ELSEVIER

European Economic Review 47 (2003) 687–710

EUROPEAN ECONOMIC REVIEW

www.elsevier.com/locate/econbase

Innovation and spillovers in regions: Evidence from European patent data☆

Laura Bottazzi[a,b], Giovanni Peri[c,*]

[a] Universitá L.Bocconi, IGIER, via Salasco 5, 20136 Milan, Italy

[b] CEPR, London, UK

[c] Department of Economics, University of California Davis, One Shields Avenue, Davis, CA 95616, USA

Accepted 17 October 2002

Abstract

This paper estimates the effect of & research externalities in generating innovation. We use R&D and patent data for European Regions in the 1977–1995 period. We find that spillovers are very localized and exist only within a distance of 300 km. The estimates are robust to simultaneity, omitted variable bias, different specifications of distance functions, country and border effects. However the size of these spillovers is small. Doubling R&D spending in a region would increase the output of new ideas in other regions within 300 km only by 2–3%, while it would increase the innovation of the region itself by 80–90%.

JEL classification: O3; R11; R12

Keywords: Innovation; R&D Spillovers; Europe; Regions

1. Introduction

Innovation is fundamental to economic growth. The process through which resources spent in research and development (R&D) generate new ideas and the process of their diffusion are at the heart of the growth mechanism of modern market economies. The aim of this paper is to identify and estimate the effect of research externalities in generating innovation across space. We do so by analyzing R&D and innovation in 86 European Regions for the period 1977–1995. Sub-national regions are interesting units

☆ This paper was partly written while Giovanni Peri was in Università Bocconi.

* Corresponding author. Tel.: +1-530-752-3033; fax: +1-530-752-9382.

E-mail address: gperi@ucdavis.edu (G. Peri).

PII: S0014-2921(02)00307-0

of analysis in the increasingly integrated European market. They are more homogeneous than countries, better connected within themselves and they are becoming increasingly important as policy units for research and innovation purposes (see European Commission, 2001, on European Research Policy).

Localized R&D spillovers exist if the productivity of R&D in a region is affected by the amount of R&D resources used in other regions in spatial proximity. This definition is easily derived as the reduced form of a model (such as Romer, 1990; Jones, 1995) in which new ideas are generated using R&D resources and existing ideas as inputs. While R&D resources are a private and excludable input, existing ideas are a public input or at least a locally public input into the innovative process. The codified part of a new idea is likely to be perfectly available to anyone who can read a patent, and therefore is a fully public good. Nevertheless part of the knowledge generated with the idea is embodied, imperfectly codified, linked to the experience of the scientists or "attached" to people. This stock of knowledge increases in a region as local inventors discover new ideas. It diffuses mostly via personal contacts and face-to-face interactions. We think of it as a "local public good" as it benefits scientists within the region or its neighborhoods but it fades farther away as contacts and interactions decrease.

We estimate the reduced form of the innovation-generating equation for European Regions in the long run, using the total number of patents granted to inventors residing in a region as a measure of that region's innovative output. R&D externalities are identified as the effect of R&D intensity from region s on the innovative output of region r. We allow such an effect to depend on several characteristics of the regions r and s. Importantly, in order to identify these externalities using cross-section data, we assume that they depend on the distance between the regions. Our method, though, allows us to "let the data speak" about the intensity and the range of diffusion of these external effects. No parametric assumption on how externalities depend on distance needs to be made.

Our basic result, robust to several specifications and controls, is that small but statistically significant R&D spillovers diffuse within 300 km from the source region. Outside this distance-range no spillovers are found. More precisely, while doubling the resources devoted to R&D in one region contributes to an increase of its own patenting activity of 80% to 90%, such an increase has a small but significantly positive effect of 2–3% on patenting of regions within a 300 km range. No effect spreads further than that.

The rest of the paper is organized as follows: Section 2 summarizes the related literature, Section 3 describes the empirical model and Section 4 considers important estimation issues related to the endogeneity of R&D and presents the data. Section 5 presents the estimates of several specifications and Section 6 concludes the paper.

2. Related literature

After an early contribution by Shankerman (1979), Jaffe (1986) was the first to estimate R&D spillovers on innovation by using the effect of a "local" pool of R&D

on the patent productivity of a firm. Jaffe considered R&D done in other firms within the same technological cluster as a potential input for the innovative activity of a firm. While distance across firms was calculated in technological (rather than geographical) space, spatial correlation of R&D and innovation was used to infer the intensity of spillovers. In this respect that seminal work is closely related to our paper.

Differently from that, most of the recent literature on knowledge spillovers in geographical space has followed the approach proposed later by Jaffe et al. (1993). Exploiting the fact that each patent cites other previous patents as "sources" of relevant information, Jaffe et al. (1993) have used such a "paper trail" to track the direction and intensity of spillovers. This approach, though, as argued in Peri (2002), can only identify intensity and direction of knowledge flows and not R&D externalities. In order to capture externalities we need to measure how these flows affect R&D productivity in generating innovation. Moreover, citations do not capture non-codified knowledge flows and embodied knowledge flows, which could be important sources of localized spillovers, as Saxenian (1991) and Audretsch and Feldman (1996) argue.

While still using patents as a measure of innovative activity of one region, we do not consider the link between patents given by citations, but we instead exploit the spatial correlation of innovation and R&D activity. Our main goal is to estimate the impact of R&D resources in a representative region on innovation done in other regions at different geographical distances. Such a question needs the frame of an "innovation function" to be addressed. Aside from the use of an "innovation function" in the place of a "production function", our approach is similar to estimating externalities in production. The technique we use and the econometric problems we face are in fact close to those addressed by the literature on the spatial variation of labor productivity surveyed in Hanson (2001).

Related to our paper, although they estimate the external effect of R&D on countries' productivity (rather than on countries' innovation), are Coe and Helpman (1995), Keller (2001, 2002). Also, in that they consider the effect of human capital density at different distances, Conley et al. (1999) and Ciccone (2002) deal with the issue of inferring externalities from the spatial correlation of productivity. Finally, our paper is closely related to Ciccone and Hall (1996) in its use of historical population density as a source of exogenous variation to identify the productivity and the externalities of R&D in innovation.

