KNOWLEDGE FLOWS AND INNOVATION*

Giovanni Peri †

University of California, Davis

October 10, 2002

Abstract

This paper estimates knowledge flows across regions in Europe and North America as revealed by patent citations. Only if these knowledge flows affect the productivity of researchers in generating new ideas do they generate externalities in innovation. We find that knowledge diffusion is strongly affected by immediate proximity, distance, national borders, linguistic and technological barriers. While some sectors such as Computer Technology are more "globalized" than others, all knowledge flows are much more far-reaching than trade or migration flows. However, once we control for R&D inputs, we do not find evidence that these flows of knowledge have an impact on innovation.

JEL Codes: F0, O3, R1

Key Words: Knowledge Flows, Innovation, Patent Citations, Regions.

*I thank Paul Beaudry, Francesco Caselli, Marianne Feldman, Roger Gordon, Bronwyn Hall, Gordon Hanson, John Helliwell, James Rauch, Manuel Trajtenberg and Michelle White , for very useful comments. I thank Paola Franceschi who provided very competent expertise in editing the paper. Shireen Al Azzawi provided excellent research assistance. I acknowledge the Institue of Governmental Affairs (IGA) for partially funding this project. Errors are mine.

†Address: Department of Economics, U.C. Davis, One Shields Avenue, Davis Ca 95616, USA. gperi@ucdavis.edu. phone: -1 530 752 3033, fax: -1 530 752 9382.

1 Introduction

The concept of knowledge externalities in innovation has pervasively populated the recent theoretical literature on economic growth. Innovation is the engine of growth, it has been argued, and externalities from existing knowledge are the "renewable" fuel that keeps this engine running. If knowledge of existing ideas makes scientists better in discovering new ideas then there are positive knowledge externalities in innovation. The strength of these externalities is a key determinant of the dynamics of innovation, and ultimately of productivity growth. Early models of endogenous growth (such as Aghion and Howitt [1] or Romer [19]) assumed very large knowledge externalities emphasizing the public-good nature of knowledge. As neither productivity growth rates nor innovation growth rates (as measured by patents) have increased in the last fifty years, while the resources devoted to research increased vastly, several authors (Jones [15], [16] analyzing productivity and Griliches [7], [8] analyzing innovation) have questioned the existence of positive and large externalities of knowledge. The historically decreasing productivity of R&D in generating output and innovation could be the consequence of the negative impact of existing knowledge on innovation. As the theoretical analysis does not impose any constraint on the sign and magnitude of the impact of available knowledge on innovation we need to learn about this crucial parameter from the empirical analysis.

The present paper aims at identifying and estimating knowledge externalities in generating innovation, using a large dataset of innovation patented in the United States. Patent data have long been used to inquire into the process of innovation. Recently, thanks to the effort of collection and analysis of Adam Jaffe, Manuel Trajtenberg and their coauthors, we have access to detailed data and information on US granted patents (see Jaffe and Trajtenberg [14] for studies and references). The present paper aggregates the information contained in about four million of those patents to estimate the flows of knowledge at the frontier of scientific innovation as revealed by patent citations. Aggregating patents at the regional level we are able to treat knowledge flows much like other macroeconomic flows (trade, migration, investments) and we analyze their determinants from an aggregate perspective.

As citations across patents mark the trail left by an existing idea involved in developing a new idea, we use them to reveal the intensity of knowledge flows across regions. With these data we can analyze to what extent distance, crossing borders, moving away from technological and innovative proximity and other hurdles affect knowledge flows at the frontier of the innovative activity. In particular, by measuring these flows, we can calculate what share of overall knowledge generated in one region is available to another, assuming a stable geographical distribution of creation and diffusion of knowledge over time.

The analysis and description of knowledge flows occupies the first half of our paper. Then we inquire into the existence of positive or negative effects (externalities) of these flows in the process of generating new ideas. To do this we estimate the elasticity of patent creation to the estimated stock of available knowledge in each region. In particular as we are able to identify, using citations, the relative flows of knowledge from each region to any other, we consider as knowledge available to a region the flow-weighted amount of ideas going into it.

Previous work analyzed technological externalities in production (as in Coe and Helpman [5] and Keller [17]) or knowledge diffusion in itself (Jaffe et al.[13], Jaffe and Trajtenberg [14]). In this paper we explicitly model and estimate the phase of knowledge diffusion first and we build on it to estimate the potential impact of knowledge on the ability of researchers to generate new ideas.

