights matrix measuring the proximity between industries i and j across the R regions. Then the entire $(R \cdot N) \times (R \cdot N)$ contemporaneous dependence structure is described by a block weights matrix $\mathbf{W}$ including $N \times N$ matrices of dimension $R \times R$:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} & \dots & \mathbf{W}_{1N} \\ \mathbf{W}_{21} & \mathbf{W}_{22} & \dots & \mathbf{W}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{N1} & \mathbf{W}_{2N} & \dots & \mathbf{W}_{NN} \end{bmatrix}$$

As we will see in the section discussing empirical results, this weights structure is actually so general to be rather difficult to handle. To concentrate on regional spillovers, we may adopt the solution taken in the construction of multi-regional input-output tables (see, *e.g.*, Hewings, 1985), assuming that only shocks affecting the same industry are correlated across regions. Under this assumption $\mathbf{W}$ is block diagonal, with spatial weights matrices $\mathbf{W}_i$ on the diagonal and zeroes elsewhere:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{W}_N \end{bmatrix} \tag{2}$$

Assume for notational simplicity spatial autocorrelation to be captured by a spatial AR coefficient ρ homogenous across industries and let $\mathbf{u}$ be a $T \times N \cdot R$ matrix of residuals. Then the error dependence may be described by the spatial autoregressive model

$$\boldsymbol{\varepsilon} = \rho \boldsymbol{\varepsilon} \mathbf{W} + \mathbf{u} \tag{3}$$

Equations (1) and (3) form what by analogy with Factor Augmented VAR's (FAVARs) we may label Factor Augmented Spatial Error Model (FASEM). According to this model value added in industry i of region r at time t, y_{irt}, can be decomposed into the sum of the common components (the k common factors F_j, $j = 1, \dots, k$, weighted by their loadings λ_{irj}) and the error ε_{irt}. In turn, the latter is the sum of the spillovers from the shocks affecting the same industry in the other regions, and a white noise u_{irt}. Denoting by w^i_{rs} the generic element of the spatial weights matrix for industry i, $\mathbf{W}_i$:

$$\varepsilon_{irt} = \rho \textstyle\sum_{s=1}^{R} w^i_{rs} \varepsilon_{ist} + u_{irt}$$

so that the factor model can be written as:

$$y_{irt} = \textstyle\sum_{j=1}^{k} \lambda_{irj} F_{jt} + \rho \sum_{s=1}^{R} w^i_{rs} \varepsilon_{ist} + u_{irt}$$

While the extension to heterogenous spatial autocorrelation is trivial, allowing for interregional spillovers across different industries is a more delicate matter, as it would require devising an exogenous contiguity matrix between industries.

To appreciate the spatial structure of the model, substitute for the error matrix $\boldsymbol{\varepsilon}$ in (3) from (1) and rearrange, obtaining

$$\begin{aligned} \mathbf{Y} &= \mathbf{YW}\rho + \mathbf{F\Lambda}(\mathbf{I} - \mathbf{W}\rho) + \mathbf{u} && (4) \\ &= \mathbf{YW}\rho + \mathbf{F}\widetilde{\mathbf{\Lambda}} + \mathbf{u} && (5) \end{aligned}$$

Model (5) can be labeled Factor Augmented Spatial Autoregressive (FASAR). Comparing (1)-(3) with (5) we can see that in the former factor loadings as well as errors are allowed to be spatially dependent, while in the latter any spatial structure is removed from the loadings and the errors are white noise; spatial dependence is modelled through spatial lags of the variable of interest. The FASAR and FASEM structures are essentially interchangeable, but since the notion of spatial spillovers is naturally applied to shocks and allowing for a spatial structure in the loadings is actractive from the intepretation point of view, the FASEM structure is arguably more appealing. It will thus be used in our empirical analysis.

4 Estimation

4.1 Factor model

In our case the combined industries×regions cross-section dimension, 159, is large enough to allow estimation of the latent factors $\mathbf{F}$ and their loadings $\mathbf{\Lambda}$ using the principal component techniques described in Bai (2004). In this approach the factor loadings are estimated as $\sqrt{RN}$ times the k eigenvectors corresponding to the k largest eigenvalues of the matrix $\mathbf{Y}'\mathbf{Y}$, and the factors by $\mathbf{F} = \mathbf{Y}\mathbf{\Lambda}/RN$.