3. The model

3.1. Spillovers

We consider regional units that produce profitable new ideas using two main inputs: R&D resources and existing ideas (A). R&D resources employed in region i are a private input available in region i only. Ideas A, originated in region i or in any other region, could be used as a public input of production as long as they are known to researchers operating in region i. While codified knowledge is accessible everywhere, embodied[1] knowledge is less accessible and harder to diffuse as it relies more on face to face contacts. In particular, the physical distance between a researcher and the place where the idea was discovered could make the implicit knowledge associated with it less accessible to the first. Translating this assumptions into the production function of new knowledge for a region i we propose the following expression:[2]

$$\Delta A_i = B(\mathrm{R\&D})_i^{e_R} A_i^{e_0} \prod_{j \neq i} A_j^{e(dist_{ij})}, \qquad i = 1, 2, \ldots, 86. \tag{1}$$

ΔA_i represents the change over the considered period of the stock of knowledge originated in region i. It is measured by the number of new patents granted to researchers in that region. Our proposed specification is a simple generalization of the function assumed in Romer (1990) and Jones (1995). Expression (1) says that innovation in region i depends on a Cobb–Douglas combination of R&D resources used in region i, and of ideas available to the region at the beginning of the period. The constant term B captures the effect of all common factors affecting innovation in regions. The elasticity of innovation to R&D resources is measured by e_R. Ideas generated in region i, (A_i), enter with elasticity e_0 while ideas generated in other regions (A_j) enter with an elasticity $e(dist_{ij})$ that depends on the distance between region i and region j. Such specification allows embodied knowledge contained in ideas A_j to have different impact on innovation of region i, depending on the distance between region j and region i. In particular we assume that embodied knowledge does not diffuse passed a maximum distance K and that its impact depends on the distance between regions as a step function. Hence the function $e(dist_{ij})$ is equal to e_k/n_{ik} for $dist_{ij} \in k$, with $k = \{[dist_0, dist_1), [dist_1, dist_2) \ldots, [K, \infty)\}$. The index k captures a sequence of distance intervals within which the step function is constant and n_{ik} is the total number of regions in the distance-interval k from region i. Our assumption of no diffusion beyond distance K implies $e_{[K,\infty)} = 0$.

3.2. Basic specification

Under the condition that the difference equation system (1) converges to a balanced growth path (BGP) in which the innovation rate is equal across regions, ΔA_i is proportional to A_i in the long run. Therefore, substituting for ΔA_i in BGP, solving for the vector $\ln(\underline{A})$ and log linearizing the system,[3] we obtain the following approximate relationship between innovation in a region and R&D resources:

$$\begin{aligned} \ln(\Delta A)_i = \beta + \varepsilon_0 \ln(\mathrm{R\&D})_i + \varepsilon_{[dist_0, dist_1)}[\underline{m}'_{i1} \ln\underline{(\mathrm{R\&D})}] + \cdots \\ + \varepsilon_{[dist_n, K)}[\underline{m}'_{iK} \ln\underline{(\mathrm{R\&D})}] + u_i, \qquad i = 1, 2, \ldots, 86, \end{aligned} \tag{2}$$

[1] Here and in the rest of the paper when we use the expression "embodied knowledge" we mean the non-codified knowledge attached to people, not the knowledge embodied in machines.

[2] We have omitted the time subscripts, as we consider only one period for the cross section.

[3] The procedure to derive Eq. (2) is shown in detail in Bottazzi and Peri (2002).

where $\varepsilon_0 = e_R/(1-e_0)$, and $\varepsilon_k = (e_R e_k)/(1-e_0)^2$. Therefore $\varepsilon_0, \varepsilon_{[dist_0, dist_1)}, \varepsilon_{[dist_1, dist_2)} \ldots \varepsilon_{[dist_n, K)}$ are strictly related to the elasticities of innovation to R&D at different distances and to spillovers from existing ideas.

Eq. (2) is our basic specification. We estimate it using the average yearly patent applications in region i as a measure of the flow of new ideas generated in that region in the period 1977–1995 (ΔA_i). The constant β captures any common effect determining innovation while $(\text{R\&D})_i$ is the average yearly employment (or real spending) of region i in Research and Development during the same period. Therefore ε_0 captures the elasticity of innovation to own R&D resources. Similarly, the coefficients ε_k for $k = \{[dist_0, dist_1), [dist_1, dist_2) \ldots, [K, \infty)\}$ capture the elasticity of innovation in region i to the average regional R&D done in the distance range k from that region. The regressors are constructed as follows: row vector $\underline{m}'_{ik}$ is an 86×1 row vector whose jth entry is zero if $dist_{ij}$ is not within the range k, while it is $(1/n_{ik})$ if that distance is within the range k. $\ln\underline{(\text{R\&D})}$ is a 1×86 column vector whose ith entry is $\ln(\text{R\&D})_i$. The product of these two vectors gives the average ln(R&D) for regions in the kth distance interval from region i. u_i is a random i.i.d. error capturing other unobservable determinants of innovative output. The appeal of specification (2) is that, if R&D externalities vary with distance, we can identify them at any distance, by allowing the elasticity to vary. In order to estimate Eq. (2) we choose $K = 2000$ km, so that $\varepsilon_{[2000-\infty)} = 0$.[4]

4. Empirical issues

4.1. Omitted variables and endogeneity bias

The estimation of Eq. (2) by OLS might be affected both by an omitted variable and a simultaneity problem. In general, the allocation of R&D across regions may be affected by some regional characteristics that also affect R&D productivity and are not considered in our estimates (omitted variables). In particular the stock of past accumulated innovations, A_i, generates an endogeneity problem as it attracts R&D and affects u_i in the reduced-form equation (2) if there is correlation of unobservable factors over time. This would cause inconsistent estimates of the coefficients ε_0, $\varepsilon_{[dist_0, dist_1)}$, $\varepsilon_{[dist_1, dist_2)} \ldots \varepsilon_{[dist_n, K)}$.

A good instrument to address this problem would be a variable correlated with the R&D distribution across regions but not to R&D productivity.[5] Such a variable could be found in the average value of a patent in a region. The value of a patent in a region depends on the demand that the average invention in that region receives, i.e. on the market potential for new goods in the region.[6] Unfortunately we do not

[4] We also checked for much smaller cut-off distances.

[5] In order to think carefully about endogeneity and potential instrumental variables in the working paper version of this article (Section 4 and appendix of Bottazzi and Peri, 2002) we frame this issue in a simple model of our multi-region economy.

[6] In the presence of transport costs and monopolistic competition it is easy to show, using a model of geography and trade (for instance Hanson, 1998), that regions with high GDP and/or that are located close to other regions with large GDP have larger potential demand for their goods.

directly observe market potential for an invention in a region and we proxy it, following Ciccone and Hall (1996) with 1930 distance-weighted historical population density for European Regions. Concentration of population in Europe was mostly determined by historical advantages of being on the coast, close to a harbor or on a river, which bear little implication for current productivity of R&D. On the contrary the correlation of population distribution with market potential is still large because where there are more people demand is higher.

As argued in Ciccone and Hall (1996), this instrument reduces the endogeneity bias if the original sources of agglomeration in Europe have only influence on the preferences of people about where to live and they are not correlated to modern differences in productivity of R&D not explained by our model. We believe this is a reasonable assumption in our set-up, namely it is quite reasonable to assume that agglomerating forces in Europe at the beginning of the XX century were not driven by differences in research productivity unexplained by our model. Finally let us add a qualification to our measure of externalities. As we use a measure of innovative output based on pure quantities (count of patents) and not involving relative prices (value of patents) we are capturing those externalities that directly affect physical productivity of R&D in generating patents and not those that work through local demand and affect the value of the produced innovation. This is important to ensure the exogeneity of our instruments.

4.2. The data

We use a cross section of long-run averages of variables relative to 86 regions in the period 1977–1995.[7] A detailed description of the regions and of the data can be found in Appendix A while a list of the Regions is in Bottazzi and Peri (2002) Table 2. Here we present the main features of our data-set.

