INNOCENZO GASPARINI INSTITUTE FOR ECONOMIC RESEARCH

Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Innovation and Spillovers in Regions: *Evidence from European Patent Data*

Laura Bottazzi and Giovanni Peri

Working Paper n. 215

June 2002

IGIER – Università Bocconi, Via Salasco 5, 20136 Milano –Italy
http://www.igier.uni-bocconi.it

Innovation and Spillovers in Regions: Evidence from European Patent Data*

Laura Bottazzi (Università Bocconi and CEPR)
Giovanni Peri (University of California, Davis and CESifo)

June 27, 2002

Abstract

The aim of this paper is to estimate the effect of research externalities across space, in generating innovation.We do so by using R&D and patent data for eighty-six European Regions in the 1977-1995 period. We find that spillovers exist for regions within a distance of 300 Km from each other. The estimates are robust to simultaneity, omitted variable bias, different specifications of distance functions, country and border effects. The size of these spillovers is small, though. 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%. Given the small size and the limited range of diffusion, we interpret these externalities as the result of local diffusion of non-codified knowledge, embodied in people and spreading via personal contacts. This interpretation is reinforced by the finding that the spillovers are somewhat weaker across national borders.

Key Words: Innovation, R&D Spillovers, Europe, Regions.
JEL Classification: O3, R11,R12.

*This Paper was partly written while Giovanni Peri was in Università Bocconi. 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 also provided useful suggestions. We thank Davide di Laurea, Elisa Faraglia and Andrea Gentilini for excellent Research Assistance. Some results of this paper were contained in a previous Working Paper "Innovation and Spillovers: Evidence from European Regions" (CESifo Working Paper # 340, October 2000). We acknowledge Università Bocconi and MURST for funding relevant parts of this project.

"The more difficult to measure and the possibly more interesting and pervasive aspect of R&D externalities is the impact of the discovered ideas ... on the productivity of the research endeavour of others" (Zvi Griliches, 1992)

1 Introduction

Innovation is a fundamental activity for economic growth. New ideas enlarge the stock of knowledge that in turn enhances productivity. The diffusion of knowledge across space and its contribution in the generation of new profitable knowledge is an important issue to evaluate the impact of R&D on productivity and growth. The aim of this work is to identify and estimate the effect of research externalities in generating innovation across space. We do so by looking at R&D and Innovation in eighty-six European Regions for the period 1977-1995 . As economic integration has moved forward in Europe regions, rather than countries, should be considered as units of analysis. They are more homogeneous and better connected within themselves and they are becoming increasingly important as policy units for research and innovation purposes (see for instance the very recent report by the European Commission [9] on European Research Policy). The productivity of R&D resources employed in each region to generate innovative output depends on the strength and the range of diffusion of knowledge spillovers.

The count of patents granted to inventors residing in a region is used as measure of the region's innovative output. Localized R&D spillovers exist if the productivity of R&D in generating innovation in a region is affected by the R&D resources used in other regions. This definition is easily derived as a consequence of a model (such as Romer [22] or Jones [17]) 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 good (hence the externality) or at least a locally public good if their diffusion is easier in spatial proximity. The codified part of a patented idea is likely to be perfectly available to anyone who could read the patent, and therefore can be considered as a fully public good. Nevertheless part of the knowledge generated with the idea is embodied, imperfectly codified, linked to the experience of the scientists and probably "attached" to people. This stock of knowledge increases in a region as local inventors discover new ideas. It could still diffuse but probably it requires much more personal contacts and face to face interactions. We can think of it as a "local public good" as it benefits scientists in the region or close to it but it is sensitive to distanceas because it reduces contacts and interactions.

We estimate the reduced form of the innovation-generating equation for European Regions in the long run. R&D spillovers are identified as the effect of R&D intensity in region s on R&D productivity of region r in generating innovative output. We allow such effect to depend on several characteristics of the regions r and s and importantly on the distance between them.

Our basic result, robust to a series of specifications and controls, is that rather small but statistically significant spillovers come from R&D within 0-300 Km distance. 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 only an effect of 2% on patenting of regions within a 300 Km range. The benefits of R&D in generating innovation are therefore extremely localized. Most of them accrue to the region that spends the R&D resources, a small positive effect diffuses to region within 300 Km distance, no effect at all 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 introduces some relevant estimation issues. Section 5 presents the data and section 6 presents the results of the estimation of several different specifications. Section 7 concludes the paper.

2 Related Literature

After an early contribution by Shankerman [25], Jaffe [15] 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. In that work only the technological dimension of the spillovers was analyzed: R&D done in firms within the same technological cluster was considered as a potential input, besides own R&D, for the innovative activity of a firm. While distance across firms was calculated using a metric in technological (rather than geographical) space, that work crucially used spacial correlation of R&D and innovation to infer the intensity of spillovers. In this respect that seminal work is importantly related to our paper.

Differently from that, most of the recent literature which identifies knowledge spillovers in geographical space using patent data, has followed the approach proposed later by Jaffe et al.[16]. Exploiting the fact that each patent refers to a number of previous patents as "sources" of relevant information, Jaffe et al.[16] and later other researchers (for instance Maurseth and Verspagen [20] for European regions and Sjoholm [26] for Sweden) have used such a "paper trail" to track the direction and extent of spillovers. This approach, though, cannot measure whether the proximity to an innovative unit affects the productivity of R&D in generating innovation in another unit. The "citation" approach can only say if, once a patent comes into existence, its inventors (or possibly its reviewers) acknowledge stronger connections to other patents invented nearby. Moreover citations do not capture non-codified knowledge and embodied know-how which could be the source of important localized spillovers, as Saxenian [23] and Audretsch and Feldman [1] among others argue.

While still using patent count as a measure of innovative activity of one region, we do not consider the link between patents given by citations, but we exploit 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 several other regions at different geographical distances. Such a question cannot be answered by looking at patent citations, but needs the frame of an "innovation function" to be addressed. Once we substitute the "innovation function", whose output are patents, for the production function, whose output is real income, it becomes clear that our approach is equivalent 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 [13].

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 [6] and Keller [18]. Also, in that they consider the effect of human capital density at different distances, Conley et al.[7] and Ciccone [5] deal with the issue of inferring externalities from spatial correlation of productivity. We are the first, though, to our knowledge, to estimate the scope and intensity of spillovers in innovative activity, without any parametric assumption on the dependence of spillovers from distance.

Finally, although indirectly, our paper is related to the literature on regional growth in Europe. While most of the works have concentrated on convergence and GDP growth (since the seminal work of Barro and Sala i Martin [2]), it has been recognized that R&D and innovation are key sectors for European growth, in particular to fill the gap which opened in the 90's with the growth rates experienced in the US (for instance Eaton et al. [8]). Estimating the innovation function in Europe, and identifying the elasticity of innovation to R&D resources is a key issue for Policy decision on R&D.

3 Empirical Model

3.1 Spillovers

Following the theoretical work of Romer [22] and Jones [17], we consider regional units that produce profitable new ideas (patents) using two main inputs: R&D resources and existing ideas. R&D resources employed in region i contribute directly to producing new ideas, as a private input available in that region only. Ideas originated in region i or in any other region could be used as a (public) input of production, as long as they are available and 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 the innovator and an existing idea (or rather its inventor) 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 European region i we propose the following expression:

[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.

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

We have omitted the time subscripts, as we consider only one period for the cross section. ΔA_i represents the change over the considered period of the stock of knowledge originated in European region i, namely the count of new patents granted to researchers in that region. Our proposed specification is a simple generalization of the function assumed in several theoretical papers (e.g. Romer [22]). Expression (1) says that innovation in region i depends on a Cobb-Douglas combination of R&D resources used in region i, $(R\&D)_i$ 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 european 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 $e(k_{ij})$ that depends on the distance between region i and region j $(dist_{ij})$. 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 (and not on the characteristics of each region) as a step function. Hence the function $e(dist_{ij})$ is equal to e_k/n_{ik} for $dist_{ij} \in k$, with $k = \{[dist_0, dist_1), [dist_1, dist_2)..., [K, \infty)\}$ .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 $e_{[K,\infty)} = 0$. The specified diffusion process implies that innovation in region i depends on the average stock of ideas generated in regions within the distance-interval k with different sensitivities (e_k) for different distance-intervals.