Our results are rather clear. Knowledge flows are, overall, more localized than anecdotal evidence, based on casual observation of the "information community", would suggest. However some specific sectors, such as "Innovation in Computers", are much more "globalized" than others. On one hand 80% of knowledge flows, as revealed by citations, stay within the boundaries of a region and, of the remaining knowledge flows only 50% pass the country border. On the other hand, knowledge flows are much farther reaching than those involving movement of goods (trade) or people (migrations). Existing estimates imply that passing from within to outside a country reduces trade flows by 95-96%. Knowledge flows, on the contrary, are reduced by not more than 30%.

Moving to estimates of knowledge externalities, knowledge flows do not reveal any positive and significant effect on productivity of R&D resources in generating new ideas. Our cross sectional evidence seems consistent with the time series evidence of small or no positive returns to existing knowledge in innovation. This may be the indication of decreasing returns in knowledge creation. The more ideas have been discovered and are known, the harder it is to produce a further relevant discovery with the same amount of research. Alternatively, it may capture the fact that as the knowledge and research "space" expands companies and government employ less talented researchers. Let us emphasize that here we are only measuring and capturing knowledge externalities in innovation. These externalities could have large effects on production but we are not capturing them in our study.

The rest of the paper is organized as follows. Section 2 presents our model of knowledge diffusion and knowledge creation. Section 3 presents the data, Section 4 estimates the intensity and extent of knowledge flows and Section 5 analyzes the impact of knowledge flows in generating new ideas. Section 6 concludes.

2 The Model

We use two tools in order to analyze knowledge flows and knowledge externalities across regions. The first is a function that models the diffusion of ideas across regions as depending on bilateral regional characteristics.

This is an adaptation to cross sectional data of a diffusion function such as the one used in Caballero and Jaffe [3] or in Jaffe and Trajtenberg [14] Chapter 7. The advantage of this approach is that it produces estimates of knowledge flows that could be compared with estimates of other flows, namely trade and migrations. The second tool used is an innovation function describing how ideas are generated by researchers using R&D spending and existing ideas available in a particular region. The description of these two equations and their estimation occupy the rest of the paper.

2.1 Diffusion of Ideas

We use here a simple framework to estimate diffusion of ideas across regions using cross-sectional data on Patent citations. We assume that the probability that one representative idea, generated in region s (sending) in year t has flown to region r (receiving) by year T is as follows:

$$\phi(r,T,s,t) = \varkappa e^{f(r,s)} e^{-\beta(t)_1(T-t)} \left[1 - e^{-\beta(t)_2(T-t)}\right]. \tag{1}$$

This function assumes that the diffusion of ideas over time between period t and T is regulated across all regions by a double exponential in which $\beta(t)_1$ determines the rate of obsolescence and $\beta(t)_2$the rate of diffusion over time of ideas. $\varkappa$ is a constant and the term $e^{f(r,s)}$ allows each region couple r, s to have a specific effect on the probability of diffusion of the average idea between them. Such effect depends on bilateral regional characteristics and scales up or down the common term capturing diffusion over time. As long as the diffusion and the obsolescence effects captured by $e^{-\beta(t)_1(T-t)} \left[1 - e^{-\beta(t)_2(T-t)}\right]$ are equal for all regions and separable from the region-couple effect captured by $e^{f(r,s)}$ it is possible to consider any couple of "sending" and "receiving" interval of years to estimate $f(r,s)$. Integrating the receiving period between T_1 and T_2 and the sending period between t_1 and t_2 we generate the cross sectional function $\phi(r,s) = \varkappa e^{f(r,s)}\Omega$. The term Ω is a common function of T_1, T_2, t_1 and t_2 and could be considered as a constant in the cross section as long as the time-intervals considered are common to all regions. In our empirical analysis we consider a group of ideas, discovered during the period t_1 =1975 to t_2 =1991 as "sending" flows of knowledge to be used in later discoveries by "receivers" during the period T_1 =1992 to T_2 =1996. We are interested in analyzing the determinants of knowledge flows between regions therefore we assume that the function $f(r,s)$ depends on several bilateral characteristics. In particular, merging all constant terms into δ, we use the following representation,

$$\phi(r,s) = \delta e^{\gamma_1 (d_1)_{r,s}} e^{\gamma_2 (d_2)_{r,s}} e^{\gamma_3 (d_3)_{r,s}} e^{f(\underline{Char}_{r,s})}. \tag{2}$$