We choose as regions for our analysis the territorial units[8] identified by Eurostat in each country, called Nomenclature Units Territory Statistics (NUTS). We use either "NUTS 1" or "NUTS 2" level of aggregation in order to identify, for each country, the territorial units with some degree of administrative and policy authority. These regions are rather homogeneous within themselves, often have a strong local identity, are administrative units and have some degree of policy independence. As a measure of the innovative output of a region we use the count of granted patent applications that inventors residing in each region filed with the European Patent Office between 1977 and 1995. We attribute each patent to the first inventor listed in the patent application, as it is generally done in this literature (see Jaffe et al., 1993). Patents have long been considered, not without controversy, as the best measure of output of the innovative activity. Although not all inventions are patented, the patented ones have to fulfill minimal standards of novelty, originality and potential use. Therefore patents can be considered as a good approximation to ideal data on "economically profitable ideas" which one would like to have for testing theories on innovation.

[7] Some variables are averaged over shorter intervals due to the unavailability of the whole series.

[8] See Appendix A for the exact choice of regions in each country.

The caveats of using patents as a measure of innovation have been discussed by others with care and detail (see the excellent survey in Griliches, 1990).[9] Some of those critiques, though, are less relevant for the present work. First, several studies have found that patent data are much more correlated with R&D spending as well as with other measures of innovation (such as TFP growth) in the long-run than in the short run and more in the cross-section rather than in the time-series dimension (see Griliches, 1990, pp. 1673–1674). We average a period of almost 20 years to be sure to capture the long-run relationship between R&D and patenting. Second, considering all patents as an equal contribution to new knowledge is likely to generate errors due to the variability in the importance of patents. This is more relevant when small units (such as firms) and short periods of time are considered. Our choice of relatively large regions and of almost two decades of data is likely to average out most of the problems arising from this source of variation.

Considering regions as the unit of analysis, we capture the intra-sector as well as the inter-sector spillovers across them. As we do not impose an arbitrary boundary to innovation spillovers (e.g. positive spillovers within a sector but none across them),[10] we avoid the question of which is the appropriate level of sector aggregation (four, three, two or less-digits SIC classification). This is an important issue, especially as we are analyzing externalities of ideas on other ideas which could happen across sectors. The patent-citation literature, for instance, finds that even for very broadly defined technological classes, up to 25% of patent citations happens between them.[11]

Table 1 reports the summary statistics for the data we use. As shown in the first row, regions significantly differ in their size. Hence in the empirical analysis of the next section we standardize variables by regional area. This standardization eliminates the issue of regional size (in terms of area) as a possible cause of distortion. Of course regions are still different in their "economic" size (population per square kilometer). We analyze and exploit that variation in our analysis.

The heterogeneity in R&D intensity and in patenting across regions is very high. The most active regions spend around 3% of their regional resources in R&D, generating thousands of successful patent applications per year. The least active employ around 0.1% of total resources in R&D and barely fill one successful patent application per year. Certainly European regions are also rather different in their economic development (GDP per capita, levels of schooling and sector specialization) ranging from low skills and low income regions in the Mediterranean belt to high tech, human capital and service-intensive regions in the Franco-German-Benelux core.

Distance between two regions is measured as the shortest air distance between their boundaries. Two regions which share a border are attributed zero distance between each other, while regions which do not share a border are given a distance measured as kilometers between their closest borders.

[9] Due to a lag in generating the innovation from R&D activity, the relation between these variables is strong only in the long-run.

[10] As done, for instance, in Bernstein and Nadiri (1989).

[11] See Jaffe et al. (1993, p. 597).

Table 1
Summary statistics (Averages 1977–1995)

Variables:	Average	Std.Dev.	Min.	Max.
Area in km^2	26132	28937	161	215025
Total employment (thousands)	1522	1471	58	8593
Total population (thousands)	3760	3266	114	17339
% of GDP in R&D[a]	1.31	0.88	0.10	3.8
% of Private R&D in Total[a]	56	24	2	100
% of Employment in R&D[a]	1.02	0.67	0.10	3.0
% of Employment in agriculture	9	8	0	44
% of Employment in industry	29	6	13	45
% of Employment in services	62	8	35	86
Patent applications per year	178	356	1	1838
Ln (per capita real GDP)[b]	2.35	0.27	1.57	3.06
% of College educated[c]	10	4	3.8	25

84 European regions
[a]Relative to 1984–1995 only.
[b]Millions of 1985 ECU.
[c]Relative to year 1986.

5. Estimation results

5.1. Basic specification

The basic specification of Eq. (2) that we estimate is the following:

$$\begin{aligned}\ln(Patent)_i = \beta &+ \varepsilon_0 \ln(R\&D)_i + \varepsilon_1[\underline{m}'_{i[0-300]} \ln\underline{(R\&D)}]\\ &+ \varepsilon_2[\underline{m}'_{i[300-600]} \ln\underline{(R\&D)}]\\ &+ \varepsilon_3[\underline{m}'_{i[600-900]} \ln\underline{(R\&D)}] + \varepsilon_4[\underline{m}'_{i[900-1300]} \ln\underline{(R\&D)}]\\ &+ \varepsilon_5[\underline{m}'_{i[1300-2000]} \ln\underline{(R\&D)}]\\ &+ D_i * (Country)_i + u_i. \end{aligned} \tag{3}$$

$(R\&D)_i$ measures employment (or real spending) in R&D per square kilometer in region i and $(Patent)_i$ measures number of patents per square kilometer in region i. Both are averages for the 1977–1995 period. The distance classes are the following five: 0–300, 300–600, 600–900, 900–1300 and 1300–2000. Such classes have been chosen so that each region has at least another European region within each range and so that the average number of regions in each interval is roughly equal to 16 (see Appendix A for details). Eleven country-dummies $(Country)_i$, capture unobserved factors such as legal protection, efficiency of the judicial system, quality of institutions common to regions of the same country. We also assume that absorption of technology from the rest of the world, mainly US or Japan, may differ across countries but is constant within each of them so that such effect could be included in the country effect. The parameters of our regression are identified only using the within-country cross-regional variation of patenting and R&D expenditure.

Table 2
Basic specifications using R&D employment

Variables	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.96** (0.06)	0.85** (0.06)	0.84** (0.07)	0.84** (0.07)	0.84** (0.07)	0.83** (0.07)
$\underline{m}'_{[-300]} \ln(\underline{R\&D})$		0.027** (0.011)	0.025** (0.011)	0.027** (0.011)	0.027** (0.011)	0.025** (0.012)
$\underline{m}'_{[30-600]} \ln(\underline{R\&D})$			0.01 (0.017)	0.004 (0.017)	0.003 (0.018)	−0.008 (0.019)
$\underline{m}'_{[60-900]} \ln(\underline{R\&D})$				0.015 (0.013)	0.015 (0.013)	0.010 (0.013)
$\underline{m}'_{[90-1300]} \ln(\underline{R\&D})$					−0.003 (0.012)	−0.005 (0.12)
$\underline{m}'_{[130-2000]} \ln(\underline{R\&D})$						−0.02 −0.017
12 Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.70	0.86	0.86	0.87	0.87	0.87
Observations	86	86	86	86	86	86

R&D = Regional employment in research and development (private and public).
Dependent variable: ln(yearly patent applications).
Cross-section using 1977–1995 averages.
Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.

