Measuring Knowledge Spillover Effects via Conditional Nonparametric Analysis

Andrea Bonaccorsi
a.bonaccorsi@gmail.com

Cinzia Daraio
cinzia.daraio@iit.cnr.it
daraio@dsea.unipi.it

School of Engineering
University of Pisa, Italy

This version updated: 7 October 2007

Abstract[1]

Knowledge spillovers are central in modern theories of innovation and growth. There is a large gap, however, between the recognition of the role of spillovers in several theories and the empirical appreciation. In this paper we explore a new approach to the measurement of spillovers. This is based on the exploitation of a recently developed family of techniques in nonparametric efficiency analysis, which allow the estimation of the impact of external factors on the technical efficiency of productive units. We advocate the use of these tools and give a demonstration of their potential, using data at territorial level for Italy.

Keywords: knowledge spillovers, manufacturing industry, growth, efficiency analysis, conditional measures, robust nonparametric estimation

JEL Classification: C14, D20, L60, L23, O30, R50

[1] We gratefully acknowledge the financial support of the AQUAMETH Project (PRIME Network of Excellence, 6th Framework Programme), of the Italian National Project FIRB *The reorganization of the public research system for technology transfer*, iRis, of the Italian National Project PRIN *Systemic spillover on competitiveness of the manufacturing industry,* and of the European project INNODEC (Interreg III C sud). We thank Alessandro Daraio for assistance in data analysis. Earlier version of this paper were presented at the *Agglomeration and Growth* Workshop in Kiel, April 2007 and at the 34th EARIE Conference in Valencia, September 2007; we thank participants for helpful discussions and comments; the usual disclaimers apply.

1. Introduction

After two decades of intense research the notion of spillover, while extremely powerful in theory, still proves elusive in the empirical analysis. There is still a large gap between the recognition of the role of spillover in several prominent economic theories and the methods used for the empirical appreciation.

In this paper we explore a new approach to the measurement of spillover, which radically departs from the econometric approaches followed in the literature.

The approach is based on the exploitation of a recently developed family of techniques in nonparametric efficiency analysis, which allow the estimation of the impact of *external factors* on the technical efficiency of productive units (Daraio and Simar, 2005; 2007). While traditionally a number of conceptual and practical issues have severely limited the use of nonparametric techniques, the new developments solve these issues and deliver a flexible and powerful set of tools.

The paper is organized as follows. In Section 2 a short discussion of the theoretical status of the notion of spillover and of existing econometric approaches is carried out. In Section 3 the notion of conditional robust nonparametric efficiency is introduced, leaving technical details to the Appendix. Section 4 proposes a definition of knowledge spillover as the combined impact of the innovative and scientific activities on the efficiency of the manufacturing industry, and applies it to data on production activity at local (province) level in Italy. Spatial dependence effects are considered in the estimation. The magnitude and distribution of knowledge spillovers from the overall (combined) volume of patenting and publication activity, from the intensity of this activity, and from the sectoral specialization of publications are investigated. The identified dimensions of knowledge spillovers are largely discussed and policy implications derived. Section 5 discusses possible lines of inquiry including applications to tertiary activities, extensions to alternative specifications of spatial dependence factors, and inclusion of robust efficiency estimates in general input-output models.

2. Issues in the identification and measurement of spillover effects

2.1 A crucial theoretical role

Spillovers are positive externalities generated by those agents that invest in certain activities and cannot prevent other agents to benefit from these activities. More specifically, knowledge spillovers refer to the fact that firms can benefit from knowledge generated by others (i.e. by other firms, including rival firms; or public institutions, in the same industry or territory, or elsewhere) without having to pay for them. Knowledge spillovers take various forms: collaboration between researchers, mobility of researchers and technicians from academia to industry and vice versa or within the industry, informal exchanges of ideas, unintended information disclosure, reverse engineering of easily available rival products. In all these cases knowledge cannot be considered as a production factor *stricto sensu*, because companies do *not* pay for them. They benefit from knowledge produced elsewhere, that is accessible without the need to pay the full cost (although some positive cost of absorption is needed).

The whole theory of spillover effects is affected somewhat by a paradox: there is no correspondence between the large theoretical role of knowledge spillovers in a number of areas, and the sophistication of empirical methods used to capture them. The larger the theoretical importance, the more elusive the identification and measurement.

On one hand, in fact, spillovers are clearly relevant in many areas of economic theory: general equilibrium, theories of growth, economics of innovation, economic geography, and international economics. Without any ambition of completeness, let us briefly recall the most important ones.

In general equilibrium theory and public economics, it is clearly relevant as an explanation of market failure and as a foundation for State intervention in research and innovation, following the classical treatment of Nelson (1959) and Arrow (1962). The existence of spillover is at the core of the discrepancy between the private and the social rate of return of the investment into knowledge production.

The notion of spillover is firmly rooted in the idea that knowledge is not fully appropriable. If knowledge were appropriable, then there would be perfect correspondence between private investment and private return, so that at the margin the market value of knowledge would be equal to the private cost. Therefore the notion of spillover is a very fundamental one, directly linked to the core of the theory of knowledge as an economic activity (Foray, 2004).

In theories of growth, it has been crucial to the attempt to explain persistence of positive and large rates of growth of advanced countries by endogenous growth theory (Romer, 1986; 1990) and neo-schumpeterian theories (Aghion and Howitt, 1992; 1998; Jones, 1995). In these fields the crucial point is that knowledge, as opposed to other factors of production, is not subject to diminishing returns. Diminishing returns appear to be an extremely general and robust feature of production factors, and are responsible for the inevitable tendency of economies to converge to a steady state rate of growth, implying that rich countries should, in the limit, grow less and less. If knowledge, on the contrary, is subject to increase in productivity with use, then an explanation is offered for the persistence of large rates of growth in advanced economies. Again, the notion of spillover is essential for the argument of increasing returns: while it is true that the individual use of knowledge generates increasing returns, the order of magnitude of this effect would be modest if each company might entirely appropriate the benefit, without generating positive externalities for other companies. The externality component is a large portion of the overall increasing return effect.

In the economics of innovation the interest for knowledge spillover follows several streams. An important literature has dealt with the benefit that the public sector research can bring to industry, by generating flows of knowledge mediated by professional roles of individuals and personal interaction. In this literature the main channels through which knowledge is expected to flow are personnel mobility (Almeida and Kogut, 1999; Saxenian and Hsu, 2001), and personal interaction between researchers and managers, as documented by paper trails in patents and location decisions of firms (Jaffe et al., 1993; Zucker and Darby, 1996; Zucker, Darby and Armstrong, 1999). Here a certain interest is in examining whether the tacit nature of knowledge requires forms of interaction between universities and firms that benefit from co-location in the same area (Mansfield, 1980; Jaffe, 1989; 1996; Anselin, Varga and Acs, 1997; Varga, 1998; Cohen, Nelson and Walsh, 2002).

Another stream of literature in the economics of innovation deals with the idea that the investment into R&D in one industry may benefit indirectly other industries by providing new intermediate goods as inputs or new production technologies, for which the recipient industry has not paid the full cost. The econometrics of inter-industry flows of knowledge (Scherer 1982; Link, 1983; Griliches and Lichtenberg, 1984; Bernsten and Nadiri, 1988; 1989) has done a great job in producing estimates of the order of magnitude of this effect.

Finally, a large stream of studies in the economics of innovation deals with the role of multinational corporations as sources of spillovers. This may take place in two main forms: flows of knowledge *from* multinational companies to host countries and viceversa (Blomstrom and Kokko, 1998; Barrell and Pain, 1999; Aitken and Harrison, 1999; Cantwell and Iammarino, 2001) in particular via human capital mobility (Fosfuri, Motta, Ronde, 2001), and flows of knowledge *within* multinational companies (Almeida, 1996; Gupta and Govindarajan, 2000; Frost, 2001; Szulanski, 2006). Here an important theoretical contribution is the idea that multinational companies act as governance mechanisms, internalizing transactions and permitting a more efficient utilization of knowledge. While a large part of this effect is intentional and planned, large spillovers still take place. Conversely, there are also powerful organizational mechanisms that prevent knowledge flows.

The idea of spillover is also central to the new economic geography (Krugman, 1990; 1991) and the theory and empirics of geographical agglomeration of industry (Ottaviano and Puga, 1997; Rosenthal and Strange, 2004). Knowledge spillovers are considered one of the mechanisms for agglomeration, alongside input sharing and labour pooling mechanisms (Rosenthal and Strange, 2004). The idea is that companies may find it convenient to co-locate with rival companies, in the expectation to benefit from unintended information leakage or informal exchange of knowledge. In this literature an important overlapping has been created with the economics of innovation, since both literatures are strongly interested in explaining the geography of innovative activities (see among others Feldman, 1994; Audretsch and Feldman, 1996; Jaffe, Trajtenberg and Henderson, 1993; Autant-Bernard, 2001; Bagella and Bechetti, 2002; Orlando, 2004). The keen interest on this issue comes from theoretical and practical considerations. From the theoretical point of view, it is interesting to observe whether the continuous creation of scientific and technological opportunities produces a geographic dispersion of activities, generating opportunities for less favoured regions and territories, or rather it reinforces existing agglomerations, or strong areas. Empirically, it has been repeatedly observed, particularly in Europe, that research activities tend to cluster in large metropolitan areas (typically around the large capital towns) and in high tech regional agglomerations of large size and intensity (Carrincazeaux, Lung and Rallet, 2001).

Finally, international spillovers of knowledge are considered as one of the main sources of growth for less developed countries, particularly for those engaged into catching up (Kokko, 1994; Coe and Helpman, 1995; Kokko, Tansini and Zejan, 1996; Gwanghoon, 2005). While earlier models of North-South divide considered technology flows only in the form of capital investment from advanced countries to less developed countries, more recent models incorporate the notion of knowledge spillover, allowing for a variety of potential effects.

As this short and highly incomplete review clearly shows, there are several large areas of economic theory that place the notion of spillover at their core.

Rebus sic stantibus, one would expect a great deal of methodological work to be carried out in order to refine the concept and develop new measurement and estimation approaches. This is not the case, however. There are, in fact, several unsolved conceptual problems.

First of all, is knowledge a production factor? Is it separated from embodiments, such as labour or machinery?

Assuming knowledge as a direct productive factor has several advantages, using proxies such as R&D expenditure or R&D stock. Pursuing this avenue of research seems promising, in the perspective of building a knowledge production function (Griliches, 1992) in which the source of spillovers is explicitly modelled. Alternatively, R&D stock, or the share of workforce allocated to research activities, can be modelled to influence total factor productivity. Whatever the specification, empirical analyses of knowledge based on production functions are somewhat puzzling: the magnitude of effect of R&D on growth, once one admits a role for imitation, is not as large as posited by the theory ((Jones, 1995; Comin, 2002).

At a fundamental level, is knowledge an input to the production process, as suggested by endogenous growth, or is rather an output, or both? If one admits that knowledge is both an input and an output, then the conventional production function approach becomes totally inadequate.

Second, which are the specific channels for knowledge spillovers? And are there differential effects across various forms of spillovers? This issue has been addressed only recently by studies that identify and map specific spillover mechanisms, such as personnel mobility. Using the notion of spillover too broadly may indeed create a situation of catch-all explanation, reducing the analytical power (Breschi, 2000).

Third, which is the level of analysis? Studies that are explicit on mechanisms of spillover are usually based on surveys on individual firms. On the contrary, econometric exercises (see infra) try to capture an aggregate effect, but are vague about the specific mechanisms. As an example, considerable evidence has been collected on paper trails based on citations in patents: suppose this is accepted evidence for a specific mechanism of knowledge flow. What would be the aggregate effect on the economy?

Finally, there are technical problems in the use of stock of knowledge measures, as proposed by Griliches (1979) and developed, for example, by Adams (1990; 1993). In fact, the measure of stock is heavily dependent on time horizon and the rate of depreciation and there is no way to normalize these variables. By changing slightly the rates, for example, one can obtain wildly different estimates of the stock of knowledge.

Therefore, despite more than twenty years of intense research, the notion of knowledge spillover has not yet found a firm foundation.

2.2 Econometric issues

In addition to conceptual and definitional issues, there are also a number of technical problems in the econometric analysis[2]. There are several approaches to the estimation of knowledge spillover effects.

Indirect estimation

An important stream of literature derives the magnitude of spillover effects from the difference between private rates of return and social rates of return in R&D investments (see for a survey Dowrick, 2003; Wieser, 2005). As it is well known, there is a large difference between the two in most published case studies and this is taken as sufficient evidence of the existence of spillovers. The idea is that the social benefits from knowledge are much larger than the private ones because knowledge flows from private firms through mobility of key employees, leakage of strategic and technological information, reverse engineering of innovative products. The typical range of private rates of return in most studies, in the order of 20-30%, does not exceed to a great extent the average rate of return from physical assets, adjusted for risk. This confirms that R&D investment is economically sound, but not exceedingly attractive per se. On the contrary, the social rate of return is usually very high, in the order of 50% and beyond, implying a large positive externality for society. If these values were to be taken into account at face value, they would imply that knowledge spillover effects are of the same order of magnitude of appropriable benefits.

[2] For an extended discussion of the limits of conventional econometrics in these fields see Bonaccorsi and Daraio (2004).

Survey

A number of studies are based on field surveys in which the evidence for knowledge spillover is directly searched (Mansfield, 1980; 1991; 1998; Beise and Stahl, 1999; Wieser, 2005).
Here problems of memorization, ex post rationalization, and categorization on the part of respondents may apply and distort the measure.
While this approach sheds light on the microfoundations of spillovers, it is difficult to generalize and to build up an estimate at aggregate level.

Direct estimation: (a) cross effects

By a large margin the most used econometric approach has been including spillover effects directly in the specification of equations, usually in the production function.
Thus, alongside the regressor representing R&D expenditure at time t for firms i, the expenditure of all other firms in the same industry, or in other industries, or in other regions or local systems, are directly included in the equation.
Another important stream of analysis has considered that a source of spillover is the possibility for firms to purchase intermediate products that incorporate technical progress, for which they have not paid. Therefore the structure of input-output flows (at the level of industries), or the structure of international trade (at the level of countries) can be used to approximate the structure of flows of knowledge.
The direct approach requires the ex ante specification of the direction of impact of spillover in order to include the cross-effect among the variables of the equations.

Direct estimation: (b) spatial dependence

One of the crucial dimensions of spillovers is, of course, the spatial one. Knowledge produced in local area A may reach firms in the region B, with a probability and/or intensity inversely proportional to the distance.
In order to take into account these effects, a dedicated literature has developed techniques for disentangling spatial dependence from other effects (Anselin, 1988; Abreu, Florax, de Groot, 2005). These techniques make the spillover effect a function of the distance matrix between any pair of points in the geographic space, associated with a series of distance functions that may reflect a variety of effects. The estimates are much more precise and robust this way. With respect to the previous approach, spatial dependence models leave open the direction of impact of spillover, observing the effects ex post, instead of modelling interaction effects ex ante.

From a methodological point of view, the approach to knowledge spillover is subject to a dilemma.
If one wants to be precise on the spillover mechanism, the only reliable technique is survey-based, but then it is almost impossible to produce an aggregate estimate.
If, on the other hand, one is interested in the overall effect of knowledge spillover there is no way other than including various proxies of knowledge production directly in the equation. But in this way we are forced to accept the notion that, in order to have a spillover effect, the source of the spillover must be a production factor.
The crucial point is that many relevant sources of spillover are *not* production factors. They are not consumed in the activity of production, nor substituted by via investment decisions. They may affect productivity of firms without being channelled through production factors, or the purchase of intermediate goods incorporating technical progress generated in other industries. This creates a difficult dilemma, one that, to the best of our knowledge, is still unsolved in the literature.