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, in the long run, equal across regions, ΔA_i is proportional to A_i. Therefore, substituting for ΔA_i in BGP, solving for the vector $\ln(\underline{A})$ and log linearizing the system [2], we obtain the following approximate relationship between innovation in a region and R&D resources:

$$\begin{aligned}\ln(\Delta A)_i &= \beta + \varepsilon_0 \ln(R\&D)_i + \varepsilon_{[dist_0, dist_1)} \left[\underline{m}'_{1i} \ln \underline{(R\&D)} \right] + ... + \varepsilon_{[dist_n, K)} \left[\underline{m}'_{Ki} \ln \underline{(R\&D)} \right] + u_i \\ i &= 1, 2...86 && (2)\end{aligned}$$

[2]The procedure is shown in detail in the Appendix B

Where $\varepsilon_0 = e_R/(1-e_0)$,and $\varepsilon_k = (e_R e_k)/(1-e_0)^2$ .The parameters $\varepsilon_0, \varepsilon_1, \varepsilon_2 ... \varepsilon_K$, that we estimate, are strictly related to the elasticities of innovation to $R\&D$ and to spillovers from existing ideas generated in other regions.

Equation (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 including that of a diffused pool of knowledge while $(R\&D)_i$ is the average yearly employment (or the real spending in Euros) 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 $k = \{[dist_0, dist_1), [dist_1, dist_2)..., [K, \infty)\}$ 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, in fact, are constructed as follows: row vector $\underline{m}'_{ki}$ is an $86X1$ row vector whose $j-th$ entry is zero if $dist_{ij}$ is not within the range k, while it is $(1/n_{ik})$ if that distance is witihn the range k. $\ln \underline{(R\&D)}$ is a $1X86$ column vector whose $i-th$ entry is $\ln(R\&D)_i$. The product of these two vectors, generates the average $\ln(R\&D)$ for regions in the $k-th$ distance interval from region i. u_i is a random i.i.d. error capturing other randomly distributed determinants of innovative out put. The appeal of specification (2) is that, if regional $R\&D$ has an external impact on innovation that varies with distance, we can identify such effect at any distance, by allowing the elasticity to vary with it.

In order to estimate equation (2) we choose $K = 2000$ Km[3], so that $\varepsilon_{[2000-\infty)} = 0$. As we will see in Section 6, the estimation results are consistent with this assumption, as the elasticity to $R\&D$ of a region's Patenting ceases to be significant at much lower distance than 2000 Km. Also we have to choose the length of each of the intervals into which we divide the whole space (from 0 to 2000 Km) around each region. These intervals should be large enough to have a reasonable number of regions within each of them, but small enough that we can consider all regions in the interval as sharing the same distance from i. Also we want to keep these space intervals rather "balanced" in terms of number of regions included, so as to estimate with the same precision the various coefficients. To meet these criteria we choose intervals of roughly 300 Km, which is about twice the size of the average regional diameter (170 Km), for regions between 0 and 1300 Km of distance and one interval for the distance 1300-2000 Km.

4 Endogeneity and out of BGP bias

Estimation of equation (2) by OLS might be affected by a potentially important simultaneity problem. Assuming that regions grew in their balanced growth path during the considered period (we deal with the issue of deviations from the BGP later) a positive relationship between regional $R\&D$ and regional stock of knowledge (proportional to innovation) exists also due to the efficient allocation

[3] We also check for much smaller cut-off distances

of $R\&D$ across regions. The equilibrium relationship, generated by this channel, would be of the following form:

$$\begin{aligned} \ln(R\&D_i) &= a + b\ln(A_i) + c\ln(Dem_i) \\ i &= 1, 2...86 \end{aligned} \tag{3}$$

We can easily give an intuition of the above equation while leaving its formal derivation to Appendix C, where we also define precisely the variable Dem_i which measures local market potential. In equilibrium, $R\&D$ resources are allocated across regions so as to equate their marginal returns. Regions with larger stock of knowledge A_i and where profitability is higher due to larger local market potential (Dem_i) attract more $R\&D$ resources. This happens because the profitabilty pf patents is higher where market potential is larger. Therefore the equilibrium, in BGP, is defined simultaneously by equation (2) and (3). The way to solve this endogeneity issue is to use the variable Dem_i, which measures the market potential in region i, as instrument. This variable is correlated with the allocation of $R\&D$ in equilibrium (equation 3) , but it does not enter directly the production function of innovation (equation 2). By instrumenting $R\&D_i$ with Dem_i in equation (2) we get an unbiased estimator of the parameters $\varepsilon_0, \varepsilon_1 ... \varepsilon_K$. Intuitively an unequal spatial distribution of demand across regions affects the distribution of $R\&D$ resources, by making more profitable to innovate in regions where potential demand for new goods is higher. Market potential across regions is therefore correlated with $R\&D$ distribution but not with other determinants of $R\&D$ productivity. For this reason it is a good instrument.

We proxy local market potential Dem_i with population density and with a population-based measure of regional market potential. The distribution of population across European regions is mostly historically determined, and related to the presence of harbors, coastal ports and rivers more than to innovation activity. Nevertheless, unobserved shocks, during the considered period (1977-1995), causing deviations from the BGP of both population and stock of ideas, could potentially induce a correlation between the instrument and the unobserved error u_i in equation (2). To avoid this problem we use historical population (measured in 1930) as instrument for current $R\&D$ resources. In this case there is a strong presumption that the only channel through which historical population is correlated to current $R\&D$ is through its long run (BGP) correlation to current population and market potential. This ensures a reduction of the endogeneity bias.

5 The Data

We use a cross section of long-run averages of variables relative to the eighty-six regions of the most important countries of western Europe. The period considered is 1977-1995, although some variables are averaged over shorter intervals due to the unavailability of the whole series. While a more detailed description of the regions and of the data can be found in Appendix D, here we present the main features of our data-set.

The choice of regions within Europe is not straightforward. Eurostat defines different levels of aggregation (called Nomenclature Units Territory Statistics NUTS) and in most cases the first (NUTS 1) or second level (NUTS 2) correspond to what is commonly called a "Region" with some degree of administrative and policy authority. These units are, for instance, the "Landers" in Germany, the "Arrondissments" in France, the "Regioni" in Italy. We choose NUTS 1 or NUTS 2 as regions for our analysis[4]. These regions are rather homogeneous within themselves, have often a strong local identity, are administrative units and therefore meaningful for a sub-national analysis. Only the four smallest countries (Denmark, Ireland, Luxemburg, Portugal) are considered as one region each.

As a measure of the innovative output of a region we use the count of patent applications that inventors residing in each region filed with the European Patent Office within the considered period. We consider as inventor, the first one listed in the patent application. The choice of his residence, rather than the company's address, as the place where the patent is originated is common to this literature (see Jaffe et al [16]) . Since the early work of Schmookler [**?**] patents have been considered, not without controversy, as a measure of the output of the innovative activity of an economic unit (a firm, a sector or a country). Although we know that not all inventions are patented we also know that patents have to fulfill minimal standards of novelty, originality and profitability. Therefore patents can be considered as a good approximation to the ideal data on "economically profitable ideas" which one would like to have for testing theories on innovation.

The pros and cons of patents as a measure of innovation have been discussed by others with care and detail (see the excellent survey in Griliches [10]). 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 or employment as well as with other measures of innovation (such as TFP growth) in the long-run and more so in the cross-section rather than in the time-series dimension (see Griliches [10] page 1673-74). Due to lag in generating the innovation from R&D activity and in transferring the effect of innovation on productivity, the relation between these variables could be appreciated only in the long-run. Second, considering all patents as an equal contribution to new knowledge is likely to generate errors due to the actual variability of their importance, if small units (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 these two sources of variation.

A third reason which makes our choice of data suitable for our analysis is related to the issue of spillovers themselves. We are estimating "technological" externalities in R&D, namely the effect of one region's R&D on other regions'

[4]See Appendix D for the exact choice of regions in each country.

physical productivity in terms of innovation. The fact that we are using the count of patents as measure of innovation rather than some estimate of their value, ensures that we are not polluting our measures with problems of pecuniary externalities, or mis-measurement of the price of a new good, as often happens for measures based on TFP growth (see the discussion of this issue in Griliches [11] page 41-42).

Some limitations of our data are mostly due to $R\&D$ figures. In particular, we do not have the breakdown of regional $R\&D$ (employment or expenditure) by sector. Considering the total patenting of a region as its innovative output, therefore, we have to account for the fact that some sectors have larger propensity to patent than others. In particular, the manufacturing sector patents more than the service sector and the machinery and the chemical sectors within it are the most active in patenting. In order to control for these different propensities, we include the size of manufacturing in the region and of machinery and chemical sectors to account for the presence of these high patenting sectors.

In focussing on the regional dimension of $R\&D$ and innovation and on their geographical distance, we are capturing both the intra-sector as well as the inter-sector spillovers. Hence we do not need to impose discrete and arbitrary boundary to innovation spillovers (e.g. positive spillovers within a sector but none across them[5]), avoiding 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 class, up to 25% of patent citations happens between them[6]. We need to account, though, for technological proximity across regions. Regions which are geographically close might be technologically similar. If spillovers are strong within sector at any distance and are weak across sectors we may observe large spillovers in close regions just because they are technologically similar. To this purpose in section 6.6 we control for a measure of technological proximity developed by Jaffe [15].