Tables 2 and 3 report OLS estimates for the basic specification with heteroskedasticity robust standard errors. The two tables differ only in the measure of R&D activity used: regional employment in public and private R&D per square km of area in the first case (Table 2), regional (private and public) R&D spending in real (1985) Ecus per square km of area, in the second case (Table 3). The results are very similar.

In both tables the first column includes only R&D done in the region itself as input of the innovation function. In the following columns we include as regressors, one at a time, all the distance-groups of average regional R&D. The coefficients are standardized by the average number of regions within each distance class, so that their interpretation is straightforward: each coefficient measures the elasticity of regional innovation to R&D resources used at the considered distance. Let us consider, for instance, specification VI in Table 3, which contains all the distance classes between 0 and 2000 km. An increase of 1% in R&D spending of the average region increases by 0.80% its own innovative activity and by 0.025% the innovative activity of regions within a range of 300 km. The effects on all other regions are not statistically different from 0.

Table 3
Basic specifications using R&D spending

Variables	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.95**	0.83**	0.82**	0.82**	0.82**	0.80**
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
$\underline{m}'_{[-300]}\ln(\underline{R\&D})$		0.030**	0.028**	0.029**	0.026**	0.025**
		(0.010)	(0.001)	(0.011)	(0.011)	(0.011)
$\underline{m}'_{[30-600]}\ln(\underline{R\&D})$			0.004	0.003	0.002	−0.007
			(0.01)	(0.011)	(0.012)	(0.013)
$\underline{m}'_{[60-900]}\ln(\underline{R\&D})$				(0.004)	0.005	−0.004
				(0.012)	(0.013)	(0.012)
$\underline{m}'_{[90-1300]}\ln(\underline{R\&D})$					−0.010	−0.007
					(0.010)	(0.012)
$\underline{m}'_{[130-2000]}\ln(\underline{R\&D})$						−0.018
						(0.012)
12 Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.81	0.88	0.88	0.88	0.89	0.89
Observations	86	86	86	86	86	86

R&D = Regional real spending (1985 ECU) in research and development (private and public).
Dependent variable: ln(yearly patent applications).
Cross-section using 1977–1995 averages. Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.

Our initial assumption that local spillovers are not significant beyond the distance of 2000 km seems a safe one in the light of the estimates we obtain. In fact we find no sign of local spillovers past the distance of 300 km. While statistically significant, the external effect of regional R&D on innovation is only a small fraction (about 3%) of the internal one. Nevertheless, as the average European region has about 15 other regions within a 300 km distance from itself, the overall effect if each of them were to increase their R&D intensity by 1%, would be a sizeable increase of 0.39% in the region's innovative output. Therefore, our results show that if each region were to double its R&D spending this would result in 80–90% higher patenting in each region due to the internal contribution of R&D, plus an extra 40% more patenting due to the cumulated external effects from R&D in regions within a 300 km distance.

5.2. Sector composition and human capital

Part of the correlation between other regions' R&D and Patenting could be spurious and due to omission of relevant variables. In particular, as we consider aggregate patenting, we should control for the fact that different sectors have different propensities to patent: the manufacturing sector patents a much larger fraction of its innovations than services and, within manufacturing, chemicals and machinery are by far the most active sub-sectors.

Table 4, columns I and IV, show the regressions of innovative output on own and other regions' R&D, controlling for the natural log of the share in total value added produced by the region's manufacturing sector, $\ln(Manufacturing)_i$, and within it, by the Chemical and Machinery sector ($\ln(Machinery)_i$ and $\ln(Chemicals)_i$). Their effect on innovation is small, positive and only borderline significant for the Machinery sector in one specification. Notice, though, that if we run the same regression without cross-regional spillovers and country-dummies (not reported), all sector variables have very significant and positive coefficients.[12] The inclusion of these controls does not change the estimates of the elasticity of innovation to own R&D and to R&D in the 0–300 km range. Externalities from close regions remain significant and very close in magnitude to their previous estimates. Externalities from regions which are further away are still not significantly different from 0.

Another potentially important determinant of innovation in a region is its endowment of human capital. Workers with higher schooling levels should be more effective in generating innovation, for a given amount spent in R&D (or for a given number of employees in R&D). For this reason in columns II and V of Table 4 we include, as explanatory variable, the log of the share of college graduates in the population of a region, $\ln(College)_i$. Specification II uses R&D Employment while specification V uses R&D spending as measure of resources employed in R&D. The effect of human capital is large and significant on innovative output: an increase of 1% in the share of college graduates in a region increases its patenting by 0.9%. However, the coefficients on the elasticity to own R&D and R&D in the 0–300 km range are basically unchanged.

Finally, we consider the possibility that human capital might have an external effect on innovation. In particular, the estimated external effect of R&D could be a spurious result of the correlation of R&D in the 0–300 km range with human capital in the 0–300 km range. If human capital intensity, rather than R&D, generates the externalities we would capture such an effect by including the intensity of human capital in the 0–300 km range as an explanatory variable. This is what we do in columns III and VI but we do not find any evidence of such human capital-externalities, while R&D externalities remain basically unchanged.

5.3. Distance intervals

In the previous two section we chose the distance classes for regions as a reasonable trade-off between a fine division of space and the accuracy of the estimates.

[12] We also run regressions controlling for the shares of the other 2-digit manufacturing sectors but their coefficients were never significant.

Table 4
Regional controls

Used as R&D measure:	R&D Employment			R&D Real spending		
Specification:	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.82** (0.07)	0.80** (0.07)	0.81** (0.07)	0.81** (0.06)	0.79** (0.06)	0.79** (0.06)
$\underline{m}'_{[-300]} \ln\underline{(R\&D)}$	0.026** (0.012)	0.029** (0.012)	0.028** (0.011)	0.025** (0.011)	0.026** (0.010)	0.029** (0.011)
$\underline{m}'_{[30-600]} \ln\underline{(R\&D)}$	−0.010 (0.019)	−0.010 (0.018)	−0.003 (0.02)	−0.010 (0.014)	−0.010 (0.014)	−0.010 (0.015)
$\underline{m}'_{[60-900]} \ln\underline{(R\&D)}$	0.006 (0.013)	0.007 (0.013)	0.004 (0.014)	−0.007 (0.014)	−0.006 (0.012)	−0.006 (0.013)
$\underline{m}'_{[90-1300]} \ln\underline{(R\&D)}$	0.001 (0.12)	−0.004 (0.012)	−0.006 (0.012)	−0.006 (0.012)	−0.007 (0.11)	−0.007 (0.011)
$\underline{m}'_{[130-2000]} \ln\underline{(R\&D)}$	−0.010 (0.015)	−0.010 (0.015)	−0.010 (0.016)	−0.015 (0.012)	−0.014 (0.012)	−0.015 (0.013)
$\ln(Manufacturing)_i$	0.32 (0.42)	0.62 (0.42)	0.61 (0.43)	0.16 (0.41)	0.46 (0.42)	0.46 (0.42)
$\ln(Machinery)_i$	0.62* (0.33)	0.41 (0.33)	0.38 (0.34)	0.25 (0.32)	0.09 (0.32)	0.10 (0.32)
$\ln(Chemicals)_i$	0.28 (0.22)	0.24 (0.22)	0.26 (0.22)	0.23 (0.21)	0.29 (0.20)	0.28 (0.21)
$\ln(College)$		0.92** (0.35)	0.95** (0.36)		0.77** (0.33)	0.76** (0.34)
$\underline{m}'_{[-300]} \ln\underline{(College)}$			0.0216 (0.044)			0.005 (0.040)
12 Country dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.88	0.89	0.89	0.89	0.90	0.90
Observations	86	86	86	86	86	86

Dependent variable: ln(yearly patent applications).
Cross-section using 1977–1995 averages. Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.