3. Introducing conditional efficiency

Nonparametric efficiency analysis has a long history in the area of productivity analysis, but has rarely been applied to the economics of growth. One reason behind this is that conventional techniques, such as Data Envelopment Analysis, suffered from various methodological shortcomings, which made their use difficult (Bonaccorsi and Daraio, 2004). A recent wave of nonparametric techniques, called robust or order-m frontiers and illustrated in detail in the Appendix, solves for most of these limitations, and in particular for the influence of outliers and extremes in the data (from this feature comes their names of "robust").

One of the main areas of development of robust techniques is the so called conditional efficiency analysis[3]. The central idea is to reformulate the production process in a probabilistic setting, so that it is possible to condition the production activity to the impact of external factors without assuming ex ante the sign of impact itself. If these factors do not exert any impact, the conditional measure of efficiency will be equal to the unconditional one. The central role in conditional efficiency analysis is then played by the ratio between conditional and unconditional efficiency. In the following application we use robust nonparametric measures, namely order-m efficiency measures (Daraio and Simar, 2005) and then analyze the ratio of robust conditional and unconditional efficiency called Q^z_m. In this paper we adopt the output oriented framework, in which the goal of the production unit is assumed to be maximizing the quantity of output given the available inputs.

For each unit of observation this ratio takes a numerical value. If the value exceeds one, it means that external factors have an impact on the productive efficiency.

As shown formally in Daraio and Simar (2005), the scatterplots of these ratios and a smoothing nonparametric regression offer an intuitive and useful representation. The plot may exhibit irregular shapes, however, fully taking into account individual cases.

Scatter diagrams that represent the impact of Z can be interpreted as follows (for more details, see Appendix). The vertical axis represents the ratio between conditional and unconditional efficiency. The horizontal axis represents the value of the external variable Z. A line is drawn representing the nonparametric regression of the ratio over the Z variable, and can be read as a local average approximation. In the output oriented framework adopted here, an increasing pattern of this nonparametric regression line points to a *positive* effect of the external factor Z on the performance of the analyzed system of units. A decreasing pattern of the nonparametric regression line points to a *negative* effect of the external factor Z. A straight pattern of the nonparametric regression line indicates *no effect* of the external factor Z. This technique is flexible enough to decompose the conditional efficiency score for each unit of observation. The technique allows also the investigation of the effect of external factors taken separately or jointly. In the latter case it is possible to capture partial and interaction effects observing the 3-dimensional plot.

The main idea of our paper is to measure the knowledge spillover effect as the *combined impact* of innovative and scientific activities on the efficiency of the production process carried out at local (province) level. In order to illustrate the potential of conditional efficiency analysis we use an original dataset, disaggregated at the Italian territorial level of province, built by combining several sources. From the total number of provinces in Italy, 109, we excluded some recently created provinces whose data for the analyzed period (2001-2003) were not available, and a number of provinces that for which some data were not available. Finally, our sample is composed by 92 observations. A description of the variables analyzed in the paper is presented in Table 1.

[3] See Daraio and Simar (2007) for a state of the art presentation of the techniques. Bonaccorsi and Daraio (2007) illustrate the potential of these methods in the microbased analysis of universities.

Role	Variable	Description
Input (X_1)	ULA IND	No. of employees in manufacturing industries – at province level (source: ISTAT - Italian National Statistical Office, *Conti pubblici provinciali*). Average 2001-2003. Thousand units.
Input (X_2)	IP	Proxy of the infrastructural stock at province level (Picci, 2002)
Input (X_2')	DPM	Alternative proxy of the infrastructural stock at province level (Di Palma and Mazziotta, 2002)
Input (X_2'')	KPUB	Share of public capital stock, allocated in proportion to total value added. Year 2001 (Picci, 2002)
Input (X_2''')	KPUB1	Share of public capital stock, allocated in proportion to number of employees (ULA). Year 2001 (Picci, 2002)
Input (X_3')	KPRIV	Share of private capital stock, allocated in proportion to total value added. Year 2001 (Picci, 2002)
Input (X_3'')	KPRIV1	Share of public capital stock, allocated in proportion to number of employees (ULA). Year 2001 (Picci, 2002)
Output (Y)	VA IND	Added value of manufacturing industries, at province level (source: ISTAT - Italian National Statistical Office, *Conti pubblici provinciali*). Average 2001-2003. Million current euro.
External factor (Z_1)	PAT TOT	Cumulate number of EPO patents, 1999-2003. (source: Unioncamere)
External factor (Z_2)	PUB TOT	Cumulate number of ISI publications, produced by universities and other research institutions at province level, 1990-2000. (source: elaboration on ISI data from CRUI)
External factor (Z_3)	PUB TECH TOT	Cumulate number of ISI publications in the Engineering and Technology fields of science, produced by universities and other research institutions at province level, 1990-2000 (excluding Construction engineering)
External factor (Z_4)	PAT INT	Patent intensity: No. of EPO patents (PAT TOT) per million of inhabitants
External factor (Z_5)	PUB INT	Total publication intensity: Cumulate number of ISI publications (PUB TOT) per million inhabitants
External factor (Z_6)	PUB TECH INT	Engineering and Technology publication intensity: Cumulate number of ISI publications in the Engineering and Technology fields per million inhabitants, 1990-2000 (excluding Construction engineering)

Table 1. Variables used in the models

In Tables 2 and 3 descriptive analyses of inputs, outputs and external factors are proposed.
The province level of analysis allows a fine grained observation of spillover effects. Clearly, the smaller the geographic size of the unit of analysis, the larger the spillover from activities that are not located in the same area, but in the close neighborhood.
In order to take into account this possibility we model spatial dependence in patents and publications by assigning to each province its own number plus the numbers of all other provinces weighted by the ratio one divided by the distance (in km). This operationalization is consistent with most literature on geographic spillovers.

Table 2 Descriptive statistics on inputs and output

	Range	Minimum	Maximum	Mean	Std. Deviation
ULA IND	509	5	514	54	67
IP	598	36	634	100	83
DPM	156	27	184	96	32
K PUB	66301	1639	67940	10219	10025
K PRIV	106026	1380	107406	9735	13465
K PUB1	63907	1859	65766	10204	9517
K PRIV1	98119	1555	99675	9717	12693

VA IND	32052	190	32242	2692	3940

Table 3 Descriptive statistics on external factors

	Range	Minimum	Maximum	Mean	Std. Deviation
PAT TOT	3927	14	3941	218	453
PAT INT	1301	22	1323	275	261
PAT TOT AVE	803	3	806	46	93
PAT INT AVE	270	4	274	58	55
PUB TOT	89486	654	90140	8384	14465
PUB INT	70904	787	71691	10150	12826
PUB TOT AVE	10879	77	10956	1003	1717
PUB INT AVE	8720	96	8816	1221	1513
PUB TEC TOT	3143	19	3163	276	489
PUB TEC INT	2665	20	2685	313	440
PUB TEC AVE	286	2	288	25	44
PUB TEC INT AVE	243	2	245	29	40
COMP MAN	2.381	0.041	2.422	0.853	0.464
COMP TOT	0.528	0.005	0.533	0.198	0.130
APE TOT	1.255	0.020	1.274	0.373	0.231

We study how the availability at province level of several sources of knowledge influences the efficiency of production activity of the manufacturing industry. We consider units of labor and physical infrastructure as inputs, and industrial value added as output. We condition the efficiency of production process to the operation of sources of spillovers: industrial innovation activity, as measured by patent applications; and research activity, as measured by international scientific publications. These two sources of knowledge can be read, with some approximations, as coming from the private and the public sector, respectively. The estimation strategy is hence to measure the knowledge spillover as the combined impact of these sources of knowledge on the technical efficiency of production.

Following the literature, we select the manufacturing industry as the main target for knowledge spillover flows, following the literature, although it is increasingly recognized that the role of tertiary activities in growth is not negligible (see for example Kay, Pratt, Warner, 2007).

The main interest is not developing new theories of knowledge spillovers, but rather to leverage on existing contributions (mainly on specific channels for knowledge flows) and estimate a global effect at a sufficiently disaggregated territorial level to capture all relevant factors. Most literature on spillovers deals, due to the omnipresence of data constraint, with data at regional level. At this aggregation level, however, it is difficult to ascribe estimated effects to the channels usually discussed in the literature, because the latter have to do with various forms of labour mobility and personal interaction, which mostly take place at infra-regional level. We believe the level of province is the most appropriate one.

Also, we use for the first time data on publications at province level, allowing to finely identify the effect of local presence of universities and research centers. For these data we take a longer time series (1990-2000) than for patents (1998-2003), following the idea that the impact on productivity and growth of scientific research has a longer time lag than industrial research, as witnessed by patents.

Using the variables described in Table 1 we estimate several simple models with an exploratory purpose. In particular, as proxy of the infrastructural stock we use the IP stock proposed by Picci (2002), in which the permanent inventory technique was applied to reconstruct the infrastructural index of the province. Alternatively, we compare the results obtained by using the DPM infrastructural index, proposed by Di Palma and Mazziotta (2002), which is based on the stock in physical terms. The obtained results (using IP or DPM as proxy of the infrastructures of the provinces) are very similar, hence in the following we report mainly those obtained using as input the IP index.

Descriptive plots of patent activity volume versus patent intensity and of publication activity volume against publication intensity are reported in the following figures.

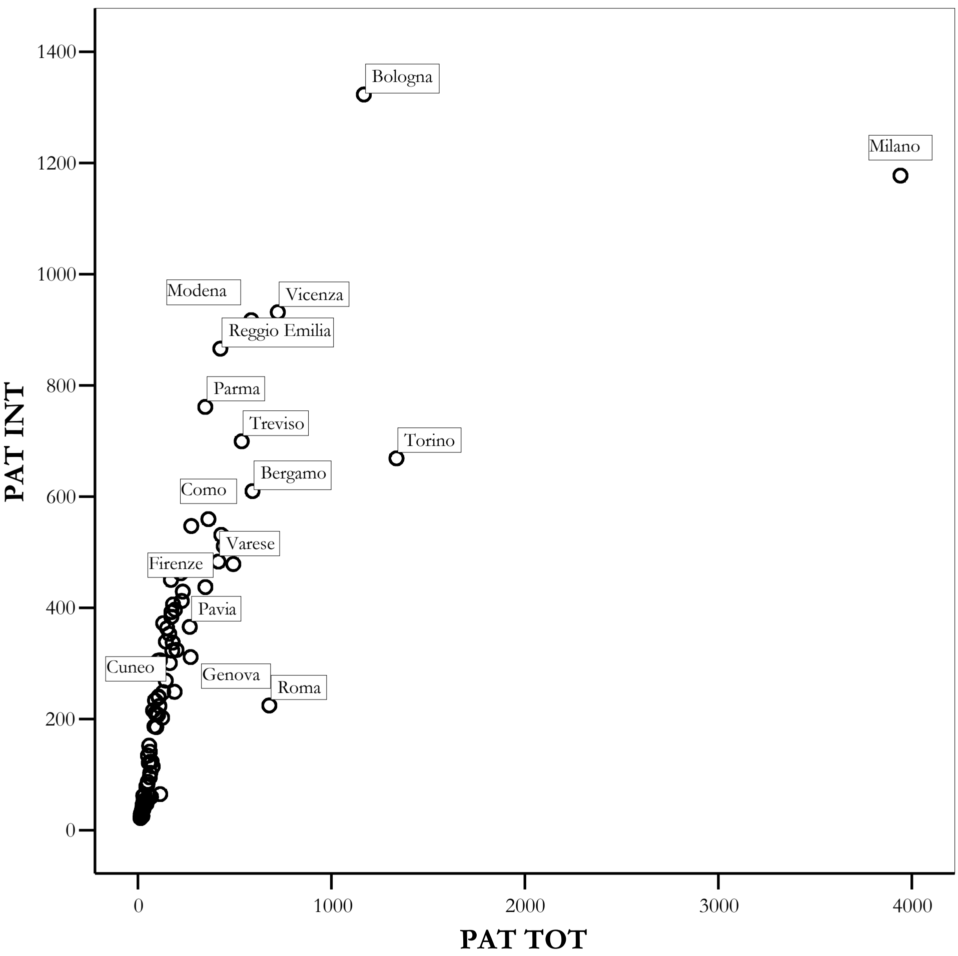

Figure 1. Plot of Patent size (PAT TOT) versus Patent intensity (PAT INT).

Figure 1 compares the size of inventive activity at province level (patent size, PAT TOT) with the relative importance of patenting with respect to the population living in the province (PAT INT). One stylized fact is immediately evident: while there are only one large inventive area (Milano) and only two large provinces (Bologna, Torino), there are many provinces where cumulative patent activity is small in size (less than 1,000) but very high in density. These provinces (Modena, Vicenza, Reggio Emilia, Parma, Treviso, Bergamo, Como) are all located in Northern Italy, are medium-sized, and their economy is largely based on a strongly competitive mechanical engineering industry.

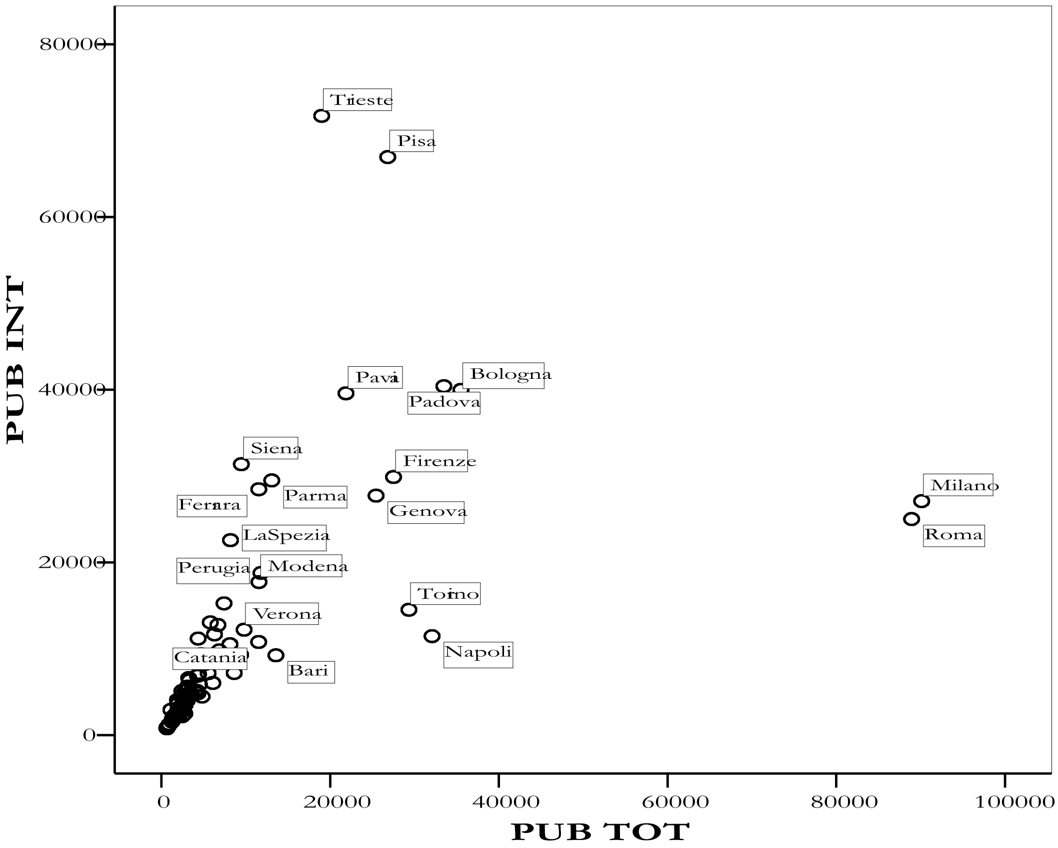

Figure 2 Plot of Publication size (PUB TOT) versus Publication intensity (PUB INT)

A very different picture comes to evidence for publications (Figure 2). Here there are two large agglomerations (Milano, Roma) and a few provinces with a significant volume of scientific activity, centered around large cities (Napoli, Torino, Genova, Firenze, Padova, Bologna). However, the stars in terms of publication intensity are small to medium-sized research oriented cities, such as Trieste, Pisa, and Pavia. These cities are not particularly strong in manufacturing activity, not have an intense patenting activity.