We use exponential functions in order to interpret the coefficients as elasticities. $(d_1)_{r,s}$, measures distance between region r and region s in geographical space, $(d_2)_{r,s}$ measures distance in technological space and $(d_3)_{r,s}$ measures distance in innovative space[1]. The vector $\underline{Char}_{r,s}$ captures other bilateral characteristics such as being the same region, being in immediate proximity, being in the same country, being in the same trade block, using a common language and so on.

We denote with ΔA_r the ideas (new knowledge) generated in region r (for "receiving") during the later period of time (1992-1996) and with ΔA_s the ideas generated in region s (for "sending") during the earlier period of time (1975-1991). Also we call $\sigma(s,r)$ the number of ideas generated in region s that diffused to and are available in region r. Therefore the definition of $\phi(r,s)$ is

$$\phi(r,s) = \frac{\sigma(s,r)/\Delta A_s}{\sum_s \left(\sigma(s,r)/\Delta A_s\right)}. \tag{3}$$

The term in the numerator measures the share of ideas generated in region s that diffused and are available in region r, this term is closely related to the "citation frequency" between r and s. The term in the denominator is a standardization so that the sum of $\phi(r,s)$ over the sending regions is equal to one. Therefore we can interpret ϕ as the probability that an idea generated in s makes it to r. There is a strict relationship between the flow of ideas across regions $\sigma(s,r)$ and the flows of citations between the same regions, $c(r,s)$, and also between the number of new ideas ΔA_s and the number of patents generated in region s, P_s. However we think it is relevant to allow for two corrections before bringing (3) and (2) to the data[2].

First, we allow the average "number of ideas" in a patent, denoted as β_s, to be different across regions. Alternatively we can interpret this parameter as the "average quality " of ideas contained in the patents of a region. We call this parameter the "intensity of ideas per patent". Denoting with P_s the total number of "sending" patents granted to region s in 1975-1991 the above assumption implies: $\Delta A_s = \beta_s P_s$.

Second, taking into account two important features of citations data, we allow a looser relationship between citations across regions $c(r,s)$ and number of ideas diffused $\sigma(s,r)$. In particular we assume the following relationship: $c(r,s) = \psi_r \sigma(s,r) \varepsilon(r,s)$. $c(r,s)$ is the total number of citations from patents of region r to patents of region s. As different "receiving" regions may have different average propensity to include citations in their patents, we allow for a region-specific term ψ_r. $\sigma(s,r)$ is the actual flow of ideas between regions. $\varepsilon(r,s)$ is a log-normal random noise due to the fact that not all citations capture the flow of ideas as some are added by patent reviewers. We assume that such noise is randomly distributed across region couples (i.e. not correlated with $\sigma(s,r)$) as reviewers include them independently of any geographic characteristic

[1] The exact measure for these three types of distances will be made clear in the empirical section.

[2] In section 5.3 we also use citations to capture directly $\sigma(r,s)$ and citation-weighted patents to capture ΔA_s as a form of robustness check.

of the citing patent. This assumption allows us to use the information in the citations to infer flows, while trying to purge it from region-specific characteristics and from random noise.

Substituting these definitions into expression (3) and equating it to expression (2) we get the following relation,

$$\phi(r,s)=\frac{c(r,s)/\psi_r\beta_s P_s\varepsilon(r,s)}{\sum_s(\sigma(s,r)/\Delta A_s)}=\frac{c(r,s)/\varepsilon(r,s)}{\mu_r P_s\beta_s}=\delta e^{\gamma_1(d_1)_{r,s}}e^{\gamma_2(d_2)_{r,s}}e^{\gamma_3(d_3)_{r,s}}e^{\underline{f}(\underline{Char}_{r,s})}, \tag{4}$$

where we have combined all the receiving-region specific effects into one: $\mu_r=\psi_r\sum_s(\sigma(s,r)/\Delta A_s)$. Isolating the observable terms of $\phi(r,s)$ on the left hand side and taking logs on both sides of (4) we obtain the following estimable equation, which serves as the empirical specification to estimate knowledge flows,

$$\ln\left(\frac{c(r,s)}{P_s}\right)=a+\ln(\mu_r)+\ln(\beta_s)+\gamma_1(d_1)_{r,s}+\gamma_2(d_2)_{r,s}+\gamma_3(d_3)_{r,s}+\underline{f}(\underline{Char}_{r,s})+u(r,s). \tag{5}$$