As for geographical distance, spillovers due to non-codified or implicit type of knowledge spread more easily over shorter distances as there are costs of traveling and commuting: business trips and contacts are much more frequent within small than large distances. For instance a distance within two-three hundred kilometers is likely to be covered frequently (sometimes daily) by people, while a distance over 2000 kilometers requires a long trip and is unlikely to be covered other than occasionally.

Table 1 reports the summary statistics for the data we use. As shown in the first row, there is rather large variation in the area of the regions. While Table 1 reports the total regional values of the variables, in the empirical analysis of the next section we standardize variables by regional area. $R\&D$ and $Patent$ are therefore going to be the employment (or real spending) in R&D and the yearly patenting per square kilometer. This standardization eliminates the issue

[5] As done, for instance, in Bernstein and Nadiri [3]

[6] See Jaffe et al. [16], pag. 597.

of regional size (in terms of area) as a cause of distortion. Of course regions are still different in their "economic" size (population or employment per square Kilometer). We analyze and exploit that variation in our analysis.

The variation in $R\&D$ intensity across regions is very high, as well as their variation in Patenting. The most active regions spend in $R\&D$ around 3% of their regional resources, generating thousands of successful patent application 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 very different in their economic development (GDP per capita, levels of schooling and sector specialization) ranging from poor, less educated, more rural regions in the Mediterranean belt to high tech, human capital and service-intensive regions, in the Franco-German-Benelux core. These characteristics are likely to have an impact on innovative activity.

In order to group regions in distance classes, as needed to estimate equation (2). We consider the shortest air distance between the boundaries of two regions as the relevant distance between them. Two regions which share a border, therefore, will have zero distance to each other, while regions which do not share a border will have a distance measured by the kilometers between their closest borders. As roads and railways (and certainly flying routes) in Europe connect cities approximating rather well the minimum air distance between them, we think this distance captures the physical distance over which people travel between regions. In order to provide some information on the position of the European regions we report in Table 2 the average distance in kilometers, of each region from the other 85. We have ranked regions by this "average distance". Low values of this variable mean that the region is rather central in Europe while large values mean that the region is peripheral. The average distance from other regions is a pretty good inverse measure of centrality of a region within Europe.

6 Estimation Results

6.1 Basic Specification

The basic specification that we estimate is exactly as described by equation (2), with five distance classes and with twelve country dummies:

$$\begin{aligned}\ln(Patent)_i \; &= \beta + \varepsilon_0 \ln(R\&D)_i + \varepsilon_1 \left[\underline{m}'_{[0-300]} \ln \underline{(R\&D)}\right] + \varepsilon_2 \left[\underline{m}'_{[300-600]} \ln \underline{(R\&D)}\right] \\ &+ \varepsilon_3 \left[\underline{m}'_{[600-900]} \ln \underline{(R\&D)}\right] + \varepsilon_4 \left[\underline{m}'_{[900-1300]} \ln \underline{(R\&D)}\right] + \varepsilon_5 \left[\underline{m}'_{[1300-2000]} \ln \underline{(R\&D)}\right] \\ &+ D_i * (Country)_i + u_i \qquad (4)\end{aligned}$$

The distance classes chosen 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 that range and so that the average number of regions in each interval would be roughly equal to 16 (see Appendix

D for details). Eleven country-dummies $(Country)_i$, capture unobserved factors (legal protection, efficiency of the judicial system, quality of institutions, absorption of technology from the US or Japan) that affect the propensity to innovate in all regions of a country. The parameters of our regression are identified only on the within country cross-regional variation of patenting and $R\&D$.

Table 3 and 4 report the OLS estimation results for the basic specification, with robust standard errors. In the first case (Table 3) we use the regional employment in public and private R&D per square Km of area as measure of $R\&D$. The value used are the 1984-1995 averages. In the second case (Table 4) we use the regional (private and public) $R\&D$ spending in real (1985) Ecus per square Km of area. The results for the two cases are very similar, denoting that Employment is a good measure of total resources used in R&D for a region.

First column of table 3 and 4 include only the effect of R&D done in the region on innovation. In each of the following columns of Table 3 and 4 we include, beginning with the closest range (0-300), one more distance-group 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 long-run elasticity of regional innovation to $R\&D$ in one region in the distance class considered. Let's consider, for instance, specification VI, 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 0.83% its own innovative activity, also increases by a statistically significant 0.025% the innovative activity of regions within a range of 300 Km while all the other regions receive very small effects which are not statistically different from 0. Besides own $R\&D$, only $R\&D$ done within 300 km range has a statistically significant, positive effect on innovation.

Our initial assumption that local spillovers are not significant beyond the distance of 2000 Km seems a safe one, considered the estimates we obtain. The point estimate of the effect of R&D is slightly negative (not significant), beyond 900 Km and, in fact, there seem to be no local spillovers passed 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 effect. Nevertheless, as the average European region has about fifteen other regions within the distance 0-300 Km from itself, the average overall effect if each of them increases by 1% their R&D intensity, is a remarkable increase of 0.39% in the region's innovative output. In particular regions that are centrally located may receive large positive spillovers if all the other regions increase their R&D spending. The best way of thinking about the overall effect of R&D on innovation is the following: 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.

6.2 Sector Composition and Human Capital

Part of the correlation between other regions' $R\&D$ and Patenting could be a spurious result due to omission of relevant determinants of patenting at the regional level. In particular, as we consider aggregate patenting, we should control for the fact that different sectors have different propensities to patent: the manufacturing sector as a whole patents a much larger fraction of its innovations than services and, within manufacturing, chemicals and machinery are, by far, the most active sub-sectors. Certainly, those sectors that patent many inventions are also those investing hevily in R&D. Therefore, sector-differences will be partly accounted for by the measure of total $R\&D$ inputs. We still want to control for the sector structure of regions in order to correct for differentials in productivity of $R\&D$ across sectors.

Table 5, columns I and IV, show the regressions of innovative output on own and other regions' $R\&D$, controlling for the regional size of the manufacturing sector, $\ln(Manufacturing)_i$, and within it, of the Chemical and Machinery sector ($\ln(Machinery)_i$ and $\ln(Chemicals)_i$). Those measures are the natural log of the share in total value added produced by the Manufacturing, Machinery and Chemicals sector respectively. 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 result very significant and positive[7]. More importantly, the inclusion of these controls does not change at all 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 $R\&D$). Highly educated people could contribute with their ideas and expertise and increase innovative output in a region. For this reason in column II and V of Table 5 we include, as explanatory variable, the log of the share of college graduates in the population of a region. The difference between the two regressions is, as usual, that specification II uses R&D Employment while specification V uses R&D spending as measure of resources employed in $R\&D$. The effect of these variables is large and significant on innovative output,confirming the important role of Human Capital for innovation. However the effect on the elasticity to own R&D and R&D in the 0-300 Km range is very small. The externality remains positive and significant, with a coefficient close to 0.03.

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 effect by including the intensity of human capital in the 0-300 Km. range as an explanatory variable.

[7]We also run regressions controlling for the shares of the other 2-digits manufacturing sectors but their coefficient was never significant.

This is what we do in column III and VI and we do not find any evidence of such human capital-externalities, while the R&D spillovers remain basically unchanged.

6.3 Distance Breakup and Centrality

In the previous two sections we chose the distance classes for regions as a reasonable compromise between a fine breakup of space and the accuracy of the estimates. Certainly, though, there is an amount of arbitrariness in those distance intervals. We are aware that increasing the "finesse" of the distance-grid would decrease the number of regions in each interval and increase the number of regressors, decreasing, as a consequence, the precision of the estimates. Nevertheless, to be sure of the robustness of our results we inquire in greater detail into what happens within the closest 1000 Km of distance, whan we define distance brackets of 100 Km's only.

In the light of the previous results we are confident that we are not missing much by limiting our analysis within 1000 Km, as we do not want to increase unnecessarily the number of regressors given our limited number of observations. All the intervals have an average number of regions between 5.2 and 6.3. As several regions have some 100 Km intervals without any region in it, in order to fill these gaps for all regions we interpolate the averages of the two closest intervals. We are, basically, making an assumption of continuity in space, and approximating the unknown $R\&D$ average intensity within a range with the average within the two ranges next to it. Once we have evaluated the average $R\&D$ at any distance with this method we estimate few specifications.

Table 6 shows the results of the analysis with 100 Km intervals considering, as usual, both Employment (in specification I-II-III) and Spending (in specification IV-V-VI) as measures of $R\&D$ intensity. In general, the only spillovers' coefficients which are always positive and frequently 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, while still positive, is not statistically different from 0. The std. 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. Sometimes std. errors are quite large on these estimates and point estimates are almost always smaller than the std. errors.