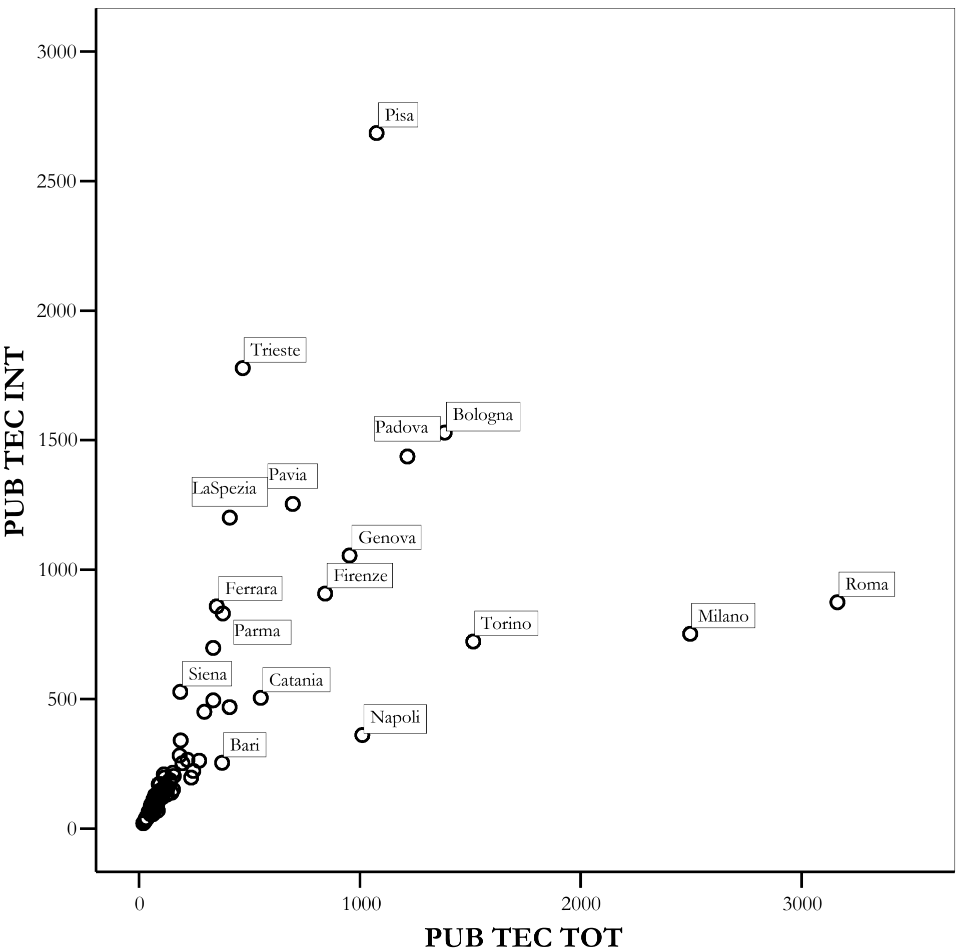

Figure 3 Plot of Total Publication in Engineering and Technology (PUB TEC TOT) versus Publication intensity in Engineering and Technology (PUB TEC INT)

This finding is confirmed, with only a few variations, for publications in Engineering and Technology (Figure 3).

Figure 4 combines the two dimensions of knowledge production in terms of intensity. It is clear that few provinces are specialized in both scientific and inventive activity.

To see this effect, one would usually position each province against the average. But since the variability in intensity is quite large, and the vast majority of provinces fall at small levels, it is better to define quadrants by taking the average value plus one standard deviation (see Table 3 for values). By doing so we see that the fourth quadrant (high high) is almost empty (Mologna, Milano, Parma) while the diagonal is densely populated. In particular, a certain dichotomy is found between provinces rich in science but relatively poor in patents (Pisa, Trieste, Padova, Pavia, Firenze, Siena, Genova, Ferrara) and provinces that follow the opposite pattern.

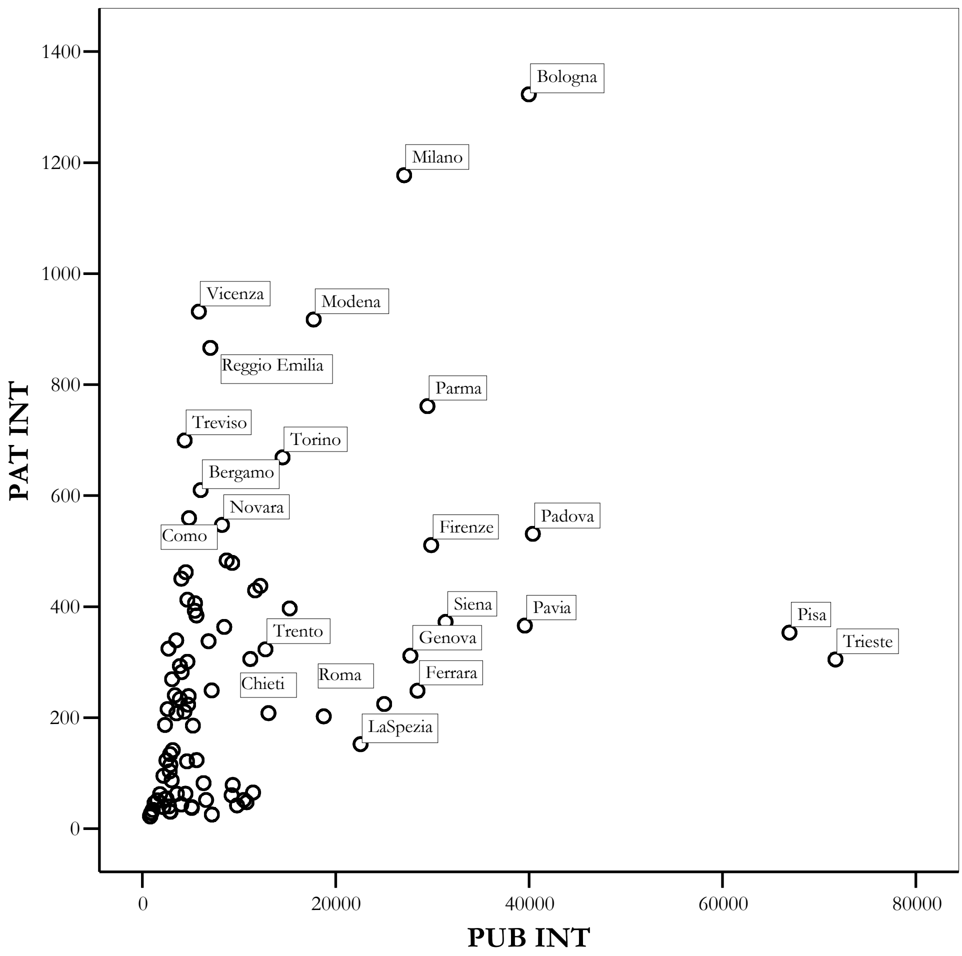

Figure 4 Plot of Publication size (PUB TOT) versus Patent intensity (PAT INT)

Interestingly a similar dichotomic pattern emerges if we focus only on technical publications and plot their intensity against patent intensity (Figure 5). Here, again, the diagonal is populated while the fourth quadrant is almost empty. This is somewhat surprising, given that scientific research in engineering disciplines should have closer relations with industrial applications. It seems that the industry located around Italian engineering schools does not benefit greatly from research activity. This mismatch is likely to have a sectoral explanation, but should be investigated further in the future.

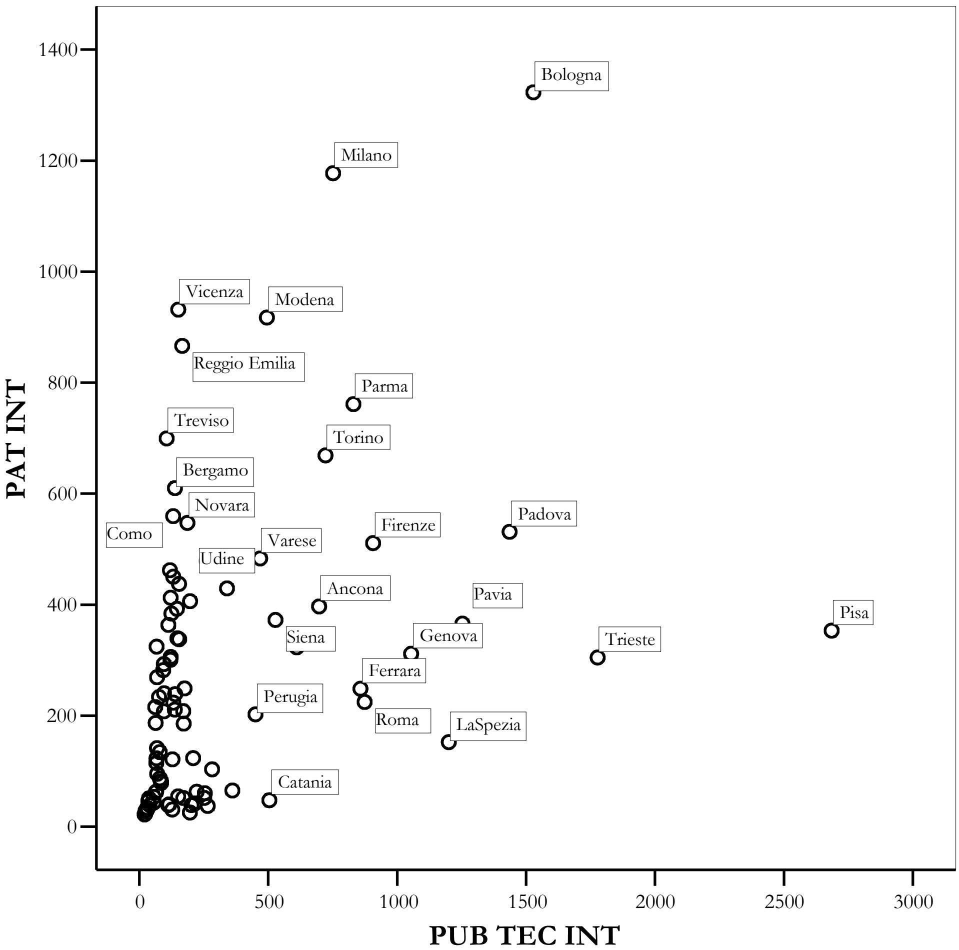

Figure 5 Plot of Publication intensity in Engineering and Technology (PUB TEC INT) versus Patent intensity (PAT INT)

Given this pattern, it is interesting to investigate the existence and magnitude of spillover effects on productive activity.

More specifically, there are several puzzling questions:

- does size matters? Do we see that spillover effects are larger in larger provinces and cities, or beyond a given threshold?
- is knowledge intensity a substitute for knowledge size? In other words, do provinces with high intensity but small size benefit from large spillovers?
- given the dichotomy between patenting and publishing for most provinces, where are spillover larger?
- are publications and patents complementary in producing spillover effects?

These questions are not only interesting for the national case at hand, but more generally for a number of theoretical and policy implications.

Before entering into the estimation exercise, let us introduce Figures 6 to 11, which offer a detailed cartography of knowledge production at province level in Italy.

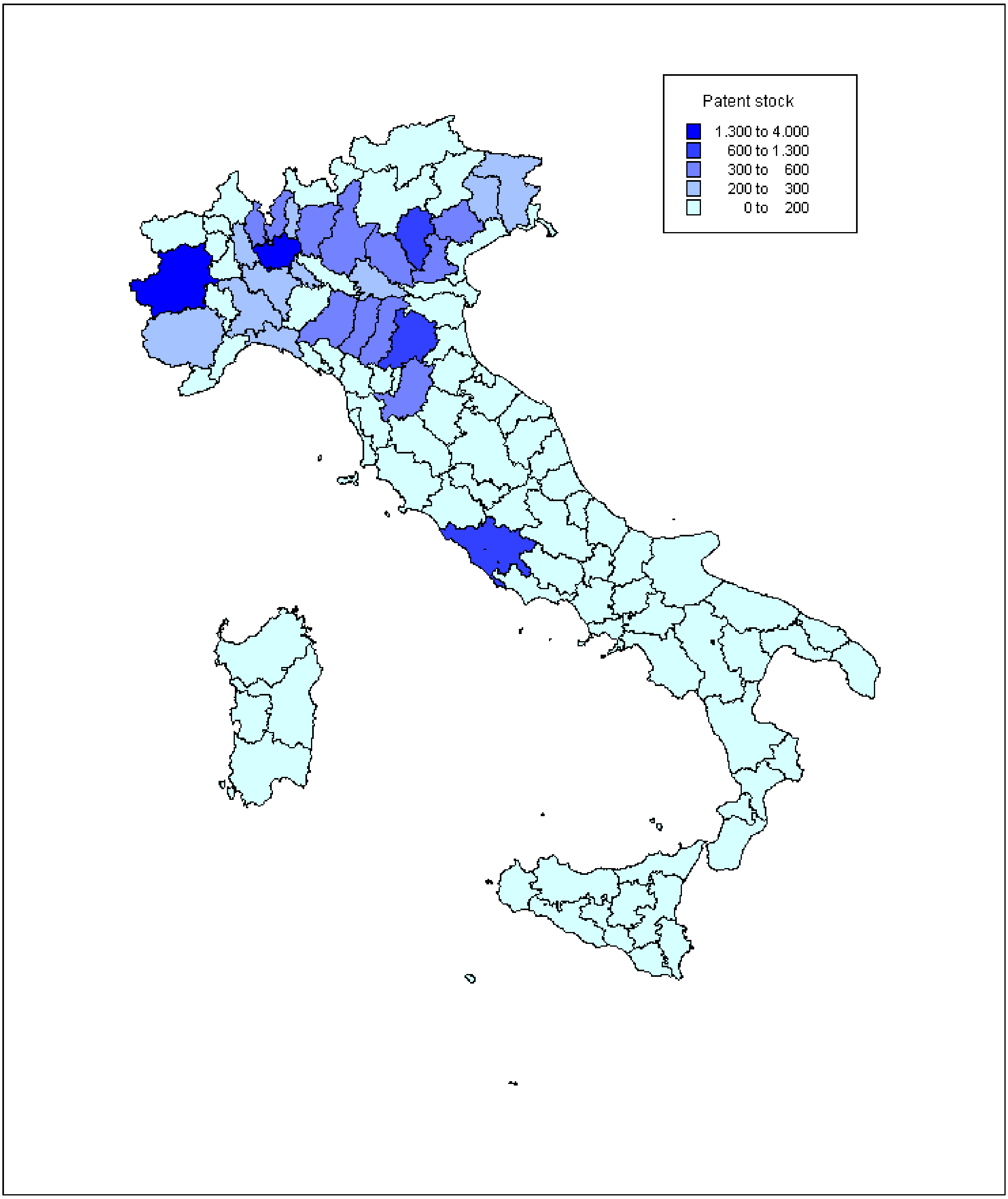

Figure 6.Territorial distribution of patent stock.