Equation (5) is derived from the framework described above, can be brought to the data and has an interesting interpretation "per se". It says that the citation frequency $\frac{c(r,s)}{P_s}$ from region r to region s depends on several characteristics. First it depends on a "citing region" fixed effect, $\ln(\mu_r)$, which controls for different propensity to cite and other fixed effects and has no interesting interpretation. Second, it depends on a "cited region" fixed effect, $\ln(\beta_s)$, which captures the average intensity of ideas per patents in the "sending" region. Third it depends on a set of bilateral distances and bilateral characteristics (described above). Finally it contains $u(r,s)=\ln\varepsilon(r,s)$ that is a normally distributed random error. The equation has the flavor of a gravity equation, popular in trade and migration analysis. The "flows" of a variable, in this case of knowledge, between two regions depend on some characteristics of the two regions and on the "distance" between them. Specification (5) allows us to estimate the parameters γ_1, γ_2 γ_3 and the vector of coefficients $\underline{f}$. Using definition (2) we can then apply this estimate to have a measure of $\phi(r,s)$, the intensity of knowledge flows.

2.2 Innovation Function and the Effect of Available Knowledge

Once we have estimated the strength of knowledge flows, we can estimate the impact of these flows on the production of new ideas. If the process of generating ideas in the analyzed regions has fluctuated, on average, around a balanced growth path, we can separate the contribution to innovation of private inputs such as scientists and R&D resources from the contribution (externality) of available ideas in the estimation of the regional innovation function.

New ideas, ΔA, are generated by scientists and researchers who exhibit different levels of productivity across regions. Employees in R&D activities in region r are indicated as $L_r^{R\&D}$.Their productivity depends on two main factors: the amount of resources available to them (such as laboratories, research funds, salaries and so on) captured by spending per-worker in R&D and denoted as h_r and on the stock of ideas available to them in region r. We denote with the symbol A_r^{AVA} and ΔA_r^{AVA} the cumulated stock and the increase in the stock of ideas available in region r. A_r and ΔA_r still denote the stock and the increase of the stock of ideas generated in region r. The knowledge available in region r is given by the knowledge generated anywhere and diffused to region r . Using the definition of knowledge flows $\phi(r,s)$ developed in the previous section, available knowledge for region r is: $A_r^{AVA} = \sum\limits_{s\in S} \phi(r,s)A_s$. The production function of innovation is, therefore,

$$\Delta A_r = f(L_r^{R\&D}, h_r, A_r^{AVA}) = f\left(L_r^{R\&D}, h_r, \sum_{s\in S}\phi(r,s)A_s\right). \tag{6}$$

As we assume that regions are on their balance growth path with a common growth rate of knowledge stocks, then flows are proportional to stocks and: $A_s = g\Delta A_s$. We devote Section 5.1 to check that our data support this assumption. Substituting this condition into equation (6) and using a log-linear expression for the innovation function we have

$$\ln(\Delta A_r) = -\ln(g) + \varepsilon_L \ln(L_r^{R\&D}) + \varepsilon_h \ln(h_r) + \varepsilon_I \ln\left(\sum_{s\in S}\phi(r,s)\Delta A_s\right). \tag{7}$$

ε_L is the elasticity of innovation to the supply of scientists. ε_h is the elasticity of innovation to the supply of resources per scientist. ε_I is the elasticity of innovation to available ideas. We can construct the available knowledge in each region using the estimates of $\phi(r,s)$ obtained from the previous section (2.1). Also as ΔA_r is equal to $\beta_s P_s$,denoting with a hat$\widehat{}$ the variable estimates from equation (5) we have the estimable specification

$$\begin{aligned}\ln(P_r) \quad = \quad & Intercept - \ln(\widehat{\beta}_r) + \varepsilon_L \ln(L_r^{R\&D}) + \varepsilon_h \ln(h_r) + \\ & + \varepsilon_I \ln\left(\sum_{s\in S}\exp\left(\widehat{\gamma_1}(d_1)_{r,s} + \widehat{\gamma_1}(d_2)_{r,s} + \widehat{\gamma_3}(d_3)_{r,s} + \widehat{f}\ (\underline{Char}_{r,s})\right)\widehat{\beta}_s P_s\right) + \nu_{r,s}.\end{aligned} \tag{8}$$