Column I and IV report the results of the basic specification which includes 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[8]. This seems true across specifications, but a qualification applies. If we run the regression including only the first three distance classes (from 0 to 300), the size of the coefficients is decreasing as distance increases. For instance in the basic specification as in column I, with only three distance classes, we have coefficients

[8]Interestingly, this result was found also in our previous analysis Bottazzi and Peri [4] and a similar finding is also in Conley et. al [7]

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 6, is 0.029, which is exactly equal to the estimate of the average effect of R&D in the 0-300 Km range from the equivalent specification II in Table 5. The corresponding average effect estimated using R&D spending, in column IV, Table 6 is 0. 042, which is somewhat larger than the 0.026 estimate in Table 5 column V. Column II and V simply include all the regional controls as in column II and V of Table 5 and confirm the same findings as the basic specification.

An important and interesting issue, concerning the location of innovative activity and its productivity within Europe, is the "absolute" versus the "relative" location of a region. So far we have considered the position of a region "relative" to the *R&D* done at several distance-ranges from it and the effect of such *R&D* on the region's innovative output. We have also suggested, in the description of the data, that Europe has a "center", located between northern France, northern Germany and Benelux, where the intensity of its economic and innovative activity is highest, and a "periphery", spread along the Mediterranean coastline, which is less advanced. It could be the case that the variables measuring *R&D* at different distances are just proxing for the "centrality" of a region within Europe. *R&D* within the 0-300 Km range is high for those regions located in the central part of Europe and their innovative output is large since the center of Europe is the best location for any economic activity. In order to distinguish this "absolute location" explanation from the spillovers' one, we include in the regression an index of a region's absolute position within the "European Island". In particular we use an index of centrality ($\ln(Centrality)$) given by the log of the inverse average distance of the region from all the other included European regions (as described at the end of Section 5). Given the high correlation of this index with *R&D* intensity, its inclusion is certainly bound to reduce the significance of the spillovers coefficients. Nevertheless we want to see if all the variation identifying externalities comes from the degree of centrality of regions or if peripheral and semi-peripheral regions with "high R&D neighbors" help identifying those externalities.

Column III and VI of Table 6 show the results of including this index in the regression with the 100 Km distance ranges. The coefficient of $\ln(Centrality)$ is highly significant and reduces somewhat the effect of *R&D* in the closest regions. Still *R&D* in the 100-200 range remains significant. If we consider the specification for the 300-Km intervals (not reported) analogous to Column II and V of Table5, with the inclusion of $\ln(Centrality)$, we obtain that the coefficient of *R&D* in the 0-300 Km range is still large and significant (0.027 with std. error equal to 0.11 for *R&D* employment and 0.028 and std. error of 0.010 for the *R&D* Spending). We take this as evidence that, although the absolute position within Europe is an important determinant of innovative output, spillovers from closer regions are relevant. The magnitude of these spillovers, which was estimated to be on average 3% in the previous section for

$R\&D$ in the 0-300 Km range, here for the regions in the 0-200 Km range seem to be potentially as high as 4-6%. The elasticity of innovation to R&D of a bordering region could be as large as 0.06.

6.4 Parametric Decay of Elasticity 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 form of diffusion (and therefore of decay with distance) of the spillovers. The data themselves suggest a rather "discontinuous" behavior of spillovers, with a positive effect within a range of 200 Km (possibly 300). We can also approximate this behavior using a parametric function to weight $R\&D$ at different distances from the region. The goal of this approach is to confirm our previous results which show a rapid decline in the effect of $R\&D$ with distance and to provide potentially more precise estimates of spillover's intensity at zero-distance. In fact, if we approximate correctly the decay-function of spillovers, 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 [18]) have estimated parametric functions of decay for R&D spillovers over space. We use, therefore, the most popular functional forms used in the literature in our estimates.

We construct a "pool" of $R\&D$ resources, potentially available to a region assuming that the "effective" $R\&D$ contribution by each of the other regions depends parametrically on the distance between that region and the one for which we are constructing the "pool". Call $F(d)$ a "decay" function of distance, d, such that $1 \geq F(d) \geq 0$ and $F(0) = 1$. We still divide the space around each region in 100 Km intervals, now considering the average $R\&D$ done in each of them and we use one hundred Km as the unit of distance. A region at distance 100 Km has $d = 1$. Then 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 function used are exactly: $F(d) = 1 - (d/20)$ in the linear case, $F(d) = (1 + d)^{-1}$ for the inverse and $F(d) = e^{-d}$ 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. In particular, the linear case which is the one where decay happens more slowly still attributes a weight of 0.75 to R&D at 500 Km and of 0.5 to $R\&D$ at 1000 Km. The inverse decay attributes a weight of 0.16 to R&D at 500 Km while at 1000 Km it gets a weight of 0.09. Finally the Exponential function implies the fastest decay and attributes a weight of 0.0067 to R&D at 500 Km and basically 0 to R&D at 1000 Km. We assign the distance of the mid-interval to each range (i.e. the 0-100 gets 50 and so on).

The coefficient of the linear 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 from the regions sharing a border. Table 7 reports the results, using the pool of $R\&D$ ($Pool(R\&D)_i$)

variable calculated using the different functions. Column I and II use the linear specification, column III and IV use the Inverse and column V and VI use the exponential specification. All specifications include own R&D and the regional controls, while the second specification for each functional form include also the centrality index. We only report the results using R&D spending as measure of R&D, those using R&D employment are rather similar.

Looking at the coefficients for $Pool(R\&D)_i$ in column I, III and V we notice that the significance and the size of the coefficient increases going from the linear to the inverse to the exponential specification. Also, the overall fit of the model (R^2) is slightly better for the exponential case. This is very consistent with our non-parametric specification: the linear functions gives too much weight to R&D done far from the region and, trying to fit a linear function 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 performs best as it attributes non negligible weight only to the first 2-3 intervals with values decreasing very fast. This specification estimates that an increase of 1% in R&D on the border generates and external effect of 0.05% on regional innovation. This spillovers decrease very quickly with distance so that at 200 Km the effect is 0.008% and at 500 Km it is null.

6.5 Simultaneity and Historical Population as Instrument

To be sure that our estimates are not affected by reverse causality between R&D and Innovation we instrument $R\&D$ intensity in a region with an exogenous source of its variation, which is the pre-determined market potential of the region. We use two measures of the historically pre-determined market potential of a region in Europe. First, simply the density of its population in 1930, $\ln(Pop_{30})_i$: denser regions have higher demand and production and therefore generate more direct and indirect demand for new goods. Moreover, probably no determinants of today's innovative ability of a region are much affected by its population in 1930. Second, we account for the fact that as goods are traded, the demand from other regions affect market potential. If there are transportation costs, farther regions have smaller effect than closer regions. Therefore we calculate "historical market potential", $\ln(MkPot_{30})_i$ as the sum of the region's population density and the other regions' density (in 1930) weighted with an exponentially decreasing function whose rate of decay is -0.03 times the distance in thousands of kilometers. This parameter is taken from the estimate in Hanson [12] on how the local market potential depends on distance.

These instruments should be correlated with $R\&D$ as they are correlated with the market size of a region and therefore with the profitability of new goods in a region, while they should be uncorrelated with other determinants of the innovation function (efficiency in the use of R&D or innovative capacity of a region). Therefore we can use them to estimate the internal and external returns to $R\&D$. Table 8 shows these estimates: specification I, II and III include own $R\&D$ and $R\&D$ in the 0-300 Km range, while column IV, V and VI estimate the specification with own $R\&D$ and the exponentially weighted $Pool(R\&D)$.

In the first case we use, as instruments 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 $Pool(Pop_{30})_i$, where the pool variable is constructed in the same way as for the $R\&D$ pool, just using historical population rather than $R\&D$. Very simply, as each region's $R\&D$ intensity should be instrumented by the region's historical population density, we instrument all the transformation of regional $R\&D$ using the analog transformation of regional historical population Pop_{30}. We choose to include only $R\&D$ in the 0-300 Km range as all the non-parametric estimates consistently 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 7. Table 9 is identical to 8, just using the market potential $MkPot_{30}$ as an instrument rather than Pop_{30}.

The instruments (both $MkPot_{30}$and Pop_{30}) are excellent as they explain between 72 and 78% of the independent variables's variations. Also they are highly correlated and the results obtained using population are very similar to those obtained using market potential as an instrument. 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 IV column 2) 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 specification I and II the effect of R&D within 0-300 Km, is still significant and very close to its OLS estimate. Only when we introduce the centrality index the coefficient on $\left[\underline{m}'_{[0-300]}\ln(\underline{R\&D})\right]$ decreases somewhat and looses significance. Similarly, the estimated effect of the $Pool(R\&D)$ is rather similar to its OLS estimates (in column V and VI of Table 7), and roughly between 4 and 6%. While the IV 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.