An often delicate point of principal components studies is the dealing with the scale of the variables. Since the key scale element here is the dimension of the regions, we decided to scale value added by regional population. Since over the period of interest this is available only for census years (1861,1871,1901,1911), annual estimates have been constructed by linear interpolation. As it can be appreciated from Table 2, allowing for a maximum of four factors the three information criteria suggested by Bai (2004) give partially contrasting results. The first information criteria, IPC_1, is minimised by three factors, but the difference with the value for two factors is marginal. The second, IPC_2, is equal at the second decimal point with two and three factors, and the third, IPC_3, favours one factor. With the maximum set at three IPC_1 and IPC_2 suggest instead two factors, and IPC_3 still one. Considering also that IPC_3 is consistent only when the N is not large relative to T, so that the suggestions of IPC_1 and IPC_2 may be considered overall as more reliable, two factors seems a reasonable choice. This conclusion will receive further support from the finding that the residuals are actually stationary, implying that two factors are sufficient to capture all non-stationarity of our dataset. From Fig. 6 we can see that the first is an essentially monotonous trend capturing long-term growth, the second a long cycle with peaks at the extremes of the sample and through in the 1890's (when the trend also slowed down, with a few years of negative growth).

Table 2

Factor Model Information Criteria

	Max=4			Max=3		
Factors	IPC_1	IPC_2	IPC_3	IPC_1	IPC_2	IPC_3
1	3.20	3.24	**3.99**	3.41	3.47	**4.53**
2	2.55	**2.64**	4.13	**2.99**	**3.11**	5.19
3	**2.51**	**2.64**	4.85	3.17	3.35	6.45
4	2.82	2.99	5.91			

min for each column in bold face

It is instructive to compare the paths followed by these two factors with historical events of the time. While economic historians are always skeptical about periodization, the prevailing literature acknowledges

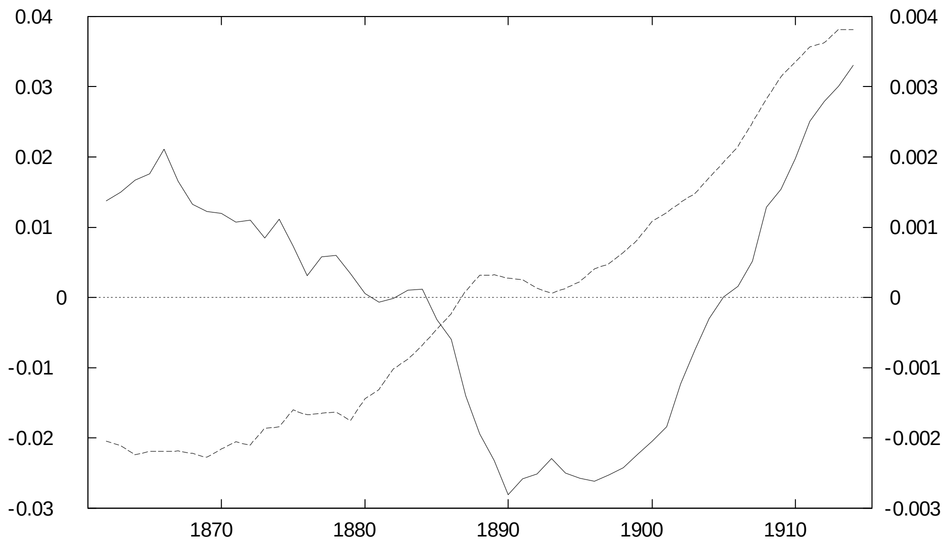

Figure 6: Non-stationary common factors.

that the years 1880-1887 ca were years of rapid industrial growth, the year 1887-1895 ca of economic crisis, and the years 1895-1907 ca of sustained growth again. An early example of this periodisation is Bachi (1919); later, Luzzato (1963, p. 263) described the 1889-1894 time interval as the "darkest years since unification [*1861*]". The works by Fenoaltea (1988, 2011, 2014) represent the most important contributions to understand the temporal evolution of Italian industry at the national level. Fenoaltea (1988) shows that the long swings in Italy's industry described above (the growth of the 1880s, the crisis of the 1890s, and the upswing of the new century) where not only an Italian phenomena. Rather, they represented a co-feature shared by the peripheral countries of the time, tied to the exogenous shift in the supply of foreign capital, and much involved the construction sector. More specifically, Fenoaltea (2011) shows that during the 1890s new constructions in the railway sector dropped dramatically, and that a similar reduction, although less pronounced, concerned other public works and urban construction. In 1887 the collaps of the real estate bubble caused a severe financial crisis, which led in 1893 to the bankruptcy of the "Banca Romana", one of the six national banks authorized at the time to issue currency. The creation in the same year of the Bank of Italy, aimed at controlling the financial markets and avoid future crisis, was the main institutional innovation of this period. Summing up, the paths of the two estimated factors appear nicely consistent with available *a-priori* evidence on the evolution of the Italian economy in the 19th century. The next step is to examine in some details the loadings of the trending factor, as these reveal the role played by the various regions in the national development. The full matrix is reported in Table A1 in the Appendix, and industry and regional averages in Tables 3 and 4 below. In some industries (most notably, Textiles, Engineering, Chemicals, Non Metallic Mineral Products) the loadings decline regularly from the northern to the southern regions, while in others they vary between regions without an obvious pattern. On the average the former pattern prevails, so that a north-south negative spatial trend stands out clearly. Setting the average over all industries for Piedmont=100, those of Liguria and Lombardy (the other two regions of the North-West) are respectively 96 and 94: the grand national average over industries and regions, 71, is almost one third lower. Both NE regions (Venetia and Emilia), have averages close to the national value. The same holds for Centre (Tuscany, Marches, Umbria, Latium), where however the regional values vary widely between a minimum of 33 (Latium) and a maximum of 93 (Umbria). Finally, the average of the southern regions is 61, with most cases well below the national mean. From Table 3 we can appreciate that these remarks largely hold if we look at the medians, rather than the means, of the industry loadings.