The dependent variable is the count of receiving patents granted in region r. The term $Intercept$ contains all the uninteresting constants. The term $\ln(\widehat{\beta}_r)$ controls for the average intensity of ideas per patent in region r; it uses the estimates obtained from equation (5) with the assumption that such average intensity remains constant in a region over time. $\ln(L_r^{R\&D})$ is the log of workers employed in the R&D sector, $\ln(h_r)$ is the log of spending in US \$ per worker in the R&D sector. The term in brackets is the estimated knowledge available in region r coming from ideas discovered in the past. It is calculated as the intensity of flows to region r (the exponential term) times the amount of ideas generated in each region $\widehat{\beta}_s P_s$. Finally $\nu_{r,s}$ is a zero average random disturbance that captures other determinants of patenting not included in our equation.

3 The Data

Our data on patented inventions, on their inventors and on citations between them are taken from the NBER Patent Citation data File described in Jaffe et al. [11]. We choose only patents granted between 1975 and 1996, for which citation data are available, and we organize them across 141 regions[3]: 51 US states including D.C., 10 Canadian Provinces and 80 regions in 17 European Countries. The regions chosen for Europe are the Territorial Units used by Eurostat and denoted with the name NUTS1. Patents are assigned to the region of residence of their first inventor in order to locate the invention as close as possible to the place where it was developed (this procedure is common to several other studies such as Jaffe et al. [13]). The original NBER file identifies the state of residence of American inventors but only the city and the zip code of foreign inventors. We mapped cities and zip codes into regions for European and Canadian inventors with the help of national Maps and Gazetteers. As the location of the first inventor is highly correlated with the location of the other inventors this method gives an accurate representation of the distribution of innovation in Europe and North America.

In our estimate of knowledge flows we want to be particularly careful. We want to measure knowledge diffused across regions and available to be used by others and not track the diffusion of knowledge within a company. While the first flow may generate externalities the second simply mirrors the existence of large multinational companies that probably compensate their departments in different regions for providing and diffusing knowledge. Therefore we do not include in the count of citations $c(r, s)$ the "self-citations", i.e. those citations done between two patents assigned to the same institution (University, Company, Research-Lab, and so on)[4].

Other studies have used patent to patent citation to track communication of knowledge across inventors (see Jaffe and Trajtenberg [14] Chapters 5, 6 and 7) and recent surveys have shown that "patent citations do provide indication of communication, albeit one that also carries a fair amount of noise" (Jaffe and Trajtenberg [14], Chapter 12 pp. 380). The use of citations in our study to extract information on knowledge flows is probably less affected by the "noise" as we aggregate over large units (regions). Moreover we explicitly allow and model a random error when using citations as proxies for knowledge flows (in section 2.1).

[3]The list of regions and countries they belong to is described in Appendix A

[4]If we estimate the main specification in Table 2 column I, including self citations the coefficient that is mostly affected, increasing in absolute value by 30%, is the one on the Out of Region dummy. The other coefficients change very little.

The data on employment in R&D and spending per employee in R&D are the averages for the 1992-1996 period. The first is measured as number of employees working in the R&D sector and the second in spending per employee in constant 1992 U.S. dollars. The averaging over four years allows us to fill some existing gaps in the regional yearly data and provides a value less affected by year to year fluctuations. In particular, by doing so, we are more confident to approximate the relative balanced growth path levels of these variables. For European regions the data are from the REGIO dataset, for Switzerland data were provided by the Swiss statistical office, for the U.S. data are from the National Science Foundation and for Canada they are from Statcan. A detailed description of the Data and their sources can be found in the Data Appendix.