6.6 Technological Distance

As we consider the aggregate effect of R&D spillovers on innovation it is hard to know which part is due to inter-sector and which 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 [15]. For each region, we divide their patents in thirty technological classes, following the IPC (International Patent Code) classification. Each region is then assigned a 30X1 "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 correlation coefficient of their technological vectors. Two regions which patent exactly in the same proportion in each class have an index equal to one. Two regions which 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 10 the effect of controlling for technological distance. The variable $\left[\underline{m}'_{TECH.DIS}\ln\underline{(R\&D)}\right]$ in Table 10 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 10 reports the estimates for the specifications with controls and country dummies when we include this variable, both using OLS (column I and II) and 2SLS (column II and IV) method. We include the control $\underline{m}'_{TECH.DIS}\ln\underline{(R\&D)}$ both with the non-parametric (specification I and III) and with the parametric specification (II and IV) of distance spillovers.

The estimates show that technological proximity matter. 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 is still existing and significant. Also when we measure it using the exponentially weighted *R&D Pool*, we get significant coefficient between 0.04 and 0.06. Table 10 confirms that spillovers are affected by geographical distance and not only by technological distance. Being close in space is a factor that genuinely affect the diffusion of spillovers. We leave to further research a more careful analysis of the interaction between these two dimension of knowledge diffusion.

6.7 Border Effect on Spillovers

It is a common finding in the literature which studies flows of people and goods, that national borders constitute a hurdle to exchange. As shown originally by McCallum [21] and confirmed in several works by Helliwell (notably in Helliwell [14]) 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. As above 600 Km of distance there are extremely few regions in the same country, we only need four groups of regional R&D intensity, which are, as reported in Table 10: $\left[\underline{m}'_{[0-300]within}\ln\underline{(R\&D)}\right]$, $\left[\underline{m}'_{[0-300]across}\ln\underline{(R\&D)}\right]$, $\left[\underline{m}'_{[300-600]within}\ln\underline{(R\&D)}\right]$ and $\left[\underline{m}'_{[300-600]across}\ln\underline{(R\&D)}\right]$. These groups capture, respectively average R&D intensity in regions within 300 Km in the same country, in regions within 300 Km but in a different country, in regions within 300-600 Km in the same country and in regions within 300-600 Km in a different country. 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 11 report the result of the basic equation, using the four groups, without and with country dummies. The country dummies could absorb part of the effect of within-country spillovers and therefore we estimate one specification also without them.

The point estimates of the coefficients is larger for the within-country spillovers in both specification. 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 column 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. Only in the specification without Country Dummies though the estimate of national regions spillovers is significantly positive. Also, in spite of their difference, the coefficients for national and foreign regions are not statistically different at standard confidence level. Column V and VI uses 2SLS estimation, using the variables 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 significance the point estimate for national regions' R&D is still quite larger than that for foreign regions' R&D.

On average we think that the data show some evidence that the spillovers are stronger, at a given distance, when coming from regions in the same country. In the 0-300 Km range they seem to be, if at all, only slightly stronger for national regions. In the 300-600 Km range, while there are no spillovers from foreign regions, there might still be some weak spillovers from regions of the same country. All in all a border effect is present but does not seem to be large. While this may seem surprising we should emphasize that the inclusion of country effects and the relatively small variation of within country R&D intensity versus the larger variation of across countries R&D intensity may result in difficulties in identifying within country versus across countries effects.

7 Conclusion

Knowledge spillovers are unobservable. While some growth theorists have evoked their pervasive presence to justify sustained growth (notably Lucas [19] and Romer [22]) it is the empirical analysis which 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 are available to other researchers in their quest for profitable innovations. We believe that new knowledge, when codified, is available to everybody and therefore is a public good which affects the potential for new ideas everywhere in the world. However, new ideas which are not perfectly codified are "embodied" in people. This could be one of the reasons why innovative activity tends to cluster (as argued in Audretsch and Feldmand [1]), and why innovators in an area tend to cite ideas from their neighbors more frequently (Jaffe et al.[16]).

Our work tackles exactly this issue by estimating what is called an "innovation generating" function at the regional level for Western Europe in the long run. In particular, we estimate the elasticity of innovation to R&D done in other regions at various distances, an important parameter for macro and policy purposes. We estimate this elasticity to be positive and significantly different from 0 only for R&D done within 300 Km of distance from a region. Its magnitude, though, is quite small: doubling R&D in a region would increase by 2-3% the patenting activity in another region within 300 Km of distance. Closer to its border (within the first 100 Km) and for regions of the same country the effect could be as large as 5-6%. The small size and the short range of these effects is consistent with the idea that such spillovers are the result of diffusion of non-codified knowledge between people who have frequent (weekly or monthly) interactions. It is reasonable to believe that in Europe people commute and interact quite frequently within regions and in their vicinity, while much less so if a longer trip is required. Moreover they probably commute and interact more within than across countries and therefore a (small) border effect on these spillovers is detected. The range of these spillovers could very well be that of frequent face-to-face interactions, while the rest of the knowledge flows in codified format and is not sensitive to distance.

References

[1] Audretsch D. and M. Feldman (1996), "R&D Spillovers and the Geography of Innovation and production", *American Economic Review* 86, 641-52.

[2] Barro R. and J. Sala i Martin (1991) "Convergence across States and Regions" *Brookings Papers on Economic Activity*, 1 1991 pp. 107-82.

[3] Bernstein J.I. and M.I. Nadiri (1989) "Research and Development and Intra-industry spillovers: an Empirical Application of Dynamic Duality" *Review of Economic Studies*, 56 249-269.

[4] Bottazzi L. and G. Peri (2001) "Innovation and Spillovers: Evidence from European Regions" CESifo Working Paper, #340, March 2001

[5] Ciccone A. (1998) "Agglomeration-Effects in Europe" Mimeo Universitat Pompeu Fabra.

[6] Coe D.T. and E. Helpman (1995) "International R&D Spillovers" *European Economic Review* 39: 859-887.

[7] Conley T., F. Flier, and G. Tsiang (1999) "Local Market Human Capital and the Spatial Distribution of Productivity in Malaysia," Mimeo, Northwestern University, February.

[8] Eaton J., E. Gutierrez and S. Kortum (1998) " European Technology Policy" *Economic Policy,* 1998.

[9] European Commission (2001) "The territorial Dimension of Research and Development Policy: Regions in the European Research Area" - by Dr Achilleas MITSOS, Director-General, Directorate-General for Research, European Commission (Feb. 2001) available at http://europa.eu.int/comm/research/area.html

[10] Griliches Z. (1990) "Patent Statistics as Economic Indicators: A Survey" *Journal of Economic Literature* Vol.XXVIII pp.1161-1707.

[11] Griliches Z. (1992) " "The Search for R&D Spillovers", *Scandinavian Journal of Economics* 94: 29-47

[12] Hanson G.H. (1998) "Market Potential, Increasing Returns and Geographic Concentration" NBER Working Paper # 6429.

[13] Hanson G.H. (2000) "Scale Economies and the Geographic Concentration of Industry" NBER Working Paper # 8013.

[14] Helliwell J. (1998) "How Much Do National Borders Matter?" The Brookings Instiution, Washington D.C.

[15] Jaffe A. (1986) "Technological Opportunities and Spillovers of R&D: Evidence from Firm's Patents, Profits and Market Value" *The American Economic Review,* Vol. 76 984-1001.

[16] Jaffe A., M. Trajtenberg and R. Henderson (1993) "Geographic localization of knowledge spillovers as evidenced by patent citation" *Quarterly Journal of Economics* 108 (3) 577-98.

[17] Jones C. (1995) "R&D based models of economic growth" *Journal of Political Economy* 103, 739-784

[18] Keller W. (2000) "Geographic Localization of International Technology Diffusion" NBER Working Paper # 7509.

[19] Lucas R. E: (1988) "On the Mechanics of Economic Development" *Journal of Monetary Economics* XXII (1988), 3-42.

[20] Maurseth P.B and B.Verspagen (2000) "Knowledge Spillovers in Europe. A patent-citation analysis" mimeographed MERIT Maastricht University, September 2000.

[21] J. McCallum (1995) "National Borders Matter: Canada-US Regional Trade Patterns" *American Economics Review* 85(3) 615-623

[22] Romer P.M. (1990),"Endogenous technological Change", *Journal of Political Economy* 98(5) part2:71-102

[23] Saxenian A. (1991) "Regional Advantage: Culture and Competition in Silicon Valley and Route 128" Harvard Publishing.