Table 3
Means and medians across industries of the loadings of the trend factor

	Piedmont	Liguria	Lombardy	*NW*	Venetia	Emilia	*NE*	
Mean	100	95.8	93.6	*96.4*	72.5	70.0	*71.3*	
Median	100	102.7	95.6	*99.4*	72.8	86.8	*79.8*	
	Tuscany	Marches	Umbria	Latium	*Centre*			
Mean	83.7	71.5	93.1	33.1	*70.4*			
Median	81.4	63.4	70.7	37.9	*63.4*			
	Abruzzi	Campania	Apulia	Basilicata	Calabria	Sicily	Sardinia	*South*
Mean	84.0	60.8	56.4	36.4	66.1	51.9	72.3	*61.1*
Median	80.0	68.9	41.6	41.6	56.6	60.8	73.4	*60.4*

The Values of the macroareas (NW, NE, Centre, South) are averages of the regional statistics.

Taking averages in the other directions, over regions, we can see from Table 4 that as to be expected the highest loadings are associated with industries producing investment and intermediate goods (Metals, Non Metallic Minerals, Chemicals, Paper). The only exception is the Engineering industry, whic has loadings close to those of most consumption goods industries (Textiles, Clothing, Leather, Wood). The Food industry clearly lagged behind, with loadings in median half those of the other consumption industries.

Table 4
Means and medians across regions of the loadings of the trend factor

	Food	Textiles	Clothing	Leather	Wood
Mean	100	98.5	229.9	212.7	227.4
Median	100	194.4	244.8	258.2	277.6
	Metal	Eng	Non Met Min Prods	Chemicals	Paper
Mean	580.1	220.0	343.7	440.1	609.4
Median	668.6	275.3	484.3	526.8	751.2

Abbreviations: see Table 1.

The finding that the loadings appear to be somehow homogeneous within macroareas although heterogeneous across industries naturally prompts the question if they may be constrained to be constant for all industries in the same macroarea. Denoting by R_M the number of regions in the macroarea M (with $M =$ NW, NE, Centre, South) and by $\lambda_{M1} = (R_M \cdot N)^{-1} \sum_{r=1}^{R_M} \sum_{i=1}^{N} \lambda_{ri1}$ the average for macroarea M, this restriction can be formally written as

$$H_0 : \lambda_{ri1} = \lambda_{M1}, \ \forall i, \forall r \in M \tag{6}$$

and the restricted factor model as

$$y_{rit} = \lambda_{M1}F^0_{1t} + \sum_{j=2}^k \lambda_{rij}F_{jt} + \varepsilon_{rit},\ r \in M$$

In matrix terms, letting $\mathbf{\Lambda}^0_1$ the vector of loadings of the restricted factor, $\mathbf{F_2} = [F_2 \dots F_k]$ the $T \times (k-1)$ matrix of the factors from 2 to k and $\mathbf{\Lambda_2}$ the corresponding $(k-1) \times (R \cdot N)$ matrix of factor loadings, the restricted model is

$$\mathbf{Y} = \mathbf{F}^0_1\mathbf{\Lambda}^0_1 + \mathbf{F_2}\mathbf{\Lambda_2} + \boldsymbol{\varepsilon} \tag{7}$$

In view of Table 3 this hypothesis entails that the North-West had a leading role in the national development, with variability of the loadings over industries mattering less than regional variability.

The problem of testing hypothesis on the loadings has not received much attention in the literature, which has largely concentrated on estimation and selection of the number of factors. Amengual and Repetto (2014) have recently proposed an $F-$type test comparing unrestricted and restricted sum of squared residuals, which appears particulary suitable for our case. Applying their approach to our null hypothesis the procedure to compute the test would be the following:

1. fit the factor model and compute the unrestricted residual sum of squares, RSS.

2. compute the trend factor under the hypothesis of known loadings[7] $\mathbf{\Lambda}_1 = \mathbf{\Lambda}^0_1$, obtaining $\mathbf{F}^0_1 = \mathbf{Y}\mathbf{\Lambda}^0_1/RN$.