4 Flows of Ideas

4.1 Diffusion

In this section we present the estimates of equation (5). We divide patents into citing patents (granted in the 1992-1996 period, a total of 919,743 patents) and cited patents (granted in the 1975-1991, a total of 3,046'869 patents). The total count of potential citing patents for the citing region r is denoted as $(P_{9296})_r$ and the total count of potential cited patents for the cited region s is denoted as $(P_{7591})_s$. In order to include in the regression the information from regional couples without any citation link we add 0.01 citations to all the region-couples[5]. In the basic specification we include several controls. We include citing and cited regions specific effects, we include five dummies, one for crossing a regional border between region r and region s, one for crossing two regional borders, one for crossing a national border, one for crossing a Common Trade Area border and one for crossing a linguistic border. For instance the dummy "Crossing Region Border" is equal to zero if the sending and receiving regions are the same ($r = s$) and equal to one otherwise, the dummy "Crossing the next region border" is zero if r and s are either the same region or they share a border and it is one otherwise, and so on. We also include measures of three types of distance. $(d_1)_{r,s}$ is simply the geographical distance between two regions. It is measured in thousands of kilometers and calculated as the shortest arc distance on the earth surface between the two regions, choosing as location of each region the location of its largest metropolitan area. Technological distance $(d_2)_{r,s}$ is a measure of the difference in technological specialization of two regions. This measure is based on an index of technological proximity first introduced by Jaffe [12]. This index is calculated as follows: first we divide the patents granted in thirty-six technological groups (defined in Jaffe and Trajtenberg [14] pp. 452-454). For each region s we construct the vector of shares of patents in each group $\underline{sh}_s = (sh_{s1}, ...sh_{s36})$. Then for each citing-cited region couple (r, s)

[5] We also redo estimates of specifications in Table 2 and 3 using only those observations with strictly positive $c(r, s)$. Results are similar to those obtained using all observations and are available upon request.

we construct the un-centered correlation coefficient between vector $\underline{sh}_r$ and $\underline{sh}_s$ as

$$\rho_{r,s} = \frac{\sum_i (sh_{si} * sh_{ri})}{\sqrt{\sum_i (sh_{si})^2 \sum_i (sh_{ri})^2}}. \tag{9}$$

The un-centered correlation is also the angular distance between the vectors: two regions with identical specialization have a correlation of one, two with orthogonal specialization have a correlation of zero. The technological distance $(d_2)_{r,s}$ is equal to $1 - \rho_{r,s}$ and is bounded between 0 and 1. Finally the distance in innovative intensity $(d_3)_{r,s}$ measures how far two regions are in technological advancement, rather than technological specialization. It is the difference in absolute value of patents per worker granted in the two regions in the 1975-1991 period. A region at the frontier of technological innovation would have a large value of per capita patents while regions that are technologically lagging behind have low values of the same variable. This distance may affect the intensity of spillovers as a region may be more effective in using knowledge flowing from another region with similar level of technological advancement rather than from a region that is much more (or much less) technologically developed[6].

Table 1
Descriptive statistics

Variable	Mean	Std. Deviation	Min	Max
Number of region to region citations without self, c(r,s)	75.2	516.7	0	68778
Geographical distance[a] (d_1)	4.44	3.22	0	13.70
Technological distance (d_2)	0.34	0.19	0	1
Innovative advancement distance[b] (d_3)	1.38	1.52	0	8.48
Number of Patents in cited region, 1975-1991, (P_{7591})	6523.2	12924	1	96804
Number of Patents in citing region, 1992-1996, (P_{9296})	2544	5186	2	44744

Notes: Citation frequencies are calculated omitting self-citations, i.e. citations between patents whose first authors belong to the same company-institution.
a: Thousands of Kilometers.
b: Difference in patents per worker (absolute value).

Table 1 reports some summary statistics for the number of citations, the number of patented innovations and for the three measures of distance. Geographical distance range from 0 kilometers, when citing and cited region are the same one, to almost fourteen thousand Kilometers (between the US Hawaiian Islands and the Greek Islands of Nisia Aigouu-Kriti). Most of the distances, though, are below 10,000 kilometers (only 2% of distances are above). Technological distances range from 0 (same region) to 1, and innovative advancement

[6]Similar results are obtained if we include only "positive" innovative distances and zeros for negative distances, assuming that regions can only receive flows from more advanced ones.

distances range from 0 to 8.48. The average number of region to region citations without self-citations is 75.2 but the variance is very large and the distribution very skewed with many couples with few citations and few couples with a very large number of citations. Our regression results, though, are rather robust to the exclusion of outliers.