[24] Schmookler J. (1954) "The level of Inventive Activity" *Review of Economics and Statistics*, 36(2) pp183-90

[25] 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.

[26] Sjoholm F.(1996) "International Transfers of Knowledge: The role of international trade and Geographic Proximity" *Weltwirthshaftliches Archiv* 132 pp. 97-115.

A Tables

Table 1: Summary statistics (Averages 1977-95)

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 $R\&D^*$	1.31	0.88	0.10	3.8
% of Private $R\&D$ in Total*	56	24	2	100
% of Employment in R&D*	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
Patents Application per year	178	356	1	1838
Ln (per capita real GDP)***	2.35	0.27	1.57	3.06
% of College educated**	10	4	3.8	25
84 European Regions				
* relative to 1984-95 only. ** relative to year 1986, *** Millions of 1985 ECU				

Nuts CODE	Capital City	Ave. Distance from other Regions	Nuts CODE	Capital City	Ave. Distance from other Regions	Nuts CODE	Capital City	Ave. Distance from other Regions in KM
FR26	Dijon	592	IT31	Trento	737	UK2	Sheffield	942
FR71	Lyon	605	BE1	Brussels	739	IT71	L'aquila	946
DEB	Mainz	608	FR25	Caen	739	ES4	Valladolid	956
FR41	Nancy	611	UK5	London	740	DEF	Kiel	959
FR21	Reims	615	IT4	Bologna	746	DE6	Hamburg	961
DE1	Stuttgart	625	FR62	Toulouse	750	UK8	Manchester	991
FR24	Orleans	629	IT32	Venice	760	DE3	Berlin	995
FR43	Besancon	641	DE9	Hannover	764	UK1	Newcastle	1017
DEA	Cologne	656	NL4	Eindhoven	765	ES1	Oviedo	1018
FR72	Clermont-Ferrant	657	IT51	Florence	766	ITB	Cagliari	1027
DE2	Munich	662	FR51	Nantes	768	ES7	Palma	1059
BE3	Liegi	664	NL3	Amsterdam	769	IT72	Campobasso	1065
FR42	Strasbourg	669	NL2	Arnhem	770	IT8	Naples	1066
FR1	Paris	671	FR61	Bordeaux	780	IE	Dublin	1079
IT2	Milan	672	NL1B	Groningen	813	DK	Copenhagen	1093
DE7	Frankfurt	673	UK4	Norwich	823	UKA	Glasgow	1129
BE34	Lussemburgo	677	IT33	Trieste	838	IT91	Bari	1150
IT11	Turin	678	IT52	Perugia	853	IT92	Potenza	1152
DEC	Saarbruchen	685	UK6	Bristol	868	UKB	Belfast	1200
FR22	Amiens	688	ES5	Barcelona	871	ES3	Madrid	1206
FR3	Lille	690	FR83	Ajaccio	879	ITA	Palermo	1214
IT12	Aosta	694	UK7	Birmingham	887	IT93	Catanzaro	1291
FR63	Limoges	695	IT53	Ancona	897	ES6	Siviglia	1392
BE2	Antwerp	696	UK3	Nottingham	905	GR1	Salonicco	1433
FR53	Poiteirs	705	FR52	Brest	905	PT1	Lisbon	1506
FR81	Nimes	710	UK9	Cardiff	908	GR2	Patrasso	1540
FR82	Marseille	730	DE5	Bremen	917	GR3	Athens	1791
IT13	Genoa	731	ES2	Zaragoza	929	GR4	Heraklion	2004
FR23	Le Havre	732	IT6	Rome	931			

Table 2: Distance from the other Regions in Km.

Table 3: 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)
$\underline{m}'_{[0-300]} \ln \underline{(R\&D)}$		0.027** (0.011)	0.025** (0.011)	0.027** (0.011)	0.027** (0.011)	0.025** (0.012)
$\underline{m}'_{[300-600]} \ln \underline{(R\&D)}$			0.01 (0.017)	0.004 (0.017)	0.003 (0.018)	−0.008 (0.019)
$\underline{m}'_{[600-900]} \ln \underline{(R\&D)}$				0.015 (0.013)	0.015 (0.013)	0.010 (0.013)
$\underline{m}'_{[900-1300]} \ln \underline{(R\&D)}$					−0.003 (0.012)	−0.005 (0.12)
$\underline{m}'_{[1300-2000]} \ln \underline{(R\&D)}$						−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. Robust std. errors in parenthesis						
* = Significant at 90%, ** = Significant at 95%						

Table 4: Basic Specifications Using R&D Spending

Variables	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.95** (0.05)	0.83** (0.06)	0.82** (0.06)	0.82** (0.06)	0.82** (0.06)	0.80** (0.06)
$\underline{m}'_{[0-300]} \ln \underline{(R\&D)}$		0.030** (0.010)	0.028** (0.001)	0.029** (0.011)	0.026** (0.011)	0.025** (0.011)
$\underline{m}'_{[300-600]} \ln \underline{(R\&D)}$			0.004 (0.01)	0.003 (0.011)	0.002 (0.012)	−0.007 (0.013)
$\underline{m}'_{[600-900]} \ln \underline{(R\&D)}$				(0.004) (0.012)	0.005 (0.013)	−0.004 (0.012)
$\underline{m}'_{[900-1300]} \ln \underline{(R\&D)}$					−0.010 (0.010)	−0.007 (0.012)
$\underline{m}'_{[1300-2000]} \ln \underline{(R\&D)}$						−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. Robust std. errors in parenthesis						
* = Significant at 90%, ** = Significant at 95%						

Table 5: 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)
$\underline{m'_{[0-300]} \ln \underline{(R\&D)}}$	0.026** (0.012)	0.029** (0.012)	0.028** (0.011)	0.025** (0.011)	0.026** (0.010)	0.029** (0.011)
$\underline{m'_{[300-600]} \ln \underline{(R\&D)}}$	−0.010 (0.019)	−0.010 (0.018)	−0.003 (0.02)	−0.010 (0.014)	−0.010 (0.014)	−0.010 (0.015)
$\underline{m'_{[600-900]} \ln \underline{(R\&D)}}$	0.006 (0.013)	0.007 (0.013)	0.004 (0.014)	−0.007 (0.014)	−0.006 (0.012)	−0.006 (0.013)
$\underline{m'_{[900-1300]} \ln \underline{(R\&D)}}$	0.001 (0.12)	−0.004 (0.012)	−0.006 (0.012)	−0.006 (0.012)	−0.007 (0.11)	−0.007 (0.011)
$\underline{m'_{[1300-2000]} \ln \underline{(R\&D)}}$	−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)
$\ln(College)_i$		0.92** (0.35)	0.95** (0.36)		0.77** (0.33)	0.76** (0.34)
$\underline{m'_{[0-300]} \ln \underline{(College)}}$			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. Robust std. errors in parenthesis

* = Significant at 90%, ** = Significant at 95%

Table 6: Distance Breakup and Centrality

Used as $R\&D$ measure:	$R\&D$ Employment			$R\&D$ Real Spending		
Specification:	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.80** (0.07)	0.77** (0.07)	0.74** (0.07)	0.80** (0.07)	0.76** (0.07)	0.71** (0.07)
$\underline{m'_{[0-100]} \ln(R\&D)}$	0.039* (0.022)	0.027 (0.20)	0.001 (0.020)	0.04** (0.02)	0.038* (0.021)	0.01 (0.02)
$\underline{m'_{[100-200]} \ln(R\&D)}$	0.042** (0.018)	0.036** (0.16)	0.038** (0.015)	0.07** (0.025)	0.061** (0.029)	0.035* (0.19)
$\underline{m'_{[200-300]} \ln(R\&D)}$	0.006 (0.02)	0.025 (0.20)	0.026 (0.021)	0.02 (0.02)	0.032 (0.021)	0.025 (0.021)
$\underline{m'_{[300-400]} \ln(R\&D)}$	−0.020 (0.025)	−0.03 (0.024)	−0.01 (0.023)	−0.01 (0.023)	−0.01 (0.021)	0.0002 (0.021)
$\underline{m'_{[400-500]} \ln(R\&D)}$	0.002 (0.015)	0.008 (0.014)	−0.007 (0.014)	−0.002 (0.012)	−0.004 (0.012)	−0.004 (0.012)
$\underline{m'_{[500-600]} \ln(R\&D)}$	−0.025 (0.030)	−0.030 (0.025)	−0.03 (0.025)	−0.03 (0.024)	−0.03 (0.024)	−0.03 (0.22)
$\underline{m'_{[600-700]} \ln(R\&D)}$	−0.03 (0.024)	−0.03 (0.023)	−0.03 (0.023)	−0.04 (0.026)	−0.04 (0.026)	−0.03 (0.025)
$\underline{m'_{[700-800]} \ln(R\&D)}$	0.02 (0.028)	0.004 (0.028)	0.015 (0.026)	0.02 (0.025)	0.006 (0.024)	0.02 (0.024)
$\underline{m'_{[800-900]} \ln(R\&D)}$	0.006 (0.026)	−0.003 (0.024)	−0.006 (0.024)	−0.002 (0.024)	−0.003 (0.023)	0.001 (0.022)
$\underline{m'_{[900-1000]} \ln(R\&D)}$	0.02 (0.02)	0.018 (0.02)	0.007 (0.02)	0.02 (0.021)	0.019 (0.020)	0.01 (0.020)
Other Controls	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
$\ln(Centrality)_i$	*No*	*No*	1.78** (0.56)	*No*	*No*	1.57** (0.62)
12 Country Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
R^2	0.83	0.86	0.89	0.85	0.87	0.
Observations	86	86	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. Robust std. errors in parenthesis