Increasing the "finesse" of the distance-grid would decrease the number of regions in each interval and increase the number of regressors, lowering the precision of the estimates. Nevertheless, to be sure of the robustness of our results we define distance brackets of 100 km within the closest 1000 km of distance.

Table 5 shows the results of the analysis with 100 km intervals considering, as usual, employment (in specification I–II) or spending (in specification III–IV) as measures of R&D intensity. In general, the only spillovers' coefficients that are always positive and statistically significant are those on R&D in the 0–100 and in the 100–200 km range. The coefficient on R&D in the 200–300 km range is still positive but not statistically different from 0. The standard errors of the estimates are almost twice the size of those in Table 5. The effect of R&D, beyond 300 km of distance, is never statistically significant, and it is often negative.

Columns I and III report the results of the basic specification when we control for country dummies besides the R&D variables. Somewhat strangely, the coefficient on the R&D at 100–200 km is slightly larger than the one on R&D at 0–100, although the difference is not statistically significant. [13] However, if we run the regression including only the first three distance classes, from 0 to 300 km (not reported), the size of the coefficients is decreasing as distance increases. For instance in the basic specification with only three distance classes (0–100, 100–200, 200–300 km), the coefficients are equal to 0.053, 0.041 and 0.017 respectively with std. errors of 0.021, 0.022 and 0.021. The inclusion of several other intervals probably reduces the precision of the estimates.

Consistently with the previous estimates, the average of the 0–100, 100–200 and 200–300 R&D coefficients in specification I, Table 5, is 0.029, which is basically equal to the estimate of the average effect of R&D in the 0–300 km range from specification II in Table 2. Columns II and IV simply include all the regional controls as in columns II and V of Table 4 and confirm the same findings as the basic specification.

In Bottazzi and Peri (2002) we also show that the positive effect from close regions' R&D not only benefits the "central" regions in Europe but also the peripheries and semi-peripheries. Controlling for a measure of centrality within Europe that puts the center roughly in the Belgium–Luxembourg area, the effect of R&D within 300 km is still significant. This means that regions in the periphery with R&D intensive neighbors benefit from it just as central regions do.

5.4. Parametric decay with distance

Part of the appeal of our approach is that we have been able to estimate the elasticity of innovation to R&D at different distances without having to assume a particular functional form of diffusion of the spillovers. The data themselves suggest a rapid decrease of spillovers with distance, with positive effect on productivity only within 200 km (possibly 300). We can approximate this behavior using a parametric function to weight R&D at different distances from the region. If we approximate the decay-function of spillovers correctly, we can use all the data to estimate one parameter, and have a potentially accurate evaluation of the function at 0 distance from the region. Other authors (for instance Keller, 2001, 2002) have estimated parametric functions of decay for R&D spillovers over space.

[13] Interestingly, this result was also found in our previous analysis, Bottazzi and Peri (2000) and by Conley et al. (1999).

Table 5
Finer distance intervals

Used as R&D measure:	R&D Employment		R&D Real spending	
Variables	I	II	III	IV
$\ln(R\&D)_i$	0.80**	0.77**	0.80**	0.76**
	(0.07)	(0.07)	(0.07)	(0.07)
$\underline{m}'_{[-100]} \ln(\underline{R\&D})$	0.039*	0.027	0.04**	0.038*
	(0.022)	(0.20)	(0.02)	(0.021)
$\underline{m}'_{[10-200]} \ln(\underline{R\&D})$	0.042**	0.036**	0.07**	0.061**
	(0.018)	(0.16)	(0.025)	(0.029)
$\underline{m}'_{[20-300]} \ln(\underline{R\&D})$	0.006	0.025	0.02	0.032
	(0.02)	(0.20)	(0.02)	(0.021)
$\underline{m}'_{[30-400]} \ln(\underline{R\&D})$	−0.020	−0.03	−0.01	−0.01
	(0.025)	(0.024)	(0.023)	(0.021)
$\underline{m}'_{[40-500]} \ln(\underline{R\&D})$	0.002	0.008	−0.002	−0.004
	(0.015)	(0.014)	(0.012)	(0.012)
$\underline{m}'_{[50-600]} \ln(\underline{R\&D})$	−0.025	−0.030	−0.03	−0.03
	(0.030)	(0.025)	(0.024)	(0.024)
$\underline{m}'_{[60-700]} \ln(\underline{R\&D})$	−0.03	−0.03	−0.04	−0.04
	(0.024)	(0.023)	(0.026)	(0.026)
$\underline{m}'_{[70-800]} \ln(\underline{R\&D})$	0.02	0.004	0.02	0.006
	(0.028)	(0.028)	(0.025)	(0.024)
$\underline{m}'_{[80-900]} \ln(\underline{R\&D})$	0.006	−0.003	−0.002	−0.003
	(0.026)	(0.024)	(0.024)	(0.023)
$\underline{m}'_{[90-1000]} \ln(\underline{R\&D})$	0.02	0.018	0.02	0.019
	(0.02)	(0.02)	(0.021)	(0.020)
Other controls	No	Yes	No	Yes
12 Country dummies	Yes	Yes	Yes	Yes
R^2	0.83	0.86	0.85	0.87
Observations	86	86	86	86

Other controls are: ln(*Manufacturing*), ln(*Machinery*), ln(*Chemicals*) and ln(*College*).
Dependent variable: ln(yearly patent applications).
Cross-section using 1977–1995 averages. Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.

Table 6
Parametric specification using R&D spending

Decay of weighting Specification	Linear I	Inverse II	Exponential III
$\ln(R\&D)_i$	0.76**	0.77**	0.78**
	(0.06)	(0.06)	(0.06)
$Pool(R\&D)_i$	0.012*	0.034**	0.072**
	(0.007)	(0.013)	(0.024)
Other controls	Yes	Yes	Yes
12 Country dummies	Yes	Yes	Yes
R^2	0.88	0.89	0.89
Observations	86	86	86

R&D = Regional real spending (1985 ECU) in research and development (private and public).
Dependent variable: ln(yearly patent applications).
Other controls are: ln(*Manufacturing*), ln(*Machinery*), ln(*Chemicals*) and ln(*College*).
Cross-section using 1977–1995 averages. Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.

We construct a "pool" of R&D resources, $Pool(R\&D)$, potentially available to region i assuming that the "effective" R&D contribution by each of the other regions depends parametrically on the distance between that region and region i. Call $F(d)$ a "decay" function of distance d such that $1 \geqslant F(d) \geqslant 0$ and $F(0)=1$. We still divide the space around each region in 100 km intervals, we consider the average R&D done in each interval and we use 100 km as the unit for distance. We weight average $\ln(R\&D)$ at distance d by the function $F(d)$. In particular we choose a linear, an inverse and an exponential specification. The functions used are: $F(d)=1-(d/20)$ in the linear case, $F(d)=(1+d)^{-1}$ for the inverse and $F(d)=\mathrm{e}^{-d}$ for the exponential. When $d=0$ we are weighting $\ln(R\&D)$ with a coefficient of one, which is the maximum, while as d increases the weight given to the research is smaller and smaller. We assign the distance of the mid-interval to each range: the 0–100 gets $d=0.5$ (i.e. 50 km), the interval 100–200 gets $d=1.5$ and so on.