3. In order to compute the restricted factor model first subtract the restricted permanent component associated with the trend factor ($\mathbf{F}^0_1\mathbf{\Lambda}^0_1$) from the data, obtaining, say, $\mathbf{Z} = \mathbf{Y} - \mathbf{F}^0_1\mathbf{\Lambda}^0_1$, then fit the factor model to $\mathbf{Z}$. Finally, compute the sum of the squared restricted residuals of the latter model, RSS_0.

4. compute the test statistic comparing RSS_0 and RSS. Since the number of observations is $T \cdot R \cdot N$, the factor model has $k{\cdot}R{\cdot}N$ parameters, and assuming the loadings of the first factor to be known we are imposing $R \cdot N$ constraints. The traditional F statistic is thus $F = (T-k)(RSS_0 - RSS)/RSS$.

Since Amengual and Repetto (2014) derive the asymptotic distribution only for the stationarity case, in order to compute the significance level of the test we need to devise a bootstrap procedure. To this end we need to construct pseudodata obeying the null hypothesis of known matrix of loadings for the trend factor; in other terms, we need pseudodata equal to the sum of

(i) the permanent component under the null hypothesis, $\mathbf{F}^0_1\mathbf{\Lambda}^0_1$, and

(ii) a component obtained resampling the $T \times N$ matrix $\mathbf{Z} = \mathbf{Y} - \mathbf{F}^0_1\mathbf{\Lambda}^0_1$.

This set-up recalls closely the task to be carried out in order to construct a bootstrap panel unit root or cointegration test with dependent units. Thus, a natural candidate is Chang's (2004) procedure based upon application of the sieve bootstrap to the differenced data. However, there is a crucial difference to be taken into account. In the general case of more than one factor $\mathbf{Z}$ is the sum of an $I(1)$ component (the permanent components of the factors with uncostrained loadings, $\mathbf{F_2}\mathbf{\Lambda_2}$), and a stationary one (the residuals, $\boldsymbol{\varepsilon}$). Taking the first difference of the latter we will obtain a non-invertible MA process with a unit root. Hence, $\Delta\mathbf{Z}$ does not have a finite AR representation and the sieve bootstrap cannot be applied.

[7] More general hypothesis can be handled using iterated constrained principal components, as in Reis and Watson (2010).

We thus followed a different route, resampling the first differences through the Stationary Bootstrap. Parker, Paparoditis and Politis (2006) showed that this resampling scheme delivers asymptotically valid unit root tests, and Di Iorio and Fachin (2014) extended this result to cointegration tests and panel cointegration tests for independent units. We thus conjecture that this holds for a test on the factors loadings as well. As we will see, this conecture is supported by some exploratory simulations. Assuming $k > 1$ non-stationary factors, the procedure is the following:

1. subtract the restricted permanent component associated with the trend factor, $\mathbf{F}_1^0\mathbf{\Lambda}_1^0$, from the data, obtaining $\mathbf{Z} = \mathbf{Y} - \mathbf{F}_1^0\mathbf{\Lambda}_1^0$.
2. resample applying the Stationary Bootstrap the differences $\Delta\mathbf{Z}$, obtaining the pseudo-differences $\Delta\mathbf{Z}^*$, and cumulate them to obtain $\mathbf{Z}^*$.
3. append $\mathbf{Z}^*$ to the trend permanent component under H_0, obtaining $\mathbf{Y}^* = \mathbf{F}_1^0\mathbf{\Lambda}_1^0 + \mathbf{Z}^*$.
4. estimate the factor model on the pseudodata $\mathbf{Y}^*$ and compute the statistic F for the true null $H_0 : \mathbf{\Lambda}_1 = \mathbf{\Lambda}_1^0$, obtaining F^*.
5. repeat 2-4 a large number of times, obtaining a distribution of statistics F^* under H_0.
6. given that the rejection region of the $F - test$ is the right tail, compute the bootstrap $p-$value of the empirical test statistic F as the proportion of statistics F^* exceeding $F : p^* = prop(F^* > F)$.

As mentioned above, we tested the performance of this procedure using the simple Monte Carlo data generating process (DGP) used by Bai (2004). The permanent component is given by two $I(1)$ factors with unit loadings, and the stationary part follows independent ARMA processes:

$$x_{it} = \lambda_1 F_{1t} + \lambda_2 F_{2t} + \varepsilon_{it} \tag{8a}$$
$$F_{1t} = F_{1t-1} + e_{1t} \tag{8b}$$
$$F_{2t} = F_{2t-1} + e_{2t} \tag{8c}$$
$$\varepsilon_{it} = \phi\varepsilon_{1t} + \nu_{it} + \theta\nu_{it-1} \tag{8d}$$
$$e_{1t},\ e_{2t},\ \nu_{it} \sim NID(0,1) \tag{8e}$$
$$\lambda_1,\ \lambda_2 \sim NID(1,1) \tag{8f}$$

The results of some exploratory simulations are reported in Table 5 below. Given the exploratory nature of the exercise the number of bootstrap redrawings and Monte Carlo simulations is rather small, only 300 in both cases. Future work will include larger simulation exercises with a more general design. Here the test appears to be severely undersized for all cross-section dimensions for $T = 25$, but this is hardly surprising given that this sample is really very small. Doubling the time sample to $T = 50$, a time span quite common in long-run studies with annual data, the size bias is still present but considerably reduced, especially for $N = 150$ (which is the case relevant for our empirical problem). In fact, taking into account Monte Carlo variance Type I errors for this N, T combination can be considered not significantly different from nominal values. The same seem to hold for all N when $T = 75$.