* = Significant at 90%, ** = Significant at 95%

Table 7: Parametric Specification Using R&D Spending

Decay of Weighting	*Linear*		*Inverse*		*Exponential*	
Specification	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.76** (0.06)	0.77** (0.06)	0.77** (0.06)	0.77** (0.06)	0.78** (0.06)	0.78** (0.06)
$Pool(R\&D)_i$	0.012* (0.007)	0.010 (0.07)	0.034** (0.013)	0.025** (0.012)	0.072** (0.024)	0.056** (0.025)
$OtherControls$	Yes	Yes	Yes	Yes	Yes	Yes
$\ln(Centrality)_i$	No	1.24** (0.51)	No	1.14** (0.51)	No	0.98** (0.51)
12 Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.88	0.89	0.89	0.89	0.89	0.90
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 $\ln(College)$.

Cross-Section Using 1977-1995 Averages. Robust std. errors in parenthesis

* = Significant at 90%, ** = Significant at 95%

Table 8: IV Estimation (Historical Population)

Variables	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.91^{**} (0.071)	0.89^{**} (0.071)	0.91^{**} (0.07)	0.89^{**} (0.07)	0.89^{**} (0.07)	0.89^{**} (0.07)
$\underline{m'_{[0-300]}}\ln\underline{(R\&D)}$	0.021^{**} (0.011)	0.0181^{*} (0.011)	0.016 (0.011)			
$Pool(R\&D)_i$				0.07^{**} (0.024)	0.066^{**} (0.024)	0.044^{*} (0.024)
Other Controls	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
$\ln(Centrality)$	*No*	*No*	1.18^{**} 0.53	*No*	*No*	1.17^{**} (0.53)
12 Country Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
R^2	0.87	0.89	0.89	0.87	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)

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. Std. errors in parenthesis

* = Significant at 90%, ** = Significant at 95%

Table 9: IV Estimation (Historical MKT Potential)

Variables	I	II	III	IV	V	VI
$\ln(R\&D)_i$	0.89^{**} (0.071)	0.87^{**} (0.071)	0.089^{**} (0.07)	0.85^{**} (0.071)	0.081^{**} (0.07)	0.084^{**} (0.07)
$\underline{m'_{[0-300]}}\ln\underline{(R\&D)}$	0.20^{*} (0.010)	0.19^{*} (0.010)	0.016 (0.010)			
$Pool(R\&D)_i$				0.07^{**} (0.026)	0.062^{**} (0.025)	0.043^{*} (0.025)
Other Controls	*No*	*Yes*	*Yes*	*No*	*Yes*	*Yes*
$\ln(Centrality)$	*No*	*No*	1.14^{**} (0.52)	*No*	*No*	1.16^{**} (0.53)
12 Country Dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
R^2	0.87	0.88	0.89	0.86	0.88	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)

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. Std. errors in parenthesis

* = Significant at 90%, ** = Significant at 95%

Table 10: technological distance

Variables	I[a]	II	III	IV
$\ln(R\&D)_i$	0.69**	0.67*	0.82*	0.86*
	(0.07)	(0.06)	(0.08)	(0.08)
$\underline{m}'_{[0-300]} \ln \underline{(R\&D)}$	0.025*		0.019*	
	(0.009)		(0.01)	
$Pool(R\&D)_i$		0.058*		0.043*
		(0.02)		(0.021)
$\underline{m}'_{TECH.DIS} \ln \underline{(R\&D)}$	0.03*	0.03*	0.021*	0.015
	(0.01)	(0.01)	(0.01)	(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. Std. errors in parenthesis

[a]In specification I we also include $R\&D$ in the other distance-intervals up to 2000 Km.

* = Significant at 90%, ** = Significant at 95%

Table 11: 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)
$\underline{m'_{[0-300]within} \ln \left(R\&D\right)}$	0.028 (0.018)	0.042 (0.025)				
$\underline{m'_{[0-300]across} \ln \left(R\&D\right)}$	0.023* (0.008)	0.030** (0.011)				
$\underline{m'_{[300-600]within} \ln \left(R\&D\right)}$	0.010 (0.013)	0.010 (0.02)				
$\underline{m'_{[300-600]across} \ln \left(R\&D\right)}$	-0.003 (0.007)	-0.002 (0.008)				
$\underline{m'_{[0--600]within} \ln \left(R\&D\right)}$			0.032** (0.015)	0.057 (0.035)	0.030 (0.020)	0.04 (0.037)
$\underline{m'_{[0-600]across} \ln \left(R\&D\right)}$			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. Std. errors in parenthesis

* = Significant at 90%, ** = Significant at 95%

B Innovation and BGP growth

Writing the function (1) under the assumption that innovation in region i depends on the average stock of ideas generated at different distances, we have the following specification:

$$\Delta A_i = B(R\&D)_i^{e_R} A_i^{e_o} \left(\prod_{j=1}^{n_1} A_j^{\frac{1}{n_1}} \right)^{e_1} \left(\prod_{j=1}^{n_2} A_j^{\frac{1}{n_2}} \right)^{e_2} ... \left(\prod_{j=1}^{n_K} A_j^{\frac{1}{n_K}} \right)^{e_K} \quad (5)$$

Dividing both sides by A_i we obtain the growth rate of the stock of knowledge $g(A_i) = \Delta A_i / A_i$ on the left hand side and then taking again the rate of change of both sides we get the following expression:

$$\Delta g(A_i) = g(A_i) \left[g(B) + e_R g(R\&D) + (e_0 - 1) g(A_i) + e_1 g(\overline{A}_1) + ... + e_K g(\overline{A}_K) \right] \quad (6)$$

$g(\overline{A}_k)$ is the growth rate of the average stock of locally generated knowledge in regions at distance k from region i. The system is in balanced growth path if all regions grow at the same rate, and this happens if the right hand side of the expression (6) is equal to 0 for all i's. In this case $g(\overline{A}_k) = g(A_i) = g$ and we can solve to find the BGP common growth rate of knowledge:

$$g = \frac{g(B) + e_R g(R\&D)}{1 - e_0 - e_E} \quad (7)$$

The growth rate of regional stocks of knowledge is proportional to the growth rate of B the world-generated common pool of knowledge plus the growth rate of $R\&D$ resources, multiplied by e_R the elasticity of innovation to own $R\&D$. Also the stronger the spillovers from existing ideas, generated in the region (e_0) or generated elsewhere ($e_E = \sum_k e_k$), the larger the growth rate of innovation in BGP. Convergence to BGP is granted, locally, if $e_0 + e_E < 1$.

If we log -linearize expression (5) in BGP, using the property that $\ln(A_i) + \ln(g) = \ln(\Delta A_i)$, and collect in one single vector $\underline{C}$ all the vectors of constant terms, we obtain the following expression:

$$\ln(\Delta A_i) = \underline{C} + e_R \ln(R\&D)_i + e_0 \ln(\Delta A_i) + e_1 \left[\underline{m_{1i}} \ln(\underline{\Delta A}) \right] + ... + e_K \left[\underline{m_{Ki}} \ln(\underline{\Delta A}) \right] \quad (8)$$

In BGP the innovation in region i depends on the $R\&D$ in region i and on the spillovers from innovation in all the other region within distance K with an elasticity that varies with distance.$\underline{m_{kj}}$ are row vectors, defined as in equation

(2) in the text. As expression (8) holds for $i = 1, 2...86$ we can stack all terms in matrix format[9] and solve for $\ln(\Delta A)$,obtaining:

$$\ln(\underline{\Delta A}) = (I - e_0 I - e_1 M_1 - ... - e_K M_K)^{-1} \left[\underline{C} + e_R \ln(\underline{R\&D})\right] \tag{9}$$

While inverting the matrix $(I - e_0 I - e_1 M_1 - ... - e_K M_K)$ would be extremely complicated, we can consider a linear approximation of the inverse (linear in the parameters $e_1, e_2...e_K$, that is) , which would be good as far as the externality parameters $e_1, e_2...e_K$ are not too far from 0. Using this linear approximation and exploiting the fact that Markovian matrices (such as $M_1, M_2, ...M_K$) when multiplied by a vector of constants return a vector of constants we obtain:

$$\ln(\underline{\Delta A}) = \underline{B} + \frac{e_R}{1-e_0}\ln(\underline{R\&D}) + \frac{e_R}{1-e_0}\frac{e_1}{1-e_0}M_1\ln(\underline{R\&D}) + ... + \frac{e_R}{1-e_0}\frac{e_K}{1-e_0}M_K\ln(\underline{R\&D}) \tag{10}$$

Equation (10), for $\varepsilon_0 = \frac{e_R}{1-e_0}$ and $\varepsilon_k = \frac{e_R}{1-e_0}\frac{e_k}{1-e_0}$ is the one reported in section 3.1 as equation (4).