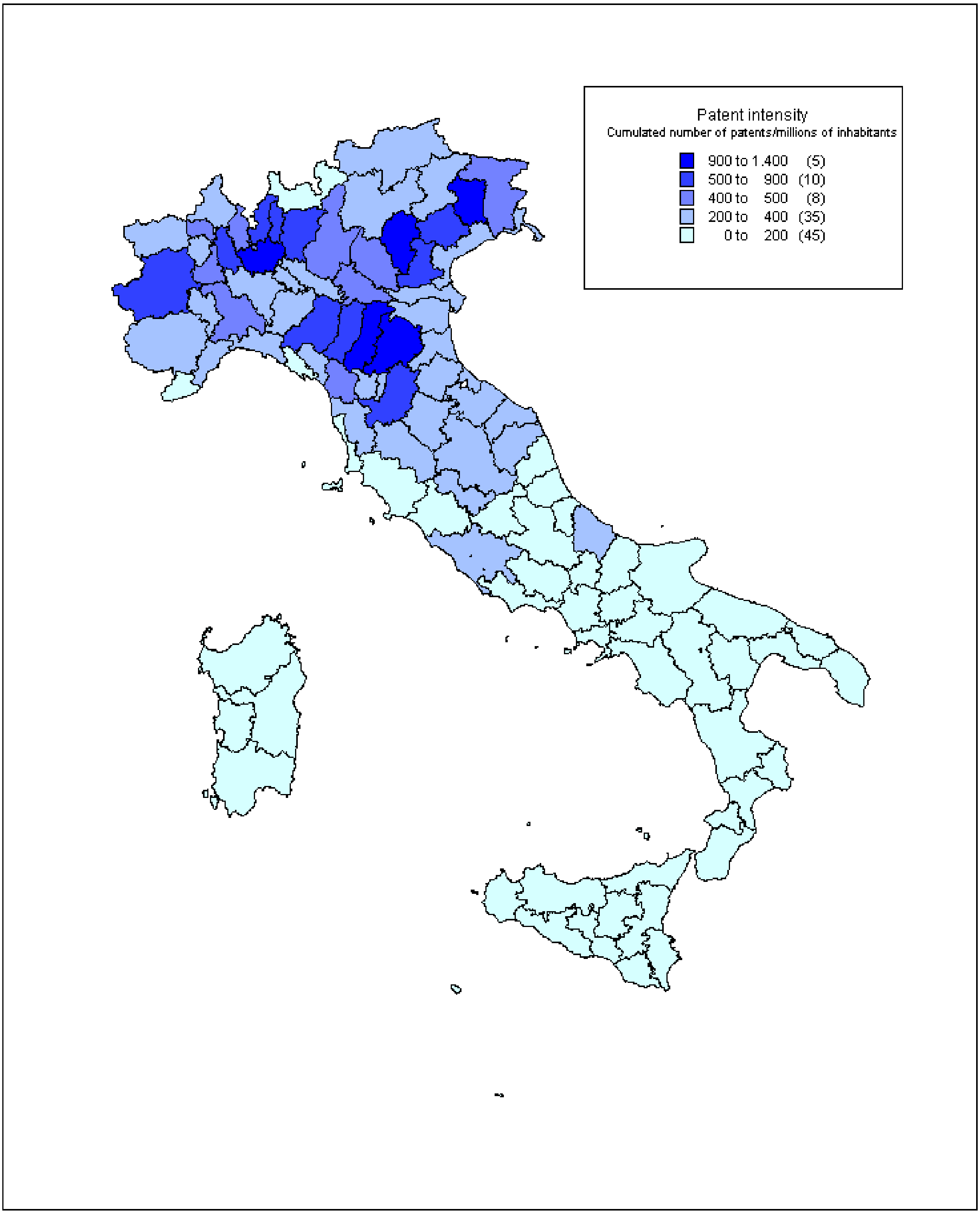

Figure 7 Territorial distribution of patent intensity..

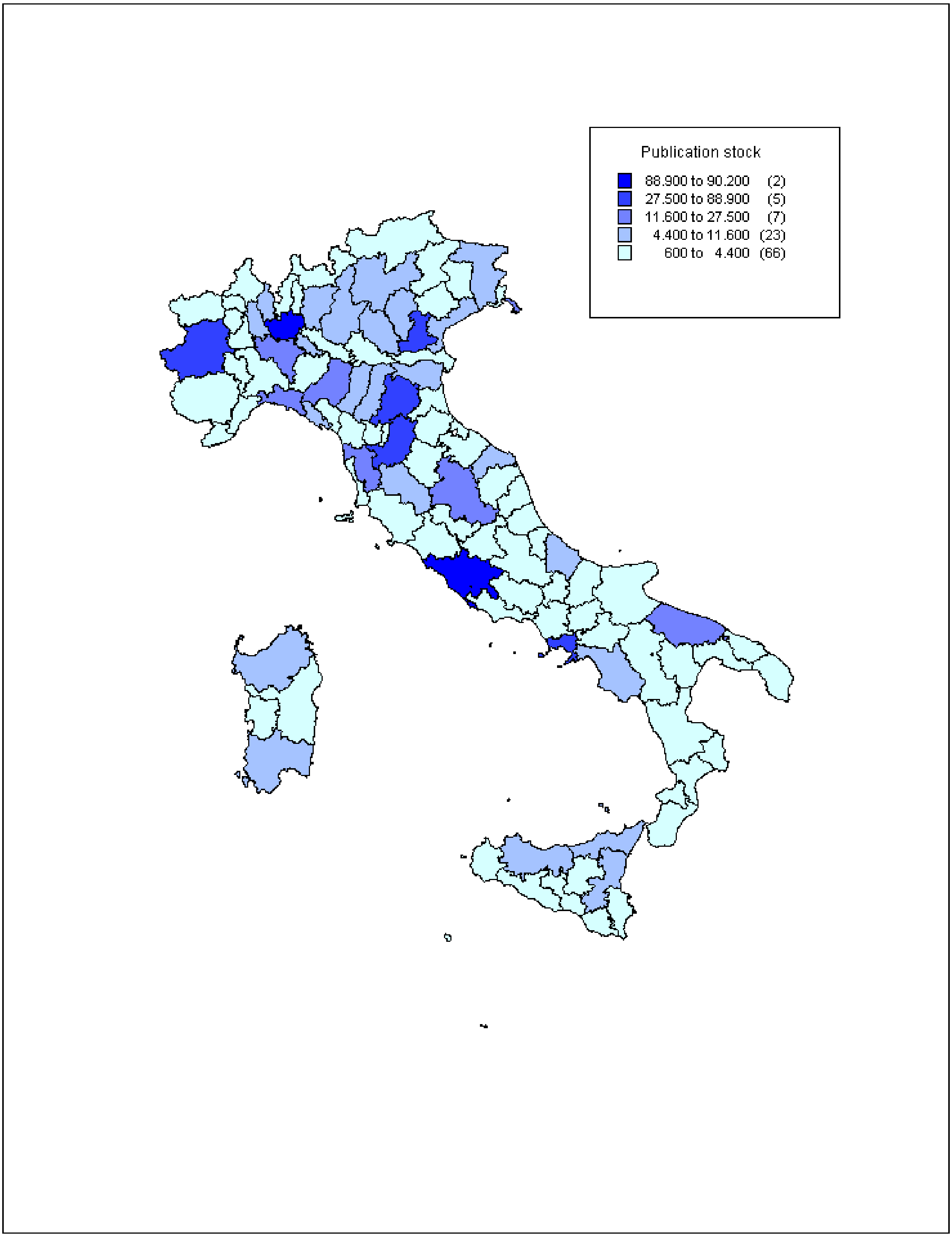

Figure 8 Territorial distribution of publication stock.

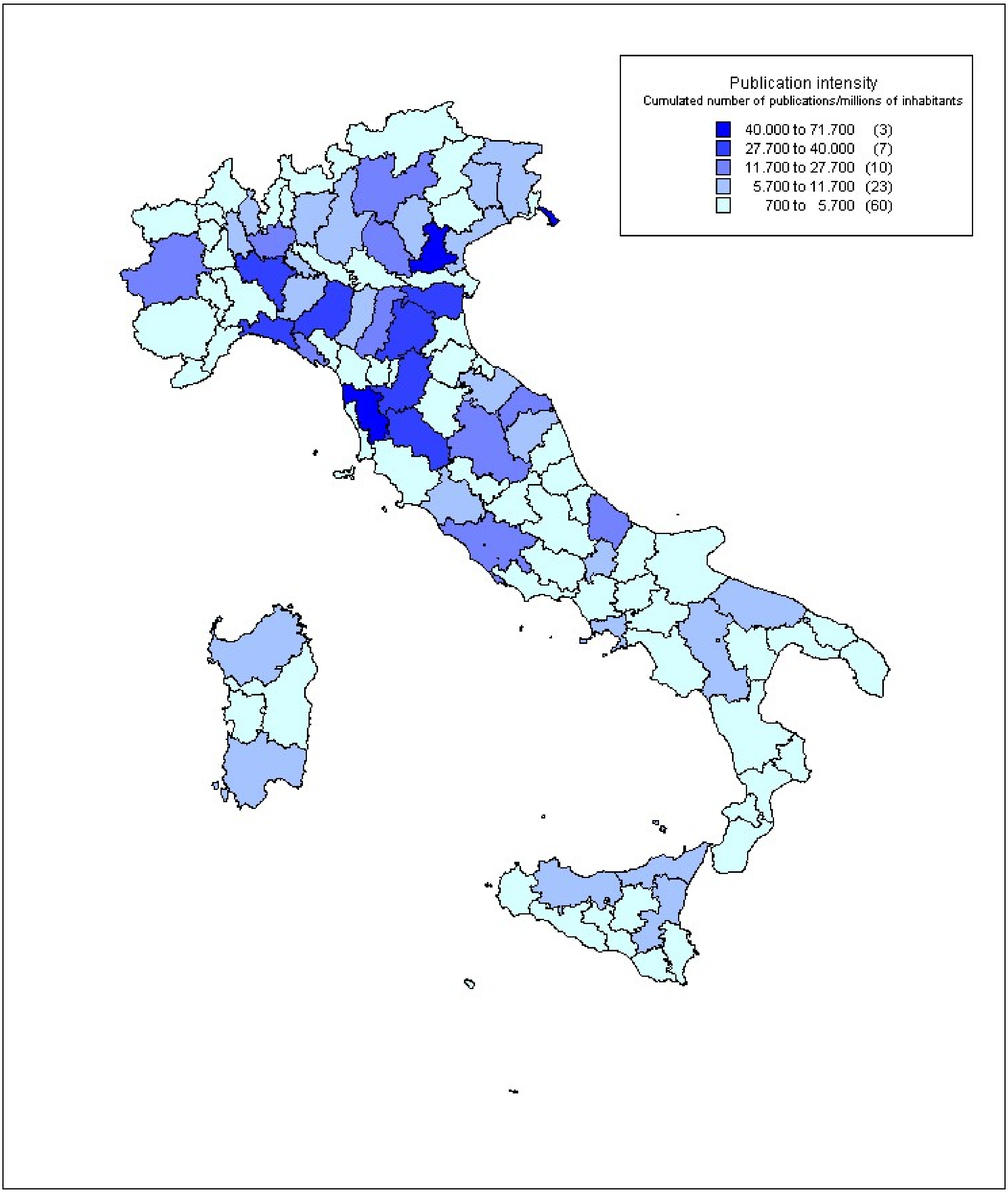

Figure 9 Territorial distribution of publication intensity.

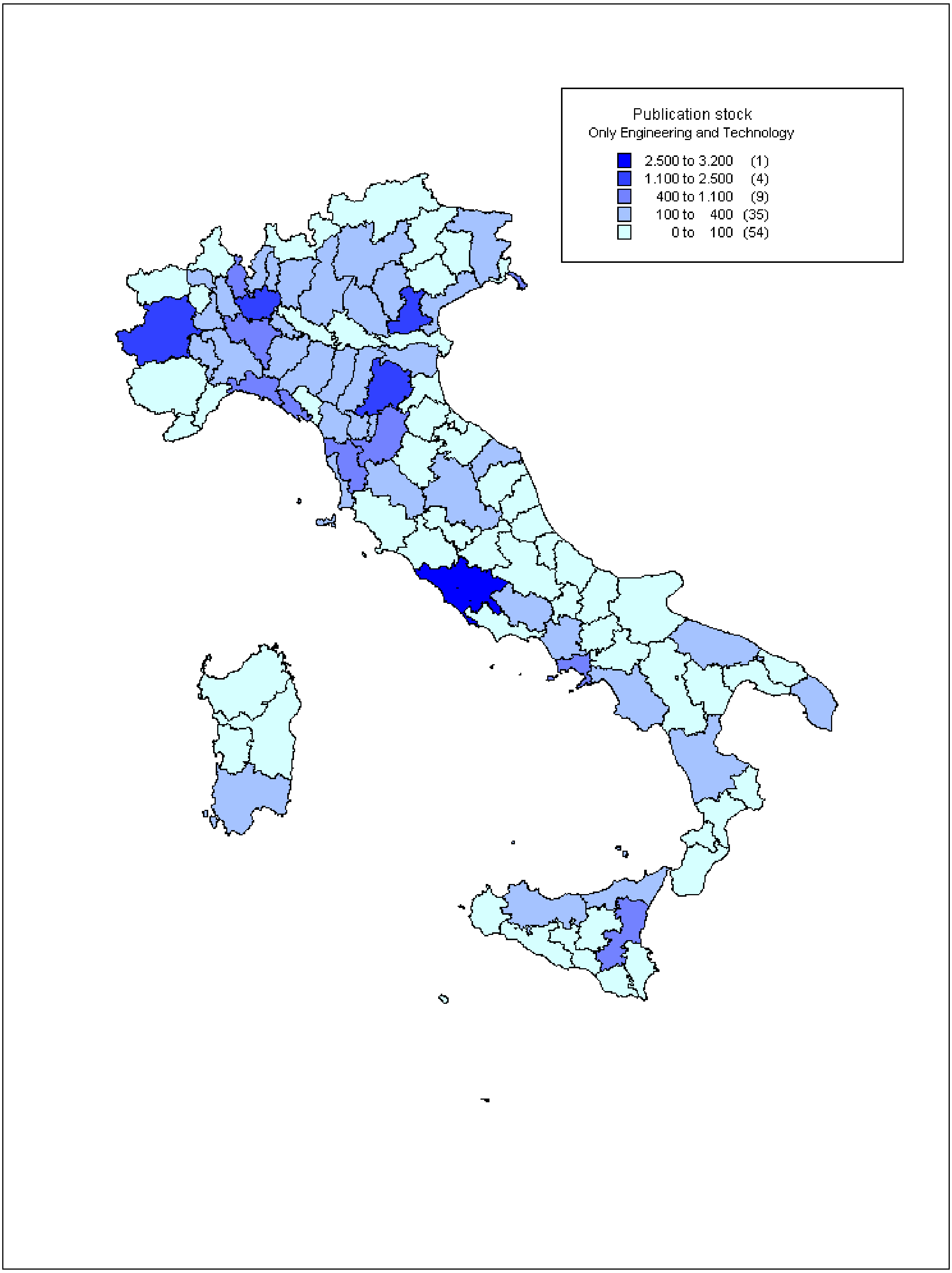

Figure 10 Territorial distribution of publication in Engineering and Technology- stock.