Table 5
Size $\times 100$ *of bootstrap* F*-tests on factor loadings*

		T					
		25		50		75	
N	α	5.0	10.0	5.0	10.0	5.0	10.0
25		0.3	2.3	2.3	4.3	7.3	11.7
50		0.7	1.3	0.2	4.7	5.3	11.0
150		1.3	2.3	3.0	7.7	4.5	7.5

DGP: equations (8a)-(8e), $\phi = \theta = 0.5$
Block length: $1.75\sqrt[3]{T}$;
Monte Carlo s.e. for $\widehat{\alpha} = 5.0, 10.0$: 1.3,1.7.

Using this procedure we tested the null hypothesis described in (6): in words, normalising to 1 the coefficients of the trends factor of all industries in the North-Western regions, those of the industries in the North-Eastern and Central regions are respectively equal to 0.71 and 0.70, and those in the Southern regions to 0.61. Although quite restrictive, this hypothesis is comfortably not rejected by the data. The F statistic computed as described above is equal to 39.40, and the bootstrap p-value 0.43 (5000 redrawings). We can then conclude that the growth gap between, on one side, the NW regions, and, on the other, essentially all the other regions, was large, significant, and pervasive across industries.

4.2 Modelling the errors

Since the estimated residuals of factor models converge asymptotically to the population unobserved values (Bai, 2004, lemma 1) their time series properties can be studied using unit root tests designed for observed data. To this end we computed standard univariate ADF tests and the bootstrap panel unit root test by Chang (2004). Consistently with *a priori* expectations of stationarity, from Fig. 7 we can see that 90% of the p-values of the statistics for the individual residual series are smaller than 0.10 (details in Table A2 in the Appendix). The impression of stationarity is fully confirmed by Chang's (2004) testing procedure. The F-statistic for H_0 : "all residuals are $I(1)$" is 165.26. Using 5000 redrawings the entire bootstrap estimate of the distribution of the statistics lies on the left of this value[8], so that the bootstrap estimate of its p-value is zero at all decimal digits. Note that since this is a bilateral test with alternative hypothesis both stationarity and explosive behaviour, it is potentially subject to power problems (in fact, Chang, 2004, proposed some unilateral tests as well). This strong rejection thus provides good support for the next step of our analysis, which relies on the stationarity of these residuals.

Having established that the de-factored residuals are stationary over time, the next step is investigating their spatial structure estimating the panel spatial autoregressive model (3). In principle this is a straightforward task; the loglikelihood is reported by e.g., Elhorst (2014). In practice, modelling all industris simultaneously there may be serious problems, as the likelihood would include a matrix of dimension $(T \cdot R \cdot N)^2$, which in our case means 70 million cells. We thus took the alternative route of estimating separate panel spatial autoregressive models for the residuals of each industry, implicitly assuming a block-diagonal contiguity matrix of the type (2). Given the rather large dimension of the

[8] The 95th percentile of the bootstrap distribution is 104.27.