C A Model of the Production side

We sketch here a minimal model which allows us to justify rigorously equation (3) in the text and also to support the idea that the variable Dem_i is meaningfully proxied by local population. The model we have in mind is a multi-region version of the one described in Jones [17]. To simplify further we assume no physical capital and also that the skilled labor is all used in *R&D* activity. Therefore the variable $(R\&D)_i$ is the employment in *R&D* in region i, and *R&D* workers move across regions freely so as to equate their returns (wages). The two equations summarizing this simple economy are:

$$Y_i = L_i^{\alpha} \int_0^{A_i} x_i(s)^{1-\alpha} ds \tag{11}$$

$$\Delta A_i = B(R\&D)_i^{e_R} A_i^{e_0} \prod_{j \neq i} A_j^{e(k)} \tag{12}$$

Equation (11) is the production function of the final, perfectly tradable consumption good (Y_i), in each region. This final good, which is also the numeraire,

[9] Matrix M_1 is obtained from stacking the row vectors $\underline{m_{11}}, \underline{m_{12}}...\underline{m_{186}}$ and similarly all the others.

is produced using all the locally invented non-tradable intermediates (s, the index of these goods, goes from 0 to A_i) in amount $x_i(s)$ and local generic labor L_i which is assumed to be exogenously distributed across regions (its distribution is given by the distribution of population). Each intermediate good, once invented, is produced using η units of the final good. Its unit cost of production is therefore ηr (where r is the interest rate). Equation (12) is the innovation function as in (1) in the text.

The marginal productivity of $R\&D_i$ in the innovative activity can be expressed as the product of the marginal contribution of one unit of $R\&D$ in generating ideas time the value of new ideas in region i. Calling P_{A_i} the value of a new idea in region i, we have the following condition, due to perfect mobility of $R\&D$ workers [10]:

$$w_R = P_{A_i} e_0 \left(\frac{gA_i}{R\&D_i} \right) \quad i = 1, 2...86 \tag{13}$$

The wage of $R\&D$ workers (w_R) is equated to the marginal productivity of these workers in each region. If patenting has different values across regions, because exogenous local conditions in the demand affect the profit obtained from a new good, then this would affect the incentives for $R\&D$ workers to move there. In fact, equation (13) states that there is a positive relation between the value of patenting in a region P_{A_i} and the number of R&D workers attracted there.

It is easy to determine P_{A_i} in BGP, the calculations are as in Romer [22]. The demand curve for a generic intermediate good in region i is:

$$x_i = \left(\frac{1-\alpha}{p_i} \right)^{\frac{1}{\alpha}}$$

where p_i is the price of that intermediate good. The profit ($\pi_i = p_i x_i - \eta r x_i$) of the representative firm in region i is maximized by the usual mark-up rule: $p_i = r\eta/(1-\alpha)$. Substituting it into the demand curve and calculating the optimal profit we have:

$$\pi_i = (1-\alpha)^{\frac{1}{\alpha}} \alpha \left(\frac{r\eta}{1-\alpha} \right)^{\frac{\alpha-1}{\alpha}} L_i \tag{14}$$

This equation says that the profit from a new good in region i depends on the size of the labor force in production of that region. A larger labor force implies that more of the new good is demanded to produce final good and therefore larger profits per firm are earned. The value of a patent is the present discounted value of profits generated by the new idea. Following Romer [22], in assuming an infinite life of the patent we have:

[10] Recall that in BGP $\Delta A = gA$

$$P_{A_i} = \frac{\pi_i}{r}. \tag{15}$$

Substituting equation (14) and (15) into equation (13), solving for $R\&D_i$ and taking logs of both sides we have:

$$\log(R\&D_i) = c + \log(A_i) + \log(L_i)$$

where c is an unimportant constant that combines all constant terms from the previous equations. This last equation is exactly as equation (3) in the text. It shows explicitly that the local exogenous labor force L_i, affecting the demand for locally generated new goods, partly determines the distribution of $R\&D_i$. In particular it determines the exogenous part of the distribution of $R\&D_i$, which is not affected, that is, by the simultaneity with the stock of locally generated knowledge. This model provides the theoretical reason for choosing local population as the instrument for $R\&D_i$.

D Data Appendix

D.1 Regions and distances

We use 86 regions in the analysis covering most of the Western European Countries, excluding Scandinavia. All countries in the EMU are included, except for Austria and Finland. UK is also included. 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 level of a 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 Kilometers.

Considering the border-to-border distance across regions, described in the text, and grouping them in class of distances we choose five intervals so that we can inquire 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

D.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 patent 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. The relatively large size of each region and the long interval of years should reduce the problems of differential "weights" of the innovative contribution of patents.

D.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 in constant 1985 ECUs 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 A. Ciccone and their sources can be found in Ciccone [5]. 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. We report the share of college educated in the population for each country in the following Table 1A.

Table 1A: College Graduates in the Labor Force

Country	*Percentage*
Belgium	9%
Denmark	13%
Germany	9%
Greece	4%
Spain	6%
France	14%
Ireland	6%
Italy	7%
The Netherlands	21%
Portugal	5%
United Kingdom	12%
Source: Ciccone (98) and National Statistical Censuses	
Years: around 1986	

D.4 Historical Population Data

Historical population of the 86 European regions comes from data of national censuses . In particular:

Belgium: "Population par Arrondissement Administratif-situation au 31 Decembre 1930" Institut National de Statistique

Denmark: Statistical Office.

France: "Recensements de 1891 aà 1962", INED Documentation.Data are from the 1931 Census and are derived from Table V.B "Population legale par departement circonscription d'action regionale. Recensement de 1891 à 1962"

Germany: " Statistiches Jahrbuch fuer das Deutche reich 1931: Laender und Landsteil and Laender und grossere Verwaltungsbezirche"." Also " Verwaltungsgrenzen in der Bundesrepublik deutchland selt Beginn des 19. Jahrhunderts (Veroeffentlichungen der Akademie fuer Raumforscung und Landesplanung-Forschungs-und Sitzungsberichte, Band 110)", Hannover 1977, ISBN 3-507-91408-5 and "Statistiches Reichsamt: Statisik des deutschen reiches, Band 451: Die Bevolkerung des deutschen reichs nach den Ergebnissen der Volkszahlung 1933, heft 1: stand, entwicklung und siedlungsweise der Bevolkerund des deutschen reichs", Berlin 1935.

Greece: "population de fait d'apres les recensements de 1839 à 1923 par departements", Table I Data are for the year 1928.

Ireland: Census of the Population of Ireland, Central Statistical Office. Dta are from the Census in 1936

Italy: "Cento Anni di statistiche sulle Regioni d'Italia"- SVIMEZ Associazione per lo sviluppo dell'industria nel Mezzogiorno. Dta are for the year 1931.

Netherlands: "Aantal inwoners van de provincien en Nederland (1830-1946)". Statistics Netherlands Data are fior the year 1930.

Portugal: Population in 1930 is derived from the "Censo da Populacao de Portugal", no Ide dezembro de 1930, Direccao Geral de Estatistica.

Spain: "Censos de Poblation", cuadro 2.20:"poblacion de espana por Comunidades Autonomas,1787-1981. Poblation de hecho". Data have been kindly provided by Antonio Ciccone and are for the year 1930.

UK: "Census of Population"- Table (A: Census Populations, density and intercensal Changes 1911-1931. England and wales, urban and Rural, Aggregates and Regions, Counties, County Boroughs and Metropolitan Boroughs.- National Statistics. Data are for the year 1931.

Particularly difficult has been the reconstruction of the population by region for Greece, Germany and UK. For all the three countries has been necessary to map the old definition of regions to the current definition, using data at counties level and aggregating[11].

[11] We are particularly grateful to the German Statistical Office for guiding our reconstruction and in particular to Dr. Thomas Helmcke .