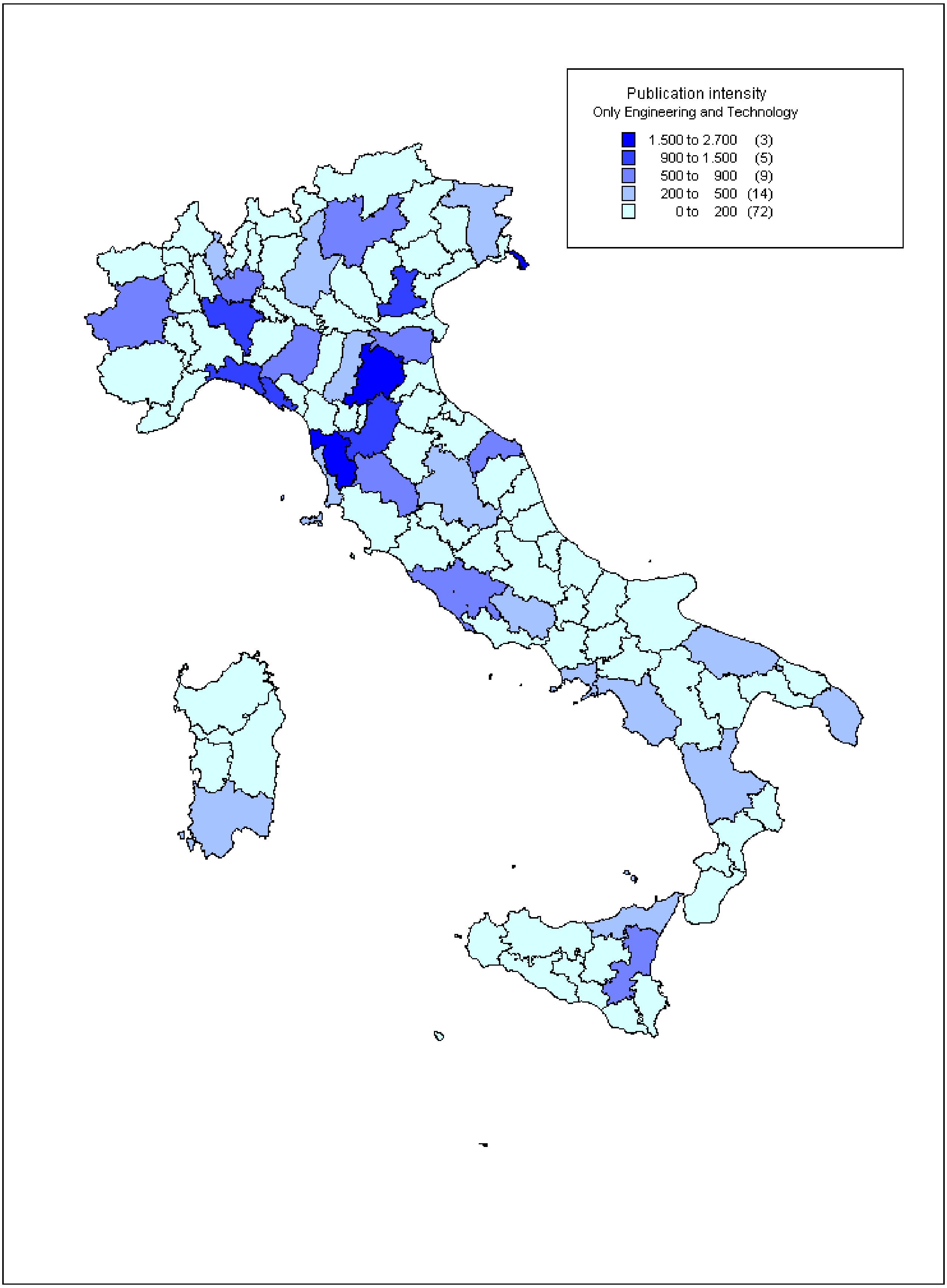

Figure 11 Territorial distribution of publication Engineering and Technology- intensity.

4. Main results

4.1 Size effects

The first group of models introduce external factors in absolute value. In particular we explore how the efficiency of manufacturing is affected by the presence in the same province of the following external knowledge factors:

(a) PAT TOT: cumulate number of EPO patents at province level (year1998-2003);

(b) PUB TOT: cumulate number of ISI scientific publications at province level (year 1990-2000);

(c) PUB ENGTECH: cumulate number of ISI scientific publications in engineering and technology fields (year 1990-2000).

By introducing these variables in absolute terms we are interested in checking size effects, or effects on production activity that may take place as result of the overall size of knowledge production activity, as observable from patents or publications.

We first explore the effects of each external factor separately, limiting to (a) and (b) and examining sectoral effects afterwards.

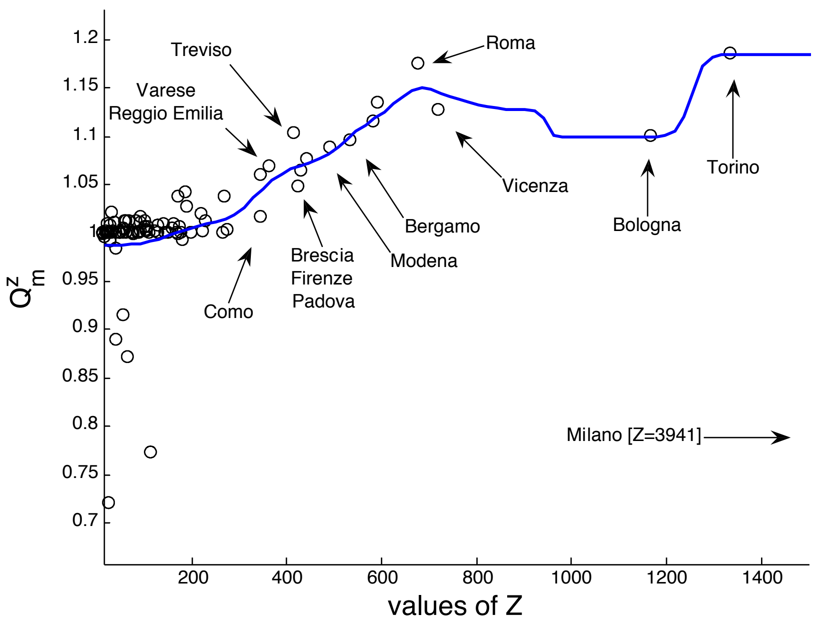

Figure 12. Impact of size of patent activities on production efficiency
Inputs: X1=ULA IND, X2=IP, X3=KPRIV, Outputs: Y=VA IND, External factors: Z= Patent size (PAT TOT)

Figure 12 shows the variation of the ratios of robust conditional and unconditional efficiency measures (Q^z_m) in relation to the total volume of patents at province level. There are several effects of interest. First, as witnessed by the upward slope in the initial region, the overall effect is positive, reaching a peak value of 1.5 for the outlier province of Milan and values beyond 1.05 for many provinces, up to 1.2. Second, most provinces are located in the first region of the plot, in which the total volume of patent activity is small (less than 200 cumulate patents per province) and the ratio takes value 1 or close to 1. This means that for most provinces the spillover effect from patenting activity is negligible. Third, there is a size effect. The ratio increases significantly after the threshold of approximately 200 cumulate patents per province. This is an important finding, that will be discussed below.

Figure 13 examines the effect of the volume of scientific publications at province level. It shows a positive effect, up to the peak value of 1.5, again for the province of Milano and Rome (not shown in the figure). However, most provinces in the positive-sloped region are located around a value of 1.1, which is lower than the one observed for patents. In addition, there seems to be a threshold around 5000 cumulate publications, given that below this value provinces are located in a flat region where Q^z_m is close to 1 (no impact of knowledge). Interestingly, in the interval beyond 15.000 cumulate publications we can find both large university cities (Roma, Milano, Napoli, Torino etc.) but also medium-sized cities with strong university and research activity such as Padova, Pavia, Pisa and Trieste. Figure 4 shows also that the peak value for these provinces is around 1.05-1.1.

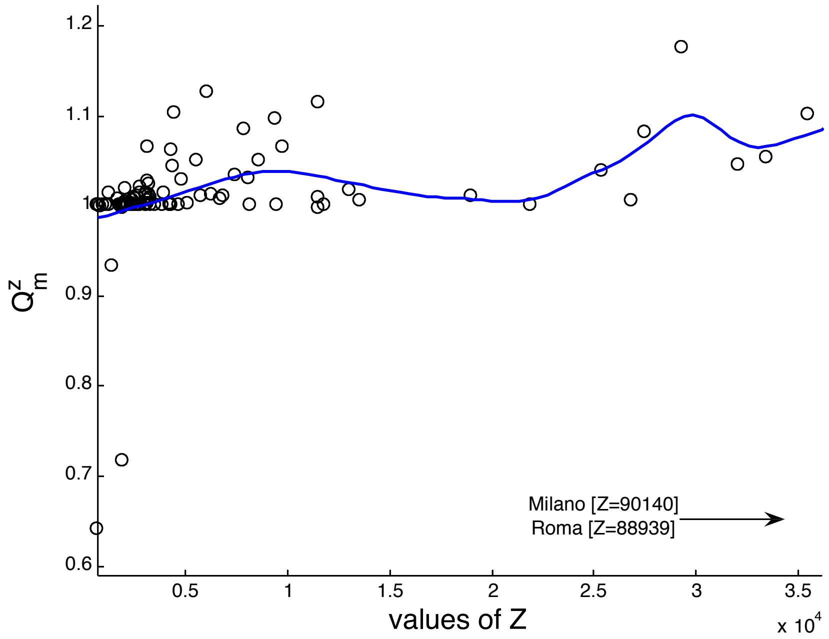

Figure 13. Impact of size of total publication activities on production efficiency
Inputs: X1=ULA IND, X2=IP, X3=KPRIV, Outputs: Y=VA IND, External factors: Z= Publication size (PUB TOT). Below, list of the provinces with PUB TOT higher than 15000.

Province	PUB TOT
Milano	90140
Roma	88939
Bologna	35515
Padova	33491
Napoli	32088
Torino	29353
Firenze	27517
Pisa	26837
Genova	25444
Pavia	21888
Trieste	19007

The examination of separate effects of patents and publications points to the existence of positive but small impact, subject to important threshold effects.
Let us examine the joint impact of the overall volume of patents and publications. This is possible in a 3-d plot, in which the two external factors are represented on the horizontal plane according to their value, and the vertical axis shows the value of the Q^z_m ratios (now the ratios are obtained by dividing the conditional measure of efficiency -in which the external factors are patents and publications- on the unconditional measure).
Figure 14 shows several extremely interesting findings. First of all, while publications and patents taken separately show that the maximum value, at around 1.5, is reached only by the outlier value (Milano) and the rest of the distribution lies in the 1.05-1.2 region, when the external factors work jointly they create a region on top of the hyperplane, where the overall ratio is around 1.4. This suggests a strong complementarity effect between the knowledge generated in the public research system and the technological knowledge incorporated in the inventive activity in companies. Second, the shape of the hyperplane suggests, again, that a certain threshold is at work. As a matter of fact the value of Q^z_m ratios drop sharply when provinces leave apart from the maximum absolute values in both publication and patent activities.
As showed by Daraio and Simar (2007), the conditional technique permits also to investigate the effect of each of the external factors when the other is fixed at a certain level. For this purpose, it may be useful to define the fixed values in terms of quartiles.

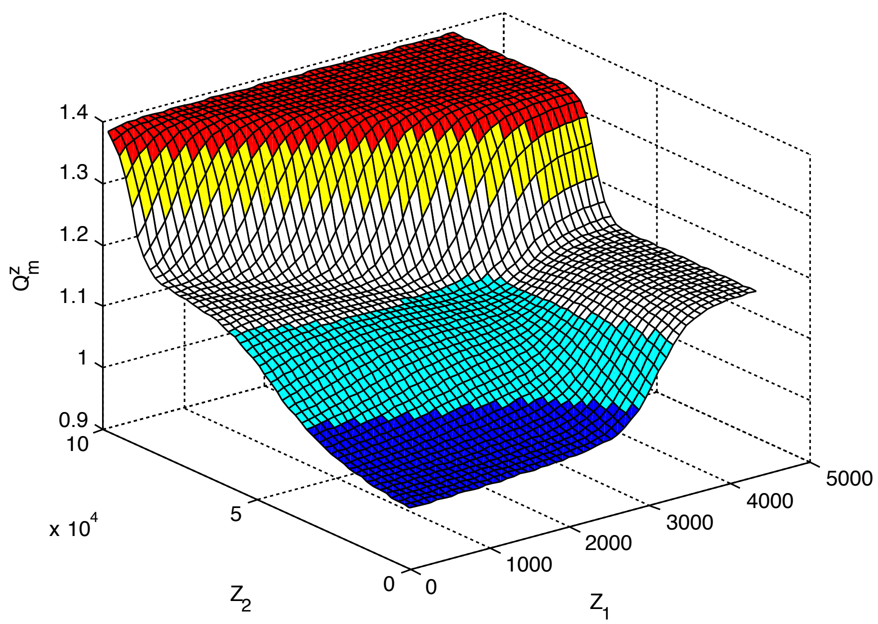

Figure 14 Combined effect of publications and patenting activities on the productivity performance. Surface of QZm on Z1 and Z2.
Inputs: X1=ULA IND, X2=IP, X3=KPRIV, Outputs: Y=VA IND, External factors: Z1= Patent size (PAT TOT), Z2=Publication size (PUB TOT) (m=35, alpha=0.97)

Figure 15 shows in the top panel the variation of the Q^z_m ratios when the external factor Z1 (PAT TOT) is allowed to vary in its interval, for each quartile of the distribution of the external factor Z2 (PUB TOT).
The picture shows that there is not much difference in the effect on efficiency of the variability of patents according to the level of publications (the three lines are close to each other in the space). Confirming the finding from previous figures, the top panel shows that a jump in the 92 Q^z_m ratios is obtained only beyond a large number of patents per province, approximately a cumulate value of 2000, meaning that very few provinces benefit from the external factors.

The bottom panel shows a different story, which is very interesting. If we allow the number of publications to vary along the horizontal axis, there is a positive but small effect in the efficiency of production if there are few patents (bottom quartiles, shown as the almost flat lines at the bottom of the figure). Interestingly, if there are instead many patents, then the overall impact of publications is very strong, and the ratios jump from 1.2 to 1.4.

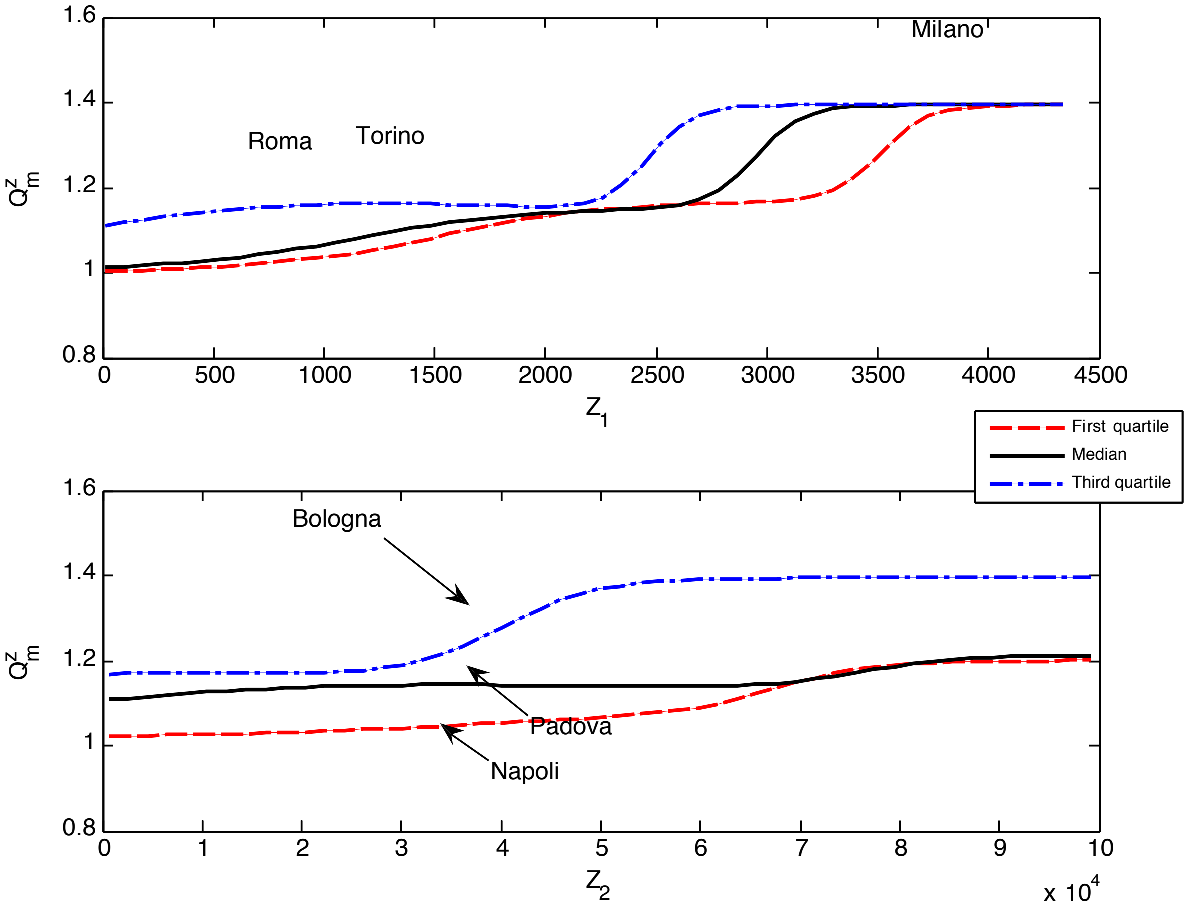

Figure 15 Combined effect of publications and patenting activities on the productivity performance Inputs: X1=ULA IND, X2=IP,X3=KPRIV, Outputs: Y=VA IND, External factors: Z1= Patent size (PAT TOT), Z2=Publication size (PUB TOT). Top panel: smoothed nonparametric regression of Q^z_m *on Z1 for Z2's quartiles. Bottom panel: Smoothed nonparametric regression of* Q^z_m *on Z2 for Z1's quartiles. Dashed line (--) first quartile, solid line (-) median, dashdot line (-.) third quartile.*

This joint analysis strongly confirms the importance of complementarity but also defines more precisely the direction of complementarity: it does not make much difference having more or less publications if the volume of patent increases, while it makes a large difference to have a large volume of scientific activity in a province, depending on whether in the same territory we also have large technological activity as represented by patents.