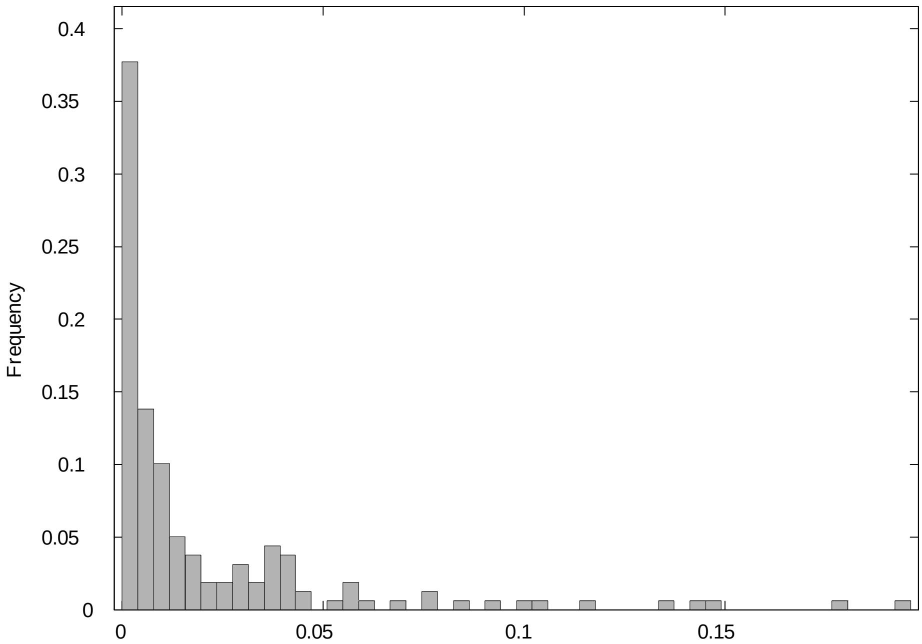

Figure 7: P-values of the ADF unit root tests of non-stationarity of the residuals of the factor model.

spatial units we adopted for all industries binary contiguity weights, with $w_{ij} = 1$ for regions i, j sharing a border and zero else. For the two main islands, Sicily and Sardinia, we proceeded as follows. Sicily has been considered a neighbour of Calabria, from which it is separated by the narrow (less than 5 km wide) strait of Messina. The regions closest to the other island, Sardinia, are Liguria, Tuscany and Latium. Although the latter is the closest in bird fly distance, it had no important ports, present instead in Liguria and Tuscany: in 1871 the total tonnage of ships registered in Liguria was 634 thousands tonnes (of which 25 steam vessells), in Tuscany 52 thousands, and in Latium only 2 thousands. We thus decided to include in the neighbourhood of Sardinia only Liguria and Tuscany. As it can be seen from Table 6, ML estimation of the panel spatial autoregressive models suggests significant spatial autocorrelation in the shocks of all industries, with all coefficients strongly significant and all but one (Wood) well above 0.50.

Table 6
ML estimates of the spatial autoregressive coefficients of the factor model residuals

Food	Textiles	Clothing	Leather	Wood
0.70	0.65	0.58	0.66	0.39
(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
Metal	Eng	Non Met Min Prods	Chemicals	Paper
0.63	0.65	0.61	0.57	0.66
(0.02)	(0.02)	(0.02)	(0.03)	(0.02)

Abbreviations: see Table 1
standar errors in brackets.

5 Conclusions

Our purpose was to study the development of the Italian manufacturing industry in 19th century exploiting a recently released dataset including 1861-1913 annual series of value added disaggregated for ten manufacturing industries and 16 regions. Long-run growth has been modelled through the approximate factor model by Bai (2004). We found just two non-stationary factors, an essentially monotonous trend and a very long cycle, sufficient to capture long-run growth of the 159 series examined. Looking at the regional averages loadings of the trend factor appear to follow a clear spatial pattern, highest in the North-West and lowest in the South. This structure has been tested using a bootstrap F-test of the hypothesis that the loadings for all industries in each macroarea (NW, NE, Centre, South) are equal to their respective means. The test found this restriction to be largely compatible with the data. Finally, spatial autoregressive panel models of the errors of the factor model suggested that spatial spillovers were significant and essentially similar in all industries. Obviously, there are many open questions under current investigation. Among them the most important are an extensive Monte Carlo evaluation of the properties of the bootstrap test on the loadings and a more thorough analysis of spatial interactions.

6 Appendix

6.1 Data

The data used in this paper are taken from a comprehensive dataset of annual time series of value added at 1911 prices, disaggregated for 12 industries and 16 italian regions from 1861 to 1913. Most of the series are included in Ciccarelli and Fenoaltea (2009, 2014) and available in the data section of the Bank of Italy website (https://www.bancaditalia.it/statistiche/storiche/), while others (including Food, Wood, Paper) are still preliminary, and are based on ongoing research (Ciccarelli and Fenoaltea, *forthcoming*). The data have been constructed on the basis of a wide set of primary historical sources, including industrial and population census. The estimation strategy varies with the industrial sector and the historical sources available, and thus cannot be described in full detail here. However, the first step has always been to obtain data for physical production of the single products at the regional level. For instance, in the case of the Chemical industry regional volume time series for about 100 products (various acids, fertilizers, rubber, etc) have been constructed. Then, each volume series has been transformed into a value added series at 1911 prices using a unit value added coefficient evaluated for 1911. These coefficients have been estimated using historical data on wage and capital (for details see Ciccarelli and Fenoaltea, 2014, chapter G). Finally, it should be remarked that the boundaries of the then Kingdom of Italy, officially proclamed in 1861, changed twice during the 1861-1913 period. Venetia was annexed in 1866, and Latium, including Rome, in 1870; both regions are however included in the dataset from the beginning. In other terms, the dataset is at the boundaries as established in 1870, which held until the end of the period[9]. Thus, the data for 1861 include Veneto and Latium, although at the time these regions belonged respectively to the Austrian-Hungarian Empire and to the Papal State.