The coefficient of the regression of patenting on the $Pool(R\&D)$ variable has an easy interpretation. It captures the elasticity of innovation to R&D done at 0 distance from the region (i.e. where $F(d)=1$) and therefore it could be considered as a measure of the intensity of the spillovers on region i from the regions sharing a border with it. Table 6 reports the results using $Pool(R\&D)_i$, calculated using the different functions. Column I uses the linear specification, column II uses the inverse and column III uses the exponential specification. All specifications include own R&D, country dummies and regional controls. We only report the results using R&D spending as measure of R&D. Those obtained using R&D employment are similar.

The coefficients for $Pool(R\&D)_i$ in columns I, II and III show that the significance and the size of the coefficient increase going from the linear to the inverse

Table 7
IV Estimation (historical population)

Variables	I	II	III	IV
$\ln(R\&D)_i$	0.91**	0.89**	0.89**	0.89**
	(0.071)	(0.071)	(0.07)	(0.07)
$\underline{m}'_{[-300]}\ln(\underline{R\&D})$	0.021**	0.0181*		
	(0.011)	(0.011)		
$Pool(R\&D)_i$			0.07**	0.066**
			(0.024)	(0.024)
Other controls	No	Yes	No	Yes
12 Country dummies	Yes	Yes	Yes	Yes
R^2	0.87	0.89	0.87	0.89
Observations	86	86	86	86

R&D = Regional real spending (1985 ECU) in research and development (private and public).
Dependent variable: ln(yearly patent applications).
Other controls are: ln(*Manufacturing*), ln(*Machinery*), ln(*Chemicals*) and ln(*College*).
Instruments: Historical regional population density per Sq. km (circa 1930).
Cross-section using 1977–1995. Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.

to the exponential specification. This is consistent with the results obtained in our non-parametric specification: the first specification gives too much weight to R&D done far from the region and, trying to fit a linear function, it estimates a very low coefficient for the zero-distance spillovers. The inverse function does better as the weights decrease more rapidly, but certainly the exponential function performs best as it attributes non-negligible weight only to the first couple of intervals. Using this specification, an increase of 1% in R&D generates an external effect of 0.07% on regional innovation at zero distance. These spillovers decrease very quickly with distance so that at 200 km the effect is of 0.008% and at 500 km it is null.

5.5. Instrumental variables estimation

For the reasons discussed in Section 4.1 we instrument R&D intensity using proxies of historical market potential for regions: the density of population in 1930, $\ln(Pop_{30})_i$ and the "historical market potential", $\ln(MkPot_{30})_i$ calculated as the sum of the region's population density and the other regions' density (in 1930) weighted with an exponentially decreasing function. The rate of decay for this function is set at −0.03 and is taken from the estimates of market potential in Hanson (1998).

Table 7 shows these estimates: columns I and II include own R&D and R&D in the 0–300 km range, while columns III and IV estimate the specification with own R&D and the exponentially weighted $Pool(R\&D)$. In the first case we use, as instruments

Table 8
IV Estimation (historical MKT potential)

Variables	I	II	III	IV
$\ln(R\&D)_i$	0.89**	0.87**	0.85**	0.081**
	(0.071)	(0.071)	(0.071)	(0.07)
$\underline{m}'_{[-300]} \ln(\underline{R\&D})$	0.20*	0.19*		
	(0.010)	(0.010)		
$Pool(R\&D)_i$			0.07**	0.062**
			(0.026)	(0.025)
Other controls	No	Yes	No	Yes
12 Country dummies	Yes	Yes	Yes	Yes
R^2	0.87	0.88	0.86	0.88
Observations	86	86	86	86

R&D = Regional real spending (1985 ECU) in research and development (private and public).
Dependent variable: ln(yearly patent applications).
Other controls are: ln(*Manufacturing*), ln(*Machinery*), ln(*Chemicals*) and ln(*College*).
Instruments: Historical MKT potential: population weighted exponentially (circa 1930).
Cross-section using 1977–1995. Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.

for $\ln(R\&D)_i$ and $\underline{m}'_{[0-300]} \ln(\underline{R\&D})$, the variables $\ln(Pop_{30})_i$ and $\underline{m}'_{[0-300]} \ln(\underline{Pop_{30}})$. In the second case $\ln(R\&D)_i$ and $Pool(R\&D)$ are instrumented with $\ln(Pop_{30})_i$ and $Pool(Pop_{30})_i$. The variable $Pool(Pop_{30})_i$ is constructed in the same way as for the R&D pool, just using historical population rather than R&D. We choose to include only R&D in the 0–300 km range as all the non-parametric estimates indicate this as the only significant variable. We choose to include the exponential "Pool" specification as it provides the best fit and the highest estimate of externality in Table 6. Table 8 is identical to Table 7, just using the market potential $MkPot_{30}$ rather than Pop_{30} as instrument.

The instruments are excellent as they explain between 72% and 78% of the independent variables' variation. As we have two endogenous variables and two instruments the system is just identified. While the estimates of the effect of own R&D in specification I is slightly higher than the OLS estimate (Table 3 column II), the estimate of the coefficient on $\underline{m}'_{[0-300]} \ln(\underline{R\&D})$ is slightly smaller than its OLS estimate (0.02 rather than 0.03). Nevertheless in all specifications the external effect of R&D (within 0–300 km or measured as $Pool(R\&D)$) is still significant and very close to its OLS estimate. While the instrumental variables estimates seem to correct for a small upward bias of the coefficient on the externality, the evidence suggests that the externality, while not large, is still there. Innovation has an elasticity of 2% to R&D in the 0–300 km range, and such elasticity could be as large as 4–6% for R&D done at 0 distance.

Table 9
Technological distance

Variables	I[a]	II	III	IV
$\ln(R\&D)_i$	0.69** (0.07)	0.67* (0.06)	0.82* (0.08)	0.86* (0.08)
$\underline{m}'_{[-300]} \ln(\underline{R\&D})$	0.025* (0.009)		0.019* (0.01)	
$Pool(R\&D)_i$		0.058* (0.02)		0.043* (0.021)
$\underline{m}'_{TECH.DIS} \ln(\underline{R\&D})$	0.03* (0.01)	0.03* (0.01)	0.021* (0.01)	0.015 (0.011)
Other controls	Yes	Yes	Yes	Yes
12 Country dummies	Yes	Yes	Yes	Yes
R^2	0.91	0.91	0.90	0.90
Observations	86	86	86	86

R&D = Regional real spending (1985 ECU) in research and development (private and public).
Dependent variable: ln (yearly patent applications).
Other controls and Instruments as usual. Method of estimation: Col. I,II OLS; Col III, IV 2SLS.
Cross-section using 1977–1995. Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.
[a]In specification I, we also include R&D in the other distance-intervals up to 2000 km.