We find this result as a strong confirmation of the role of absorptive capacity at territorial level. If the local economy and society do not have sufficient accumulation of human capital, organizational capital, and entrepreneurship, the local impact of scientific activity may be negligible. There is no reason, in principle, why local research activity should benefit exclusively the local economy.

4.2 Intensity effects

The previous analysis has shown that the volume of public and private knowledge production in a given territory influences positively the technical efficiency of the industrial system located in the same territory. Let us address here a different question- whether a positive spillover effect is visible once we do not consider the absolute volume of knowledge production, but its intensity, that is, considering the relation between the volume and the underlying population.
As a measure of intensity we consider the ratio between the number of patents and publications and the total resident population, i.e. patents per million of inhabitants (PAT INT), and publications per million of inhabitants (PUB INT). This gives a rough measure of the relative importance of knowledge production over the general social and economic activity of a territory[4].
The difference between size and intensity effects is particularly important in the case of small provinces in terms of population and value added, in which however a large concentration of scientific and technological activity takes place. Do these territories produce knowledge spillovers of the same magnitude of large cities?
Figure 16 examines the effect of a single external factor, the intensity of patents over the population. The effect is clearly positive, reaching a peak value of 1.4.

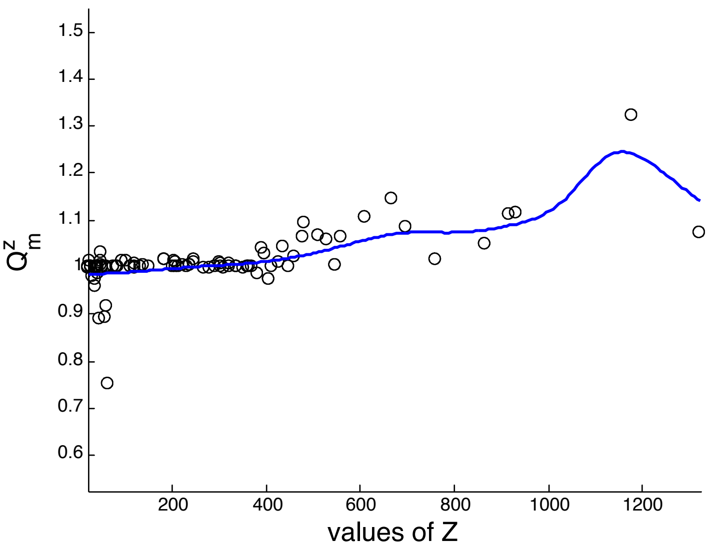

Figure 16. Impact of the intensity of patent activities on production efficiency

Figure 17 shows the effect of total publication intensity. The overall effect, as represented by the fitting line, is only weakly positive. The maximum observed value is 1.4 and a few provinces are located around the value of 1.2. The LOESS line, however, is almost flat. This finding suggests, perhaps not surprisingly, that patent intensity is quantitatively more important, on the average, than publication intensity, for the efficiency of the manufacturing system.
At the same time several medium-sized, research-oriented provinces, such as Parma, Siena and Padova are located above the fitting line, suggesting a larger positive effect (although the magnitude is very small).
To our surprise, this finding also holds for publications in engineering and technical fields (Figure 18), whose profile almost completely overlaps with the general type of publications. Again, a small number of very active provinces are located around the value of 1.2, suggesting that local development can indeed be fostered by research intensity in technical fields, but as an exception rather than a rule.

[4] In future developments we will introduce other intensity measures, using GDP or industrial value added.

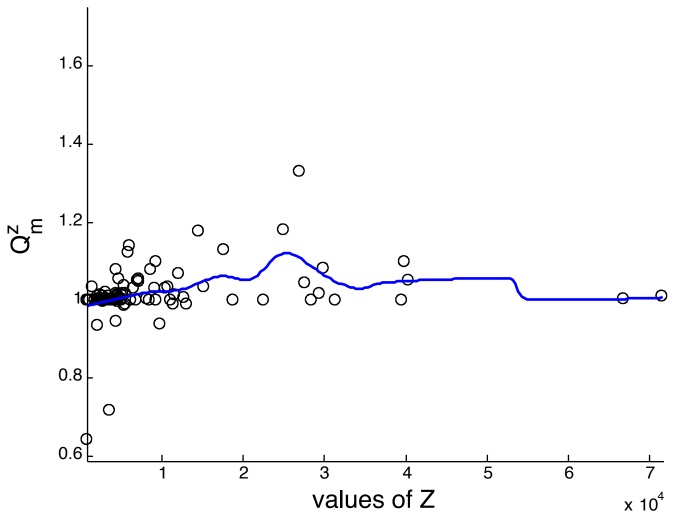

Figure 17. Impact of the intensity of total publications (TOT PUB INT) on production efficiency

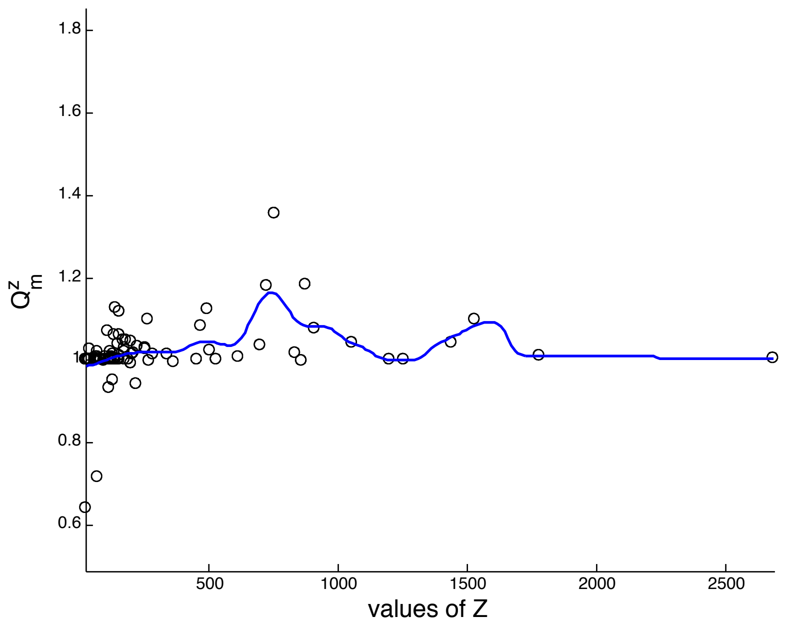

Figure 18. Impact of the intensity of engineering publications on production efficiency

Figure 19 investigates the joint effect of patent and total publication intensity. Surprisingly, here the maximum value of the $Q^z{}_m$ ratios is not obtained when both external variables are at their maximum. Rather, the peak value is located in the region in which publication intensity is at the maximum but patent intensity is at an intermediate level, somewhat below the maximum level. This means that, when the two external factors of knowledge spillover are taken together in terms of intensity, and not absolute value, the threshold effect is smaller. In addition, the region at the top is smaller.

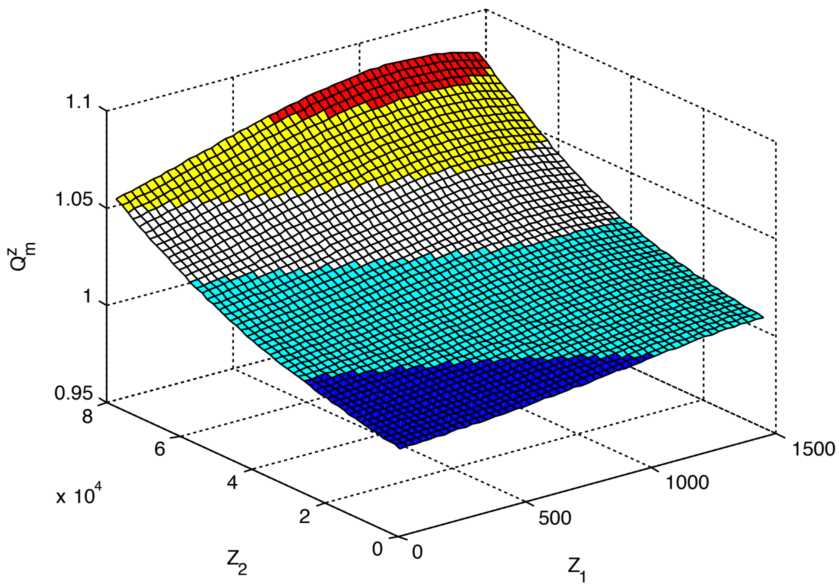

Figure 19. Combined effect of total publications and patenting intensities on the productivity performance. Surface of Q^z_m *on Z1 and Z2.*
Inputs: X1=ULA IND, X2=IP,X3=KPRIV Outputs: Y=VA IND, External factors: Z1= Patent intensity, Z2=Publication intensity (m=35, alpha=0.97)

Figure 20 gives a view of complementary effects. The overall effect of patent intensity is positive for all quartiles of publication intensity (top panel), with few distinctions among quartiles. On the contrary, the impact of publication intensity is moderately positive, but it becomes steady or even slightly negative after a certain threshold, at around 6000 cumulate publications per million of inhabitants.

Summing up, we find an overall positive effect of publication and patent intensity, but this effect is smaller than the one observed for the volume of knowledge spillover. The two effects are complementary. This means that territories in which the total volume of knowledge production is limited but the intensity is strong can exploit knowledge spillovers.

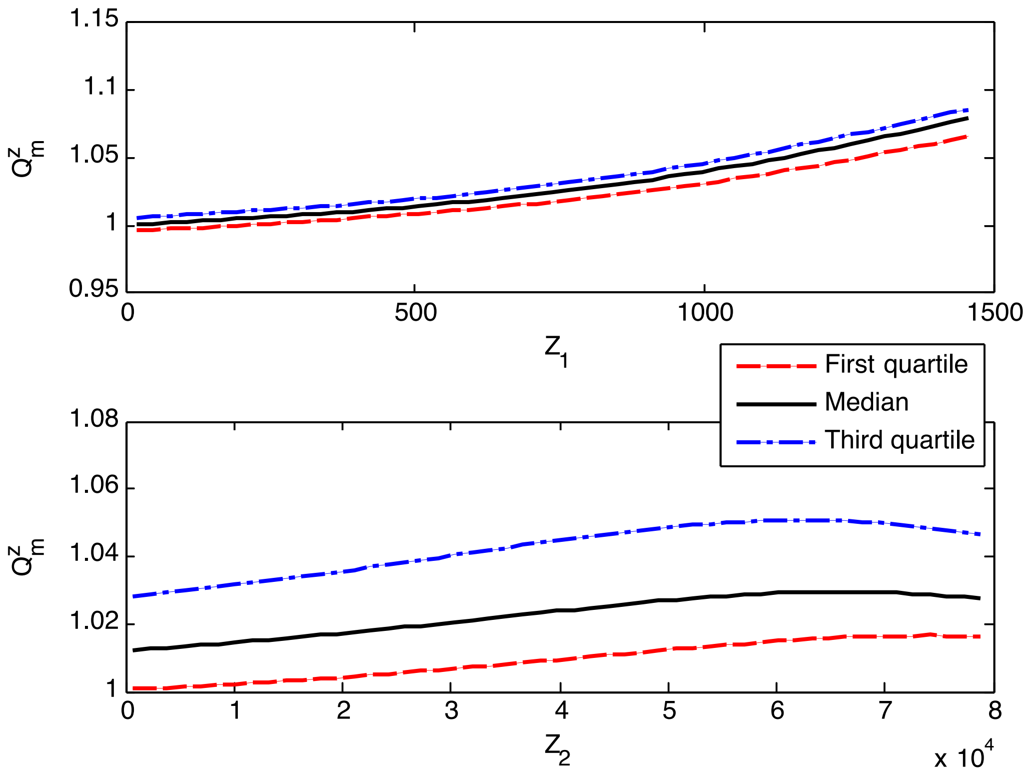

Figure 20. Combined effect of total publications and patenting intensities on the productivity performance Inputs: X1=ULA IND, X2=IP, Outputs: Y=VA IND, External factors: Z1= Patent intensity, Z2=Total publication intensity. Top panel: smoothed nonparametric regression of $Q^z{}_m$ *on Z1 for Z2's quartiles. Bottom panel: Smoothed nonparametric regression of* $Q^z{}_m$ *on Z2 for Z1's quartiles. Dashed line (--) first quartile, solid line (-) median, dashdot line (-.) third quartile.*

4.3 Specialization effect

Finally, we are interested to understand whether the production of knowledge is subject to specialization effects. In particular, we investigate the effect of scientific publications in engineering and technical fields together with patent activity. For the sake of brevity we omit the figures (available from the authors) and report only the findings.
Basically the impact of engineering publications intensity (PUB ENGTECH INT), when associated to patent intensity, is similar to the general case. This suggests the notion that the bulk of effect from scientific activity in a give territory is captured by engineering and technical fields. This is confirmed by the joint effect, whose shape closely follows the one found for all publications. Summing up, it seems that the knowledge spillover from engineering and technical fields follows the same pattern of general publications.

4.3 Summary of findings and policy implications

We have investigated the magnitude and distribution of knowledge spillovers from the overall volume of patenting and publication activity, from the intensity of this activity, and from the sectoral specialization of publications.
The introduction of conditional efficiency analysis has offered a powerful tool to examine the spillover effect without including knowledge as a separate input, given that knowledge is at the same time an input and an output.

We have found strong positive size effects and moderate positive intensity effects. In both cases patenting and publications are strategic complements, but the effect is stronger in volumes than in intensity.
Provinces with maximum volume in both patents and publications exhibit the strongest spillover effect. Alternatively, medium-sized provinces with high knowledge intensity also witness knowledge spillovers, but with less magnitude.
This new technique, therefore, permits to identify and quantify several important dimensions of spillover.
First, large metropolitan areas in which both publication activity and patenting are intense and exceed a given threshold benefit from the largest spillover. These areas are the only ones candidate for real agglomeration effects. Agglomeration requires both size and intensity effects. This finding offers a key to understand the massive process of geographic concentration of innovative activities observed in Europe. To be attractive for public research and for industrial R&D a territory must exhibit *both* large absolute volume of activity and high intensity. The size effect constitutes a large talent pool in the labour market and promotes the division of innovative labor. The intensity effect ensures that the social and economic system are conducive to innovation, since the administrative and political system, the interest groups, the cultural forces will be influenced by the relative strength of research activities.
Second, territories that lack a volume effect may still benefit from knowledge spillovers, but on a smaller scale. This effect is unlikely to generate agglomeration and attractiveness. Provinces with high research intensity benefit from a small to moderate spillover to industry, but this effect is conditional on a minimum level of activity in patents. Provinces with a strong industrial base are those that benefit more from the abstract and general knowledge produced at universities. This is confirmed by specialization effects: the spillover is mainly generated by scientific activity in engineering. This is usually associated with large and/or research-oriented engineering faculties, that have a positive effect on the quality of graduates in local labor markets.
Third, spillover effects disappear at low levels of scientific and technological activity. Most provinces are located in the flat region of the conditional efficiency ratios, suggesting the absence of or extremely low impact of knowledge spillovers.
These findings have strong policy implications. In recent times many local and regional governments have invested into innovation policies with the implicit or explicit goal of creating conditions for local agglomerations of high tech activities, usually associated with attraction of foreign investment, or so called territorial marketing. Our findings suggest that these policies may be totally misleading.
There are only three appropriate policies that try to link knowledge spillovers with an impact on local development:
(a) concentrate innovative investment into large metropolitan areas in order to be able to compete internationally in the battle for attractiveness of public and private research locations and large talent pools;
(b) consolidate public investment in medium-sized research oriented territories, and at the same time invest heavily into the complementarity between public research and technological private activity in order to maximize the spillover;
(c) do nothing. Or, better: consider that the university systems is mainly aimed at creating generic good quality human capital and capabilities, without significant spillover from research and even less attraction and agglomeration effects.