[9] In fact, until the end of WWI in 1918, when Italy acquired the provinces of Trento, Bolzano and Trieste, in the North-East.

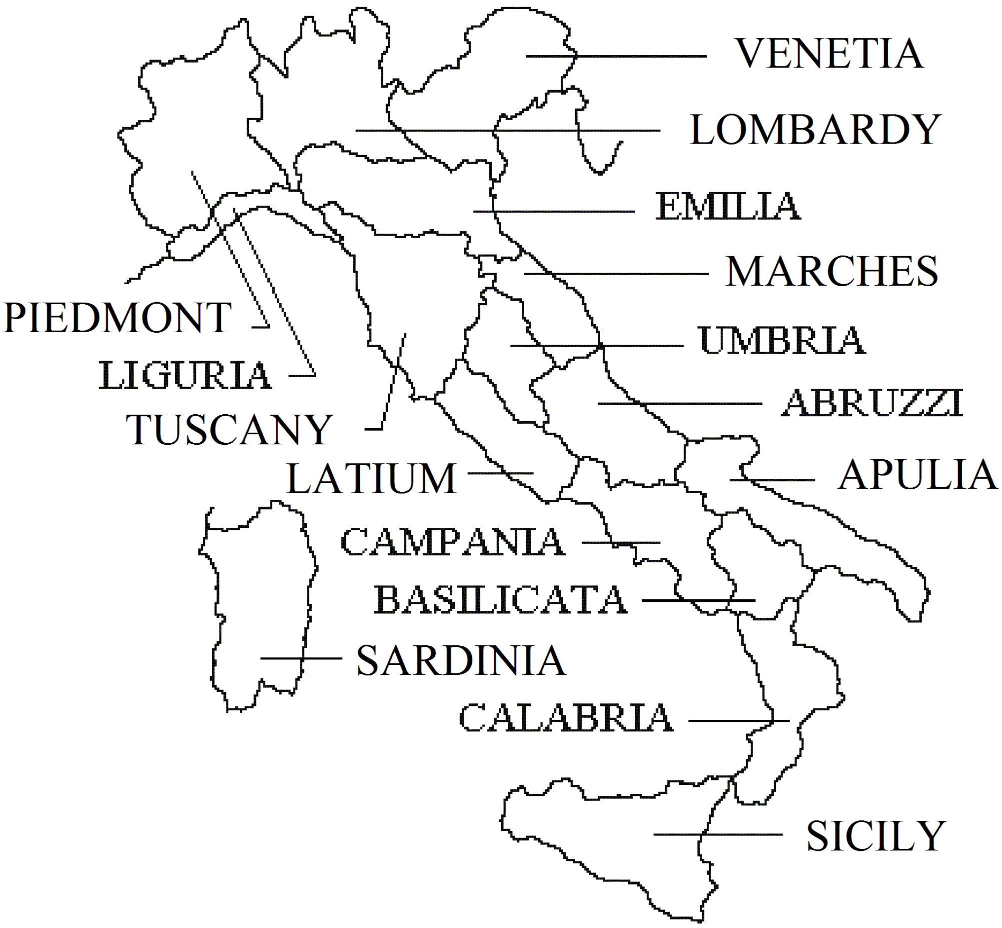

Figure 8: The Italian regions after 1870.

Table A1

Loadings of the trend factor

	Food	Text	Cloth	Leath	Wood	Met	Mach	Non Met Min Prods	Chem	Paper
Piedmont	6.1	22.0	16.5	9.1	9.6	37.6	21.5	24.4	26.6	30.8
Liguria	9.7	12.2	14.4	9.4	10.8	44.6	16.6	25.0	25.4	18.0
Lombardy	5.4	22.6	12.4	8.6	11.7	25.7	21.4	27.2	32.3	29.7
Venetia	3.3	19.1	5.4	5.4	8.9	24.6	8.4	19.6	23.6	20.9
Emilia	9.8	-10.1	15.3	8.7	10.2	19.6	12.8	21.8	19.6	25.9
Tuscany	4.1	15.3	8.5	91	10.8	44.5	13.6	18.5	26.7	26.7
Marches	2.8	-0.6	8.9	9.8	9.0	18.5	6.6	26.8	16.2	32.1
Umbria	0.9	11.1	6.2	9.4	7.8	71.5	10.9	14.3	43.7	28.5
Latium	0.9	-20.8	12.5	6.2	5.0	-1.8	10.0	5.7	15.0	31.4
Abruzzi	2.2	4.0	8.4	11.2	11.3	46.4	3.0	14.8	21.4	31.5
Campania	5.7	-6.5	8.8	12.5	9.9	33.4	13.2	2.1	12.7	22.5
Apulia	2.0	0.2	8.7	9.3	11.7	12.5	10.7	5.4	13.3	33.6
Basilicata	2.5	-1.1	7.2	11.6	9.6	-	2.5	-1.3	-1.1	25.8
Calabria	3.9	10.1	8.1	11.1	12.8	-18.6	2.5	19.7	9.1	27.8
Sicily	3.1	-21.5	9.6	9.6	9.8	9.4	8.1	16.6	10.3	24.8
Sardinia	8.8	14.0	12.7	10.5	12.9	23.9	6.6	3.9	18.6	24.0

7 References

A'Hearn, B., Venables, A. (2013) "Regional Disparities: Internal Geography and External Trade" in G. Toinolo (ed.), *The Oxford Handbook of the Italian Economy Since Unification*, New York, Oxford University Press.