5.6. Technological distance

As we consider the aggregate effect of R&D spillovers on innovation it is hard to know how much is due to inter-sector and to intra-sector spillovers. It is important though to control for technological proximity between regions as geographical distance could be just proxying for technological proximity. To do this we use the Technological Proximity Index (*TPI*) developed by Jaffe (1986). We divide patents in thirty technological classes, following the International Patent Code (IPC) classification. Each region is then assigned a 30×1 "technology vector" which measures the share of patenting in each of the technological classes for the region. The technological proximity between two regions is given by the uncentered correlation of their technological vectors. Two regions that patent exactly in the same proportion in each class have an index equal to one. Two regions that only patent in different sectors have an index equal to 0. This index is appealing as it allows for a continuous measure of technological space and avoids the problem of defining technological distance between sectors.

We use this index to weight R&D done in other regions and we explore in Table 9 the effect of controlling for technological distance. The variable $[\underline{m}'_{TECH.DIS} \ln(\underline{R\&D})]$ in Table 9 captures, for each region, the sum of average $\ln(\underline{R\&D})$ in the other regions, each weighted by the *TPI* between the two regions. This variable controls for spillovers which are stronger between technologically similar regions. Table 9 reports the estimates for the specifications including controls and country dummies using OLS (column I and II) or instrumental variables (column III and IV) method. We include the control $\underline{m}'_{TECH.DIS}\ln(\underline{R\&D})$ with the non-parametric (I and III) or with the parametric specification (II and IV) of distance spillovers.

The estimates show that technological proximity matters. In particular increasing R&D in a technologically identical region ($TPI = 1$) by 1% has an impact between 0.01% and 0.03% on the region. The effect of spillovers from close regions (0–300 km) is slightly reduced, confirming that part of those spillovers are intra-industry. Nevertheless that effect still exists and is significant. Similarly when we use exponentially weighted $Pool(R\&D)$, we get a significant coefficient between 0.04 and 0.06. We leave to further research a more careful analysis of the interactions between these two dimensions of knowledge diffusion.

5.7. Border effect on spillovers

It is a common finding in the literature that national borders constitute a hurdle to exchange. As shown originally by McCallum (1995) and confirmed in several works by Helliwell (notably in Helliwell, 1998) migration and trade flows are much more intense between regions of the same country than of different ones. As we are considering flows of ideas across regions the same border effect could exist.

To assess the size of the border effect, we consider a grouping which divides regions into 300 km intervals (as above) but now for each interval we group separately regions within the same country and regions in different countries, and then we calculate the average R&D intensity in each group. Above 600 km of distance there are extremely few regions in the same country, therefore we only need four groups of regional R&D intensity. These groups reported in rows 2–5 of Table 10 capture, respectively, average R&D intensity in regions within 300 km in the same country, within 300 km but in different countries, within 300–600 km in the same country and within 300–600 km in different countries. Including these regressors in the basic equation and considering the estimates of their coefficients allows us to estimate indirectly a “border effect” on externalities. The first two columns in Table 10 report the result of the basic equation using the four groups, with and without country dummies. As the country dummies absorb part of the effect of within-country spillovers we estimate also one specification without them.

The point estimates of the coefficients is larger for the within-country spillovers in both specifications. Nevertheless the estimates are not very precise and the effect of foreign regions in the 0–300 km range is still significant. At 300–600 km distance the effect of foreign R&D becomes slightly negative, while the regions of the same country still have a positive, but not significant effect. In order to increase the precision of the estimates in columns III and IV we merge R&D for regions in the 0–600 km range, while still keeping national and foreign regions separate. Now the point estimate of spillovers from national regions is twice to three times as large as the spillovers from foreign regions in the same range, but standard errors are large. Only in the

Table 10
Border effect

Variables	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.79** (0.05)	0.77** (0.07)	0.79** (0.05)	0.78** (0.07)	0.91** (0.06)	0.90** (0.08)
$\underline{m}'_{[-300]within}\ln\underline{(R\&D)}$	0.028 (0.018)	0.042 (0.025)				
$\underline{m}'_{[-300]across}\ln\underline{(R\&D)}$	0.023* (0.008)	0.030** (0.011)				
$\underline{m}'_{[30-600]within}\ln\underline{(R\&D)}$	0.010 (0.013)	0.010 (0.02)				
$\underline{m}'_{[30-600]across}\ln\underline{(R\&D)}$	−0.003 (0.007)	−0.002 (0.008)				
$\underline{m}'_{[-600]within}\ln\underline{(R\&D)}$			0.032** (0.015)	0.057 (0.035)	0.030 (0.020)	0.04 (0.037)
$\underline{m}'_{[-600]across}\ln\underline{(R\&D)}$			0.014 (0.010)	0.018 (0.013)	0.005 (0.008)	−0.01 (0.009)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
12 Country dummies	No	Yes	No	Yes	No	Yes
R^2	0.87	0.90	0.86	0.89	0.85	0.87
Observations	86	86	86	86	86	86

R&D = Regional real spending (1985 ECU) in research and development (private and public).
Dependent variable: ln(yearly patent applications).
Other controls are: ln(*Manufacturing*), ln(*Machinery*), ln(*Chemicals*) and ln(*College*).
Method of estimation: Col. I-IV OLS, Col. V-VI 2SLS.
Cross-section using 1977–1995. Heteroskedasticity robust std. errors in parentheses.
*Significant at 10% level.
**Significant at 5% level.

specification without Country dummies is the estimate of national regions' spillovers significantly positive. Columns V and VI use 2SLS estimation, using the variables $\ln(Pop_{30})$, $\underline{m}'_{[0-600]within}\ln(Pop_{30})$ and $\underline{m}'_{[0-600]across}\ln(Pop_{30})$ as instruments. In spite of some decrease in statistical significance the point estimate for national regions' R&D is still quite larger than that for foreign regions' R&D. All in all, a border effect is present but not precisely estimated. We should point out that the relatively small variation of R&D intensity within countries versus the larger variation of R&D intensity across countries may result in less precise estimates of within country spillovers compared to cross country spillovers. This affects the precision of a test of the "border effect" and may be observed in the estimates of columns III–VI in which cross country spillovers are estimated with smaller standard errors.

6. Conclusion

Knowledge spillovers are unobservable. While some growth theorists have evoked their pervasive presence to justify sustained growth (notably Lucas, 1988; Romer, 1990) it is the empirical analysis that will ultimately provide us with an answer regarding their existence and size. Spillovers are due to the partially public good nature of knowledge, as new ideas increase productivity of researchers in their quest for profitable innovations. If at least part of the new knowledge (the "embodied part") is sensitive to distance in its diffusion, externalities are spatially localized. This could be one of the reasons why innovative activity tends to cluster (as argued in Audretsch and Feldman, 1996), and why innovators in an area tend to cite ideas from their neighbors more frequently (Jaffe et al., 1993).