Our findings suggest that public resources are misallocated if governments try to obtain goals (a) in those territories that, in reality, are in position (b), and try to obtain goals (b) in those territories that, at a better exam, are in position (c). Even worse, often regional and local governments claim they can reach goals in (a) when they are barely at (c).
It must be clear that our results do not assume that the rationale for creating universities or research centers is local spillovers. Scientific and higher education institutions have, by nature, a national and international scope. They have enormous value.

Public investment should be directed to maximize spillovers that can be realistically expected given the size and intensity of both scientific and patenting activity. Trying to change the territorial trajectory is of course legitimate, but the resources needed must be made consistent with long term changes.

5. Conclusions and future developments

The issue of knowledge spillover has received a crucial role in several theories of innovation, growth and geography. Despite the large theoretical importance, the identification and measurement of spillovers has proved elusive. We suggest that one reason for this state of affairs is a methodological weakness of currently used econometric techniques. In particular, the widespread use of production functions and related regression-based estimation techniques do not permit the identification of local effects and of external effects.
Local effects may take place because the impact of knowledge on productivity depends very much on several contextual factors, so that any average estimation (for example across firms, industries, or territories) does not capture the inherent heterogeneity. In practice, we have seen that the impact of knowledge spillovers is of little magnitude on the aggregate, but exhibits large differences across units of analysis. Capturing the average effect via regression analysis does not allow the examination of individual cases, that are of the largest interest for the literature and policy makers. Conditional efficiency permits the calculation of spillover effects attached to each units of observation, opening room for a rich interpretation.
Furthermore, the problem of spillover is subject to a dilemma in the econometric specification. Either we include some proxy of knowledge (R&D expenditure, R&D stock, knowledge stock and the like) directly in the equations, or we derive spillover effects indirectly (for example, by computing the difference between private and social rates of return). In the former case the notion of spillover is equated to the cross-unit elasticity, introducing a strong (but unwarranted) causality assumption.
Including knowledge as an input means assuming that the level of production (value added) is determined, inter alia, by the level of investment into knowledge, whatever the specification. But knowledge is at the same time an input to the production process and an output. More generally, we do not know whether knowledge acts mainly on the input or the output side. In particular, if R&D investments are decided by companies following financial accounting rules (such as x% of turnover per year) then an endogeneity problem applies.
In the latter we are left with an order of magnitude of spillover effects, which may be of great interest but leaves the econometric problem unsolved.

The introduction of robust conditional efficiency analysis addresses some of these problems and offers an attractive alternative. By eliminating any assumption about the shape of production function, nonparametric efficiency analysis offers more flexibility. By producing scores of conditional efficiency at individual level, it makes it possible to examine local effects, and at the same time to compute aggregate spillover effects. By placing proxies of knowledge as external factors, it does not ask to make assumptions about the role of knowledge as factor of production or input, as opposed to a joint product of productive activity or output.
We have shown how this flexible tool allows the exploration of some of the most debated issues in the literature on knowledge spillovers.
This technique opens the way to a number of future developments. First, we might extend the geographic scope of the analysis at European level (NUTS 2 or NUTS 3). Second, an extension of the model to tertiary activities would be an interesting development. Third, the external efficiency analysis might be combined to other spatial dependence models, introducing as external factors measures of geographic proximity with more complex formulae than the simple one used for this paper. After these extensions are tested, an interesting area of research would imply including

robust efficiency estimates in input-output models of the economy, allowing for a direct estimate of spillovers. This would be a great advancement in the literature.

References

Abreu M., Florax R.J.G.M., de Groot H.L.F. (2005) Space and Growth. Tinbergen Institute Discussion Paper, TI 2004-129/3

Adams J. D. (1990) Fundamental stock of knowledge and productivity growth, *Journal of Political Economy*, vol. 98, n.4, 673-703

Adams J. D. (1993) Science, R&D and invention potential recharge: US evidence, *American Economic Review*, vol. 83, n.2, May, 458-462.

Aghion P., Howitt P. (1992) A model of growth through creative destruction, *Econometrica*, Vol.60, 323-351.

Aghion P., Howitt P. (1998)*Endogenous growth theory*, MIT Press, Cambridge, Mass.

Aitken B., Harrison A. (1999) Do Domestic Firms Benefit from Direct Foreign Investment? Evidence from Venezuela,*American Economic Review*, 89(3): 605-18.

Almeida P. (1996) Knowledge sourcing by foreign multinationals: patent citation analysis in the US semiconductor industry,*Strategic Management Journal*, 17: 155-165.

Almeida P., Kogut B. (1999) Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, vol. 45, pp. 905-917.

Anselin L. (1988) *Spatial Econometrics: Methods and Models*. Kluwer Academic Publishers, Dordrecht.

Anselin, L., Varga, A., and Acs, Z. (1997): "Local geographic spillovers between University research and high technology innovations", *Journal of Urban Economics* 42, pp. 422-448.

Audretsch D.B., Feldman M.P. (1996) R&D spillovers and the geography of innovation and production. *American Economic Review*, vol. 86, pp. 630-640.
Autant-Bernard C. (2001): "The geography of knowledge spillovers and technological proximity", *Economics of Innovation & New Technology* 10(4), pp237-255.

Bagella M., Becchetti L. (2002) The geographical agglomeration-private R&D expenditure effect: Empirical evidence on Italian data, *Economics of Innovation and New Technology*, 11, no. 3 (June), 233-248.

Barrell R., Pain N. (1999) Domestic institutions, agglomerations and foreign direct investment in Europe, *European Economic Review*, vol 43, pp. 925-934.

Beise M., Stahl H. (1999) Public research and industrial innovations in Germany. *Research Policy*, vol. 28, pp. 397-422.

Bernstein, J. I e Nadiri, M. I. (1989), Research and Development and Intra-industry Spillovers: An Empirical Application of Dynamic Duality, *Review of Economic Studies*, 56, 249-269.

Bernstein, J. I. e Nadiri, M. I. (1988), Interindustry R&D Spillovers, Rate of Return, and Production in High-Tech Industries, *American Economic Review*, 78(2), 429-434.

Blomström M., Kokko, A. (1998) Multinational corporations and spillovers, *Journal of Economic Surveys*, 12: 247-277.

Bonaccorsi A., Daraio C. (2004), "Econometric approaches to the analysis of productivity of R&D systems. Production functions and production frontiers", in H.F. Moed, W. Glanzel and U. Schmoch (edited by), *Handbook of Quantitative Science and Technology Research, Kluwer Academic Publishers*, 51-74

Bonaccorsi A., Daraio C. (2007), edited by, *Universities and Strategic Knowledge Creation. Specialization and Performance in Europe*, Edward Elgar Publisher, Cheltenham, PRIME Series on Research and Innovation Policy in Europe, *forthcoming* July 2007.

Bottazzi L., Peri G. (2003) Innovation and spillovers in regions: evidence from european patent data. *European Economic Review* 47, pp. 687 – 710

Breschi S. (2000) The geography of innovation. A cross-sector analysis. *Regional Studies*, vol. 34, pp. 213-229.

Cantwell J.A., Iammarino S. (2001) EU regions and multinational corporations: Change, stability and strenghtening of technological comparative advantage, *Industrial an Corporate Change*, Vol. 10, pp. 1007-1037.

Carrincazeaux C., Lung Y, Rallet A. (2001) Proximity and localisation of R&D activities, *Research Policy*, vol. 30, pp. 777-789.

Coe, D.T., e Helpman, E. (1995), International R&D Spillovers, *European Economic Review*, 39, 859-887.

Cohen W.M., Nelson R., Walsh J. (2002) Links and impacts: The influence of public research on industrial R&D, *Management Science*, vol. 48, January, 1-23.

Comin D. (2002), R&D? A Small Contribution to Productivity Growth, Economic Research Report, New York University.

Daraio C., Simar, L. (2005), "Introducing Environmental Variables in Nonparametric Frontier Models: a Probabilistic Approach", *The Journal of Productivity Analysis*, 24 (1), pp. 93-121.

Daraio C., Simar L. (2007), *Advanced Robust and Nonparametric Methods in Efficiency Analysis. Methodology and Applications*, Springer, New York.

Di Palma M., Mazziotta, C. (2002), "La dotazione di capitale pubblico in Europa e in Italia: un quadro di ricontri empirici", in *L'Italia nella Competizione Globale - Regole per il Mercato*, edited by Baldassarri M., G. Galli and G. Piga, Edizioni Il Sole 24 Ore, Milano.

Dowrick S. (2003) A review of the evidence on science, R&D and productivity. Paper prepared for the Department of Education, Science and Training, Australian Government, Mimeo, August.

Feldman, M. P. (1994): *The geography of innovation*, Kluwer Academic Publishers, Boston.

Feser, E. J. (2002) Tracing the sources of local external economies. *Urban Studies* 39(13), pp. 2485-2506

Foray D. (2004) *Economics of knowledge*. Cambridge and London: MIT Press

Fosfuri A., Motta M., Ronde T. (2001) Foreign direct investment and spillovers through workers' mobility, *Journal of International Economics*, 53: 205-222.

Frost T.S. (2001) The geography sources of foreign subsidiaries' innovation. *Strategic Management Journal*, vol. 22, pp. 101-123.

Griliches Z. (1979) Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics*, vol. 10, no. 1, pp. 92-116.

Griliches Z. (1992) The search for R&D spillovers. *Scandinavian Journal of Economics*, 94(0), 29-47.

Griliches, Z. e Lichtenberg, F. (1984), Interindustry Technology Flows and Productivity Growth: A Reexamination, *The Review of Economics and Statistics*, 66(2), 324-329.

Gupta A.K., Govindarajan V. (2000) Knowledge flows within multinational corporations, *Strategic Management Journal*, 21: 473-496.

Gwanghoon L. (2005) Direct versus Indirect International R&D Spillovers *Information Economics and Policy* 17(3), pp. 334-48

Jaffe A., Trajtenberg M., Henderson R. (1993) Geographical localisation of knowledge spillovers, as evidenced by patent citations. *Quarterly Journal of Economics*, vol. 58, pp. 577-598.

Jaffe A.B. (1996) Economic analysis of research spillovers. Implications for the Advanced technology Program, http://www.atp.nist.gov/eao/gcr708.htm

Jaffe A.B., Trajtenberg, M., and Henderson, R. (1993) Geographical localisation of knowledge spillovers, as evidenced by patent citations, *Quarterly Journal of Economics*, 58(3): 577-598.

Jaffe, A. (1989) Real effects of academic research. *The American Economic Review*, vol 79, n. 5, pp. 957-970.

Jones C. (1995). R&D based models of economic growth. *Journal of Political Economy* 103, 739–784.

Kay D.L. Pratt J.E., Warner M.E. (2007) Role of services in regional economic growth. *Growth and Change*. Vol. 8, No. 3, September, 419-442.

Kokko A. 1994. Technology, Market Characteristics and Spillovers, *Journal of Development Economics*, 43(2): 279-293.

Kokko A., Tansini R., Zejan M. (1996) Local Technological Capability and Productivity Spillovers from FDI in the Uruguayan Manufacturing Sector, *Journal of Development Studies*, 32(4): 602-611.

Krugman P.R. (1990) *Geography and trade*. Cambridge, Mass.,The MIT Press.

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

Link A.N. (1983) Inter-firm technology flows and productivity growth. *Economic Letters*, vol. 11, no. 2, pp. 179-184.

Mansfield E (1998) Academic research and industrial innovation. An update of empirical findings. *Research Policy*, vol. 26, pp. 773-776.

Mansfield E. (1980) Basic research and productivity increase in manufacturing, *American Economic Review* 70, no. 5 (December), 863-873.

Mansfield E. (1991) Academic research and industrial innovation. *Research Policy*, vol. 20, pp. 1-12.

Orlando M. J. (2004): Measuring spillovers from industrial R&D: on the importance of geographic and technological proximity, *RAND Journal of Economics* 35(4), pp. 777-786.

Ottaviano G., Puga D. (1997) Agglomeration in the global economy: a survey of the new economic geography, CEPR Discussion Papers, n.169.

Picci, L. (2002), "Le infrastrutture in Italia. Le differenze territoriali e l'efficienza della spesa", in *L'Italia nella Competizione Globale - Regole per il Mercato*, edited by Baldassarri M., G. Galli and G. Piga, Edizioni Il Sole 24 Ore, Milano.

Romer P. (1986) Increasing returns and long run growth, *Journal of Political Economy*, 95, 1002-1037.

Romer P. M. (1990) Endogenous technological change, *Journal of Political Economy*, XCVIII, S71-S102.

Rosenthal S.S., Strange W. (2004) Evidence on the nature and sources of agglomeration economies. In Henderson J.V. and Thisse J.F. (eds.) *Handbook of Regional and Urban Economics*, vol. 4, Amsterdam, North Holland.

Saxenian A.L., Hsu J.Y. (2001) The Silicon Valley- Hsinchu connection : Technical communities and industrial upgrading. *Industrial and Corporate Change*, vol. 10, pp. 893-920.

Scherer F. M. (1982) Inter-industry technology flows and productivity growth. *Review of Economics and Statistics*, vol. 64, no. 4, pp. 627-634.

Szulanski G. 1996. Exploring internal stickiness: impediments to the transfer of best practice within the firm, *Strategic Management Journal*, 17: 27–43.

Varga A. (1998) *University research and regional innovation*. New York, Springer.

Wieser R. (2005) Research and development productivity and spillovers: Empirical evidence at the firm level. *Journal of Economic Surveys*, vol. 19, no. 4, pp. 587-621.

Zucker L., Darby M.R. (1996) Star scientists and institutional transformation. Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences*, pp. 709-712.

Zucker L., Darby M.R., Armstrong J. (1999) Intellectual capital and the firm: The technology of geograpically localized knowledge spillovers. NBER Working Paper 4946, National Bureau of Economic Research.

Appendix on robust nonparametric methods in efficiency analysis

In efficiency analysis the main purpose is the study of how firms combine their inputs to obtain their outputs. More generally, in an activity analysis framework (see e.g. Debreu, 1951; Shephard, 1970), the management of a Decision Making Unit (DMU) is characterized by a set of inputs $x \in R_+^p$ used to produce a set of outputs $y \in R_+^q$. The set of technically feasible combinations of (x, y) is defined as:

$$\Psi = \{(x, y) \in I\!R_+^{p+q} \mid x \text{ can produce } y\}.$$

In this setting, the Farrell measure of output-oriented efficiency[5] for a firm operating at the level (*x, y*) can be defined as:

$$\lambda(x, y) = \sup\{\lambda \mid (x, \lambda y) \in \Psi\},$$

where $\lambda(x, y) \geq 1$ is the proportionate increase of outputs a DMU working at the level (*x, y*) should perform to achieve efficiency, and $1/\lambda$ is the Shepard output distance function (Shephard, 1970). The efficient frontier corresponds to those firms where $\lambda(x, y) = 1$. See Figure 1A for a graphical representation of the production set and the situation of a DMU A which produces an output y_A using x_A input. For sake of representation we illustrate a simple univariate frontier, even if an advantage of nonparametric methods is their multi-input multi-output description of production technology. Of course the *true* efficient frontier is not known and has to be estimated using a sample of production observations.