Amengual, D., L. Repetto (2014) "Testing a large number of hypotheses in approximate factor models" Working Paper, CEMFI.

Bachi, R. (1919) "Le fluttuazioni stagionali nella vita economica italiana" *Annali di statistica*, 5, 1-261.

Bai, J. (2004) "Estimating cross-section common stochastic trends in nonstationary panel data" *Journal of Econometrics,* 122, 137–183.

Chamberlain, G., Rothschild, M. (1983) Arbitrage, factor structure and mean-variance analysis in large asset markets" *Econometrica,* 51, 1305–1324.

Chang, Y. (2004) "Bootstrap unit root tests in panels with cross-sectional dependency" *Journal of Econometrics*, 120, 263-293.

Ciccarelli, C., Fenoaltea, S. (2009) "La produzione industriale delle regioni d'Italia, 1861-1913: una ricostruzione statistica. 1 Le industrie non manifatturiere" Rome, Bank of Italy.

Ciccarelli, C., Fenoaltea, S. (2014) "La produzione industriale delle regioni d'Italia, 1861-1913: una ricostruzione statistica. 2 Le industrie estrattivo-manifatturiere" Rome, Bank of Italy.

Ciccarelli, C., Fenoaltea, S. (forthcoming) "La produzione industriale delle regioni d'Italia, 1861-1913: una ricostruzione statistica. 3 Le industrie agricolo-manifatturiere" Rome, Bank of Italy.

Di Iorio, F., Fachin, S. (2012) “A Note on the estimation of long-run relationships in panel equations with cross-section linkages” *Economics: The Open-Access, Open-Assessment E-Journal*, 6, 2012-20. http://dx.doi.org/10.5018/economics-ejournal.ja.2012-20

Elhorst, J.P. (2014) *Spatial Econometrics* Berlin, Springer.

Elliot, G., Rothenberg, T.J., Stock, J. (1996) "Efficient tests for an autoregressive unit root" *Econometrica,* 64, 813-836.

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

Fenoaltea, S. (1988) "International resource flows and construction movements in the atlantic economy: the Kuznets cycle in Italy, 1861–1913" *Journal of Economic History*, 48, 605-637.

Fenoaltea, S. (2011) "The Reinterpretation of Italian Economic History: From Unification to the Great War" Cambridge, Cambridge University Press.

Fenoaltea, S. (2014) "The fruits of disaggregation: the general engineering industry in Italy, 1861-1913" Carlo Alberto Notebooks, no. 358.

Geary, F., Stark, T. (2015) "Regional GDP in the UK, 1861–1911: new estimates" *The Economic History Review*, 68, 123–144.

Gerschenkron, A. (1962) "Economic backwardness in historical perspective, a book of essays" Cambridge, Mass.„ Belknap Press of Harvard University Press.

Henning, M., Enflo, K., Andersson, F.N.G. (2011) "Trends and cycles in regional economic growth: How spatial differences shaped the Swedish growth experience from 1860–2009" *Explorations in Economic History,* 48, 538–555.

Hewings, G.J.D. (1985) *Regional Input-Output Analysis* Scientific Geography Series vol. 6, Beverly Hills, CA, Sage Publications.

Luzzato, G. (1963) *L'economia Italiana dal 1861 al 1894* Milano, Banca Commerciale Italiana.

Kuznets, S. (1971) *Economic Growth of Nations: Total Output and Production Structure* Cambridge, Mass., Belknap Press of Harvard University Press.

Maddison, A. (1995) *Monitoring the World Economy*, 1820-1992 Paris, OECD Development Centre.

McCloskey, D.N. (1991) “Kinks, tools, spurts, and substitutes” in *Patterns of European Industrialization: The Nineteenth Century*, R. Sylla, Toniolo, G. (Eds.), London, Routledge, 99-100.

Parker, C., E. Paparoditis, Politis, D.N. (2006) "Unit root testing via the stationary bootstrap" *Journal of Econometrics*, 133, 601-638.

Pritchett, L. (1997) “Divergence, Big Time” *Journal of Economic Perspectives*, 11, 3-17.

Reis, R., Watson, M.W. (2010) "Relative goods' prices, pure inflation, and the Phillips correlation" *American Economic Journal: Macroeconomics*, 2, 128-157.