Our work tackles exactly this issue by estimating an "innovation generating" function at the regional level for Europe. In particular, we estimate the elasticity of innovation to R&D done in other regions at various distances. We find that the effects of R&D in generating innovation are quite localized: most of the benefits accrue to the region that employs the R&D resources and small positive externalities accrue to regions within 300 km from it. Such findings of strongly localized effects of R&D confirm, at the regional level, the cross-country findings of Keller (2002). It is hard to compare our results to the estimates of knowledge spillovers from patent citations that are found (see Jaffe et al., 1993) to be much less localized than the effects we find here. Citations, though, are by no means a measure of knowledge externalities but only of knowledge flows. It is possible that codified knowledge, captured by citations, diffuses widely but does not affect productivity much. Vice-versa, the part of knowledge that really helps productivity and generates externalities relies on informal and close contacts and remains more localized. The externalities we estimate could be the result of diffusion of non-codified knowledge between people who have frequent (weekly or monthly) interactions.

Acknowledgements

We thank Antonio Ciccone for very useful comments. Erick Brackonier, Rudi Dornbusch, Jordi Gali, Giorgia Giovannetti, John Helliwell, Adam Jaffe, Jaume Ventura and two anonymous referees provided useful suggestions. We thank Davide di Laurea, Elisa Faraglia and Andrea Gentilini for excellent Research Assistance. We acknowledge Università Bocconi and the Italian Ministry of Research (MURST) for funding relevant parts of this project. Errors are ours.

Appendix A. Data appendix

A.1. Regions and distances

We use 86 regions in the analysis covering all EU15 countries except for Austria, Finland and Sweden. The distribution of regions by countries is as follows:

Belgium 3, Luxemburg 1, Germany (West) 11, Denmark 1, Spain 7, France 22, Greece 4, Ireland 1, Italy 20, Holland 4, Portugal 1, UK 11

We have chosen the level of aggregation NUTS 1 (in Belgium, Germany, Spain, Greece, Hooland and UK) or NUTS 2 (Italy and France) depending on the availability of the data on R&D, only for four small countries (Denmark, Ireland, Luxemburg and Portugal) the whole country represents only one region (NUTS 0). Regions vary in size, but most of them (63) have a radius (once their area is made circular) between 0 and 100 km, while only 5 have a radius larger than 150 km. The average radius is 85 km.

Considering the border-to-border distance across regions described in the text, and grouping them in classes of distances, we choose five intervals so that we can inquire about the effect of R&D at different distances. The following are the intervals in Km and the average number of regions that each region has within a circle of that distance:

[0–300) average number of regions: 15.7
[300–600) average number of regions: 18.3
[600–900) average number of regions: 17.8
[900–1300) average number of regions: 16.7
[1300–2000) average number of regions: 13.3
more than 2000, average number of regions: 3.3

A.2. Patents

The Data on Patents are a random 1/100 extraction from the European Patent Office (EPO) Data relative to granted patents whose application date is between 1977 and 1995. The initial year, 1977, has been chosen as it is the year in which the EPO was established. The final year was chosen in order to have a large percentage of patents granted even for the last year. The total number of patents, used to construct the regional variable $Patent_i$ was 3010 out of a total of 6010 patents extracted. The other 3000 patents had applicants residing out of the European Countries (mainly US and Japan). We attributed each patent to the region of its first inventor and we simply added the number of patents in each region and divided it by 18 (the number of years considered) to have the average yearly patenting.

A.3. R&D and other regional data

The data on R&D are the total employment (in thousands) in R&D obtained by adding private R&D, government R&D and high education R&D within the region. The period covered by the regional series in the dataset REGIO, from EUROSTAT, is in general 1984–1995. We use the average employment for the region as the balanced growth path level of regional resources used in R&D. Population and GDP (measured in constant 1985 ECU's) are also taken from the REGIO dataset, as well as the share of employment in Industry and in Services and the total employment in each region. These series often cover the whole 1977–1995 period. In any case we consider, for each variable, the average over the longest series available within that period. The data on the share of value added produced in each region by the "machinery" sector and by the "chemical" sector are from the Unido Industrial Statistic Database.

The regional data on schooling (human capital) are from national statistical agencies of the five major countries considered (Germany, France, UK, Italy and Spain). They are therefore limited to the regions of these countries. They were kindly provided by Antonio Ciccone and their sources can be found in Ciccone (2002). For the other countries we have attributed to each region the national average of college graduates in the population from national censuses. We take the share of college graduates (or equivalent degree) in the labor force for the year 1986 (or around it) as measure of human capital. The data on average schooling by country are reported in the appendix of Bottazzi and Peri (2002). The historical population data used to construct the instrumental variables for the 86 regions are from the national census of each country. The exact sources are also reported in detail in Bottazzi and Peri (2002).

References

Audretsch, D., Feldman, M., 1996. R&D spillovers and the geography of innovation and production. American Economic Review 86, 641–652.

Bernstein, J.I., Nadiri, M.I., 1989. Research and development and intra-industry spillovers: An empirical application of dynamic duality. Review of Economic Studies 56, 249–269.

Bottazzi, L., Peri, G., 2000. Innovation and spillovers: Evidence from European Regions, CESifo Working Paper #340, Munich.

Bottazzi, L., Peri, G., 2002. Innovation and spillovers in regions: Evidence from European Patent Data, IGIER Bocconi University Working Paper # 215, Milan.

Ciccone, A., 2002. Agglomeration effects in Europe. European Economic Review 46 (2), 213–227.

Ciccone, A., Hall, R., 1996. Productivity and the density of economic activity. American Economic Review 86 (1), 54–70.

Coe, D.T., Helpman, E., 1995. International R&D spillovers. European Economic Review 39, 859–887.

Conley, T., Flier, F., Tsiang, G., 1999. Local market human capital and the spatial distribution of productivity in Malaysia. Northwestern University, Evanston, IL.

European Commission, 2001. The territorial dimension of research and development policy: Regions in the European research area, Directorate-General for Research, European Commission, February 2001.

Griliches, Z., 1990. Patent statistics as economic indicators: A survey. Journal of Economic Literature XXVIII, 1161–1707.

Hanson, G.H., 1998. Market potential, increasing returns and geographic concentration, NBER Working Paper # 6429, Boston, MA.

Hanson, G.H., 2001. Scale economies and the geographic concentration of industry. Journal of Economic Geography 1, 255–276.

Helliwell, J., 1998. How Much Do National Borders Matter? The Brookings Institution, Washington, DC.

Jaffe, A., 1986. Technological opportunities and spillovers of R&D: Evidence from firms' patents, profits and market value. American Economic Review 76, 984–1001.

Jaffe, A., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citation. Quarterly Journal of Economics 108 (3), 577–598.

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

Keller, W., 2001. International technology diffusion, NBER Working Paper 8573, Boston, MA.

Keller, W., 2002. Geographic localization of international technology diffusion. American Economic Review 92 (1), 120–142.

Lucas, R.E., 1988. On the mechanics of economic development. Journal of Monetary Economics XXII, 3–42.

McCallum, J., 1995. National borders matter: Canada–US regional trade patterns. American Economic Review 85 (3), 615–623.

Peri, G., 2002, Knowledge flows and knowledge externalities, CESIfo Working Papers # 765, Munich.

Romer, P.M., 1990. Endogenous technological change. Journal of Political Economy 98 (5), 71–102.

Saxenian, A., 1991. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard Publishing, Boston, MA.

Shankerman, M., 1979. Essays on the economics of technical change: The determinants, rate of return and productivity impact of research and development. Harvard Ph.D. Thesis, Boston, MA.