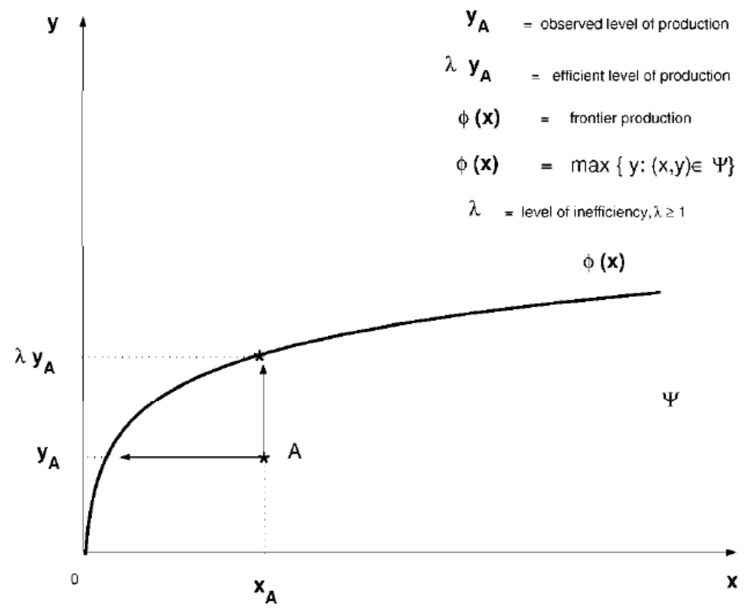

Figure 1A. Production set and efficient frontier: an illustration.

In efficiency analysis, the nonparametric approach is based on envelopment techniques, whose main estimators are Data Envelopment Analysis (DEA, see Farrell, 1957, and Charnes, Cooper and Rhodes, 1978) and Free Disposal Hull (FDH, see Deprins, Simar and Tulkens, 1984). These estimators rely on the idea that the attainable set is defined by the set of minimum volume

[5] To save space, in this appendix we only present the output oriented case that we applied in the empirical analysis.

containing all the observations. The DEA estimator relies on the free disposability[6] and on the convexity of the set Ψ, whereas the FDH relies only on the free disposability assumption. The FDH estimator of Ψ, based on a sample of n observations (x_i, y_i), is the free disposal closure of the reference set $\{(x_i, y_i) | i = 1, ..., n\}$. It can be defined as:

$$\widehat{\Psi}_{FDH} = \Big\{(x, y) \in I\!R_{+}^{p+q} \mid y \leq y_i, x \geq x_i, i = 1, ..., n\Big\}.$$

The DEA estimator of Ψ, is the convex closure of ψ_{FDH}:

$$\widehat{\Psi}_{DEA} = \Big\{(x, y) \in I\!R_{+}^{p+q} \mid y \leq \sum_{i=1}^{n} \gamma_i\, y_i \ ; \ x \geq \sum_{i=1}^{n} \gamma_i\, x_i, \quad \text{for } (\gamma_1, ..., \gamma_n) \quad \text{s.t.} \sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geq 0, i = 1, ..., n\Big\},$$

The estimated output oriented FDH efficiency score of a firm (x, y) is given by:

$$\widehat{\lambda}_{FDH} = \max\{\lambda \mid (x, \lambda y) \in \widehat{\Psi}_{FDH}\}.$$

Similarly, the estimated output oriented DEA efficiency score of a DMU (x, y) is given by:

$$\widehat{\lambda}_{DEA} = \max\{\lambda \mid (x, \lambda y) \in \widehat{\Psi}_{DEA}\}.$$

The nonparametric approach in efficiency analysis offers several advantages, among whose:
- absence of specification of the functional form for the input-output relationship;
- measurement of the efficiency with respect to the efficient frontier which measures the best performance that can be practically achieved;
- appropriate benchmark to be used for comparison: non requirement of any theoretical models as benchmarks;
- production of *multi-inputs multi-outputs* performance indicators.

One of the main drawbacks of DEA/FDH nonparametric estimators is their sensibility to extreme values and outliers in the data.
To overcome this methodological limitation, Cazals, Florens and Simar (2002) propose a nonparametric estimator of the frontier, more robust to extreme values and outliers. It is based on the concept of the *expected maximum output function* of order-*m*.
Extending these ideas to the full multivariate case, Daraio and Simar (2005) define the concept of *expected* order-*m* output efficiency score. This robust approach is based on a probabilistic formulation of the model. The production process is described by the joint probability measure of (X, Y) on $R_+^p \quad R_+^q$. In this formulation, the support of (X, Y) is the attainable set Ψ and the Farrell output efficiency can be characterized, under the free disposability assumption, as:

$$\lambda(x, y) = \sup\{\lambda \mid S_Y(\lambda y \mid x) > 0\},$$

where $S_Y(y \mid x) = \mathrm{Prob}(Y \geq y \mid X \leq x)$. A nonparametric estimator of $\lambda(x, y)$ is provided by plugging in the equation above the empirical version of $S_Y(y \mid x)$ given by:

[6] A set Ψ is free disposal if $(x, y) \in \Psi$ implies $(x', y') \in \Psi$ for any $x' \geq x$ and $y' \leq y$.

$$\widehat{S}_{Y,n}(y \mid x) = \frac{\sum_{i=1}^{n} \mathbb{I}(X_i \leq x, Y_i \geq y)}{\sum_{i=1}^{n} \mathbb{I}(X_i \leq x)}.$$

It has been shown that the resulting nonparametric estimator coincides with the FDH estimator defined above.
The order-*m* output efficiency[7] can be defined as in Daraio and Simar (2005): for a given level of inputs *x* in the interior of the support of *X*, consider *m* i.i.d. random variables $Y_i = i = 1,\ldots,m$ generated by the conditional *q*-variate distribution function $F_Y(y \mid x) = \mathrm{Prob}(Y \leq y \mid X \leq x)$ and define the set:

$$\Psi_m(x) = \{(x', y) \in \mathbb{R}_{+}^{p+q} \mid x' \leq x, Y_i \leq y, i = 1, \ldots, m\}.$$

Then, for any *y*, we may define:

$$\tilde{\lambda}_m(x,y) \quad = \quad \sup\{\lambda \mid (x, \lambda y) \in \Psi_m(x)\}$$

The variable $\tilde{\lambda}_m(x,y)$ is a random variable because the $\Psi_m(x,y)$ is random.
For any y R_+^q the (expected) order-*m* output efficiency measure denoted by $\lambda_m(x,y)$ is defined for all *x* in the interior of the support of *X* as:

$$\begin{aligned} \lambda_m(x,y) &= E(\tilde{\lambda}_m(x,y) \mid X \leq x) \\ &= \int_0^{\infty} [1 - (1 - S_Y(uy \mid x))^m]\, du \end{aligned}$$

It has been shown that $\lambda_m(x,y)$ $\lambda(x,y)$ as m . A nonparametric estimator of $\lambda_m(x,y)$ is given by:

$$\hat{\lambda}_m(x,y) \quad = \quad \int_0^{\infty} \left[1 - (1 - \widehat{S}_{Y,n}(uy \mid x))^m\right] du.$$

Hence, in place of looking for the upper boundary of the support of $S_Y(y \mid x)$, as was typically the case for the full-frontier and for the efficiency score $\lambda(x,y)$, the order-*m* efficiency score can be viewed as the expectation of the maximal output efficiency score of the unit (*x; y*), when compared to *m* units randomly drawn from the population of units producing using less inputs than the level *x*. This is certainly a less extreme benchmark for the unit (*x; y*) than the "absolute" maximal achievable level of outputs: it is compared to a set of *m* peers (potential competitors) producing using less than its level *x* of inputs and we take as benchmark, the *expectation* of the maximal achievable output instead of the absolute maximal achievable output.
Then, for any y R_+^q, the expected maximum level of outputs of order-*m* is defined as $y_m^{\partial}(x) = \lambda_m(x,y)\, y$ which can be compared with the full-frontier $y^{\partial}(x) = \lambda(x,y)\, y.$.
The robust nonparametric methodology we applied in this paper adds some new advantages to the traditional nonparametric approach (DEA/FDH):

- As the robust indicators are based on estimators that do not envelop all firms, they are more robust to outliers and noise in the data which may strongly influence the nonparametric estimation of efficiency. The level of robustness can be set by means of *m* (tuning parameter). The level of

[7] Here again, we only describe the output-oriented case. For a general presentation of the probabilistic approach in efficiency analysis, see Daraio and Simar (2007).

robustness may be tuned by the percentage of points remaining above the order-*m* frontier. Clearly, when $m \to \infty$ this percentage goes to zero.

The robust nonparametric indicators avoid the *curse of dimensionality*, typically shared by non-parametric estimators, meaning that to avoid huge confidence intervals and imprecise estimation thousands of data are required. The order-*m* indicators are $\sqrt{n}$-consistent estimators whereas the DEA are only $n^{2/(p+q+1)}$-consistent estimators ($n^{1/(p+q)}$ for the FDH). This indicates for the DEA/FDH the necessity of increasing the number of observations when the dimension of the input-output space increases to achieve the same level of statistical precision;

- The order-*m* indicators allow to compare samples with different size, in an indirect way, avoiding the *sample size bias*, of which nonparametric indicators (DEA/FDH) suffer. In this case, *m* plays an important role. The benchmark, in fact, is not made against the most efficient units in the group, but against an appropriate measure drawn from a large number of random samples of size *m* within the group. In this way size-dependent effects are eliminated.
- The possibility of explaining efficiency, considering the *conditional influence* of *external factors Z* on the full frontier and on its robust counterpart (see below).

- Using this approach, it is possible to decompose the performance of a firm (*x; y*), as measured by the *Conditional Efficiency* index in three main indicators: an indicator of the *internal* or managerial efficiency, an *externality index*, and finally, an *individual* index.

- Moreover, we can evaluate the effect of external/environmental *Z* variables on the performance of firms in different economic *scenarios*, contemplating various numbers of *potential competitors*, using the parameter *m* in its dual meaning.

- Parametrization and robust elasticities are available by applying the two-steps semiparametric estimators (based on FDH and order-*m* frontiers) introduced by Florens and Simar (2005) and extended to the multioutput case by Daraio and Simar (2007).

External or environmental conditions may strongly influence the productive efficiency evaluation.
In the efficiency literature, mainly three approaches have been developed[8]: a *one-stage* approach, a *two-stage* approach and a bootstrap-based two stage approach as in Simar and Wilson (2007). Nevertheless, all of them are based on restrictive assumptions on the data generating process and/or on the role of these *external* factors on the production process.
Based on the probabilistic formulation presented above, Daraio and Simar (2005) propose a general full nonparametric approach that overcomes most drawbacks of previous approaches. The probabilistic formulation allows an easy introduction of additional information provided by external- environmental variables $Z \in R^r$. Hence, the joint distribution on (*X; Y*) conditional on *Z* = *z* defines the production process if the external factor *Z* = *z*.
The output efficiency measure under the condition *Z* = *z* can be defined as:

$$\lambda(x, y \mid z) = \sup\{\lambda \mid S_Y(\lambda y \mid x, z) > 0\},$$

where $S_Y(y \mid x, z) = \mathrm{Prob}(Y \geq y \mid X \leq x, Z = z).$ A nonparametric estimator of $\lambda(x, y \mid z)$ that requires some smoothing in *z*, is provided by the following kernel estimator of the empirical version of $S_Y(y \mid x, z)$:

[8] See Daraio and Simar (2005) and the references cited there.

$$\widehat{S}_{Y,n}(y \mid x, z) = \frac{\sum_{i=1}^{n} I\!I(X_i \leq x, Y_i \geq y) K((z - z_i)/h_n)}{\sum_{i=1}^{n} I\!I(X_i \leq x) K((z - z_i)/h_n)}.$$

where $K(.)$ is the kernel and h_n is the bandwidth of appropriate size.[9]
In a similar way, *mutatis mutandis*, Daraio and Simar (2005) introduce also the conditional order-*m* measures of efficiency with their nonparametric estimators. For the output-oriented case, the order-*m* measure of efficiency is defined as:

$$\lambda_m(x, y|z) = \int_0^{\infty} [1 - (1 - S_Y(uy \mid x, z))^m] \, du$$

and its nonparametric estimator is obtained as follows:

$$\hat{\lambda}_{m,n}(x, y|z) = \int_0^{\infty} \left[1 - (1 - \widehat{S}_{Y,n}(uy \mid x, z))^m\right] du$$

When m , we recover the full frontier conditional measures, but for finite *m*, $\hat{\lambda}_{m,n}(x, y \mid z)$ provides a more robust estimator of the frontier, robust to extremes or outliers.

The procedure for evaluating the effect of *Z* on the production process is based on the comparison of the conditional FDH measure $\hat{\lambda}_n(x, y \mid z)$ with the unconditional FDH measure $\hat{\lambda}_n(x, y)$. Accordingly, the same comparison is done for the robust order-*m* efficiency measures. In particular, the ratios $Q^z = \hat{\lambda}_n(x, y \mid z) / \hat{\lambda}_n(x, y)$ (and their robust version $Q^z_m = \hat{\lambda}_{m,n}(x, y \mid z) / \hat{\lambda}_{m,n}(x, y)$) are useful to investigate on the effects of *Z* on performance: if $Q^z = 1$, then the conditional and unconditional efficiency measures are equal: this means that *Z* does not affect the performance of the analysed firm; if Q^z is much lower than 1, this means that the firm has been *highly* influenced by *Z*.
When *Z* is univariate, the scatterplot of these ratios against *Z* and its smoothed non-parametric regression line is also very helpful in describing the effect of these external-environmental variables on the production process. In the output oriented case (used in this chapter), we have:
- An *increasing* smoothed nonparametric regression line denotes a *Z* that is *favourable* to the production process. In this framework, a favorable *Z* means that the environmental variable operates as a sort of “extra" input *freely available*: for this reason the environment is “favourable” to the production process.

- A *decreasing* smoothed nonparametric regression line indicates a *Z* that is *unfavourable* to the production process. In this case, the environmental variable works as a “compulsory” or *unavoidable* output to be produced to face the negative environmental condition. *Z* in a certain sense penalizes the production of the outputs of interest.
-A straight nonparametric regression line shows that Z does not have any influence on the production process.

Daraio and Simar (2007, section 5.4.1 and section 6.4) have shown the usefulness of robust ratios in reveiling the impact of external factors when extreme points or outliers in Z are present and mask the effect of Z applying full frontier ratios.

Here we point out the clear interpretability of these kind of scatterplots. As we have seen in the empirical results section, the analyst has an immediate view on the global effect of external factors

[9] For more information see Daraio and Simar (2006) that propose a simple data-driven method for the choice of the bandwidth useful also in the case of multivariate external factors.

on the performance: an *increasing* line indicates a positive influence of the factor, a *decreasing* line points to a negative effect and a *straight* line reveals no influence of the factor on the performance.

This nonparametric approach, with its robust counterpart, offers a rigorous methodology to identify the factors that might influence firms efficiency by measuring their *global* effect on the performance. It gives also the possibility to analyse the effect of these variables on each individual firm. The performance of a DMU (*x, y*), measured by the *Conditional Efficiency* index, $\hat{\lambda}_n(x, y \mid z)$, can be decomposed in three main indicators (see Daraio (2003) for more details):

1. An *unconditional* efficiency score $\hat{\lambda}_n(x, y)$ that represents the *internal or managerial efficiency*;

2. An *externality index* defined as the *expected value* of the ratios Q^z given the value of *z* owned by the firm.

3. An *individual index* defined as $Q^z / E(Q^z \mid Z = z)$. It compares the value of Q^z with the value we would expect for it, given its value of *Z*. It represents the firm's expected *intensity* in catching the opportunities or threats by the environment (external factor).