Table 2
Determinants of Knowledge Flows
(Aggregating All Sectors)

Specification	**I (All sectors) Citing:92-96 Cited: 75-91**	**II (All sectors) Citing:92-96 Cited: 75-84**	**III (All sectors) Citing:92-96 Cited: 85-91**
Crossing Region Border	-1.56*	-1.54*	-1.99*
	(0.15)	(0.23)	(0.22)
Crossing next-Region Border	-0.50*	-0.80*	-0.65*
	(0.05)	(0.10)	(0.09)
Crossing Country Border	-0.30*	-0.40*	-0.44*
	(0.05)	(0.08)	(0.07)
Crossing Trade-Block Border	0.06	0.01	0.08
	(0.04)	(0.06)	(0.06)
Crossing Linguistic Border	-0.31*	-0.19*	-0.42*
	(0.04)	(0.07)	(0.06)
1000 Km farther (γ_1)	-0.036*	-0.045*	-0.06*
	(0.005)	(0.009)	(0.009)
1 std deviation of technological	-0.66*	-0.82*	-0.81*
distance farther (γ_2)	(0.03)	(0.04)	(0.04)
1 std deviation in innovation	-0.19*	-0.19*	-0.17*
activity farther (γ_3)	(0.02)	(0.03)	(0.02)
Citing Region Fixed Effects	Yes	Yes	Yes
Citied Region Fixed Effects	Yes	Yes	Yes
Observations	19881	19740	19740
R^2	0.21	0.22	0.26

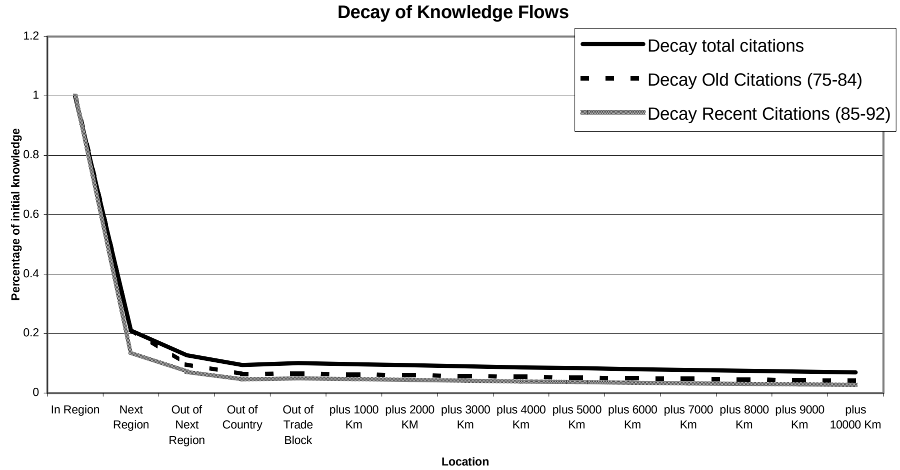
Notes: Citation frequencies are calculated omitting self-citations. To all region-couples is added 0.01 citations to avoid zeroes. Heteroskedasticity Robust Std errors in parenthesis. *= significant at 1% confidence level.

Table 2 reports in column I the estimates of the basic specification using patents in the period 1975-1991 as the "sending" cohort and in the period 1992-1996 as the "receiving" cohort. We pool all sectors and obtain 19,881 regional couples. Using the reported coefficients we can calculate the percentage decrease of the dependent variable as the independent variables change. For instance passing from within the region to outside it (first row), the knowledge flow drops to ($e^{-1.56}$ =) 21% of its initial value. Crossing another regional border by moving out of the bordering region causes a further drop to ($e^{-0.50}$ =) 60% of the previous value, which corresponds to about 12% (=0.6*0.21) of the initial value. Crossing the national border causes a further drop to 9.4% of the initial value. As we proceed by row moving down Table 2 we see that exiting the trade block has no significant effect (if anything it is positive), while moving out of a linguistic border decreases knowledge flows by 30% of their value. Also controlling for these bilateral characteristics geographical distance still has an effect on knowledge flows. For each thousand Km of distance, knowledge flows drop an additional 4%. Each effect is incremental, therefore at a distance of 6000 Km (this is roughly the distance between say New York and Paris) after moving out of the region, next-region, country and trade-block, the flow of knowledge between two regions is reduced to about 8% of its initial intensity. Technological differences as well as differences in the innovative advancement have rather strong effect on the flow of knowledge too. An increase of the first distance by a standard deviation decreases flows to ($e^{-0.66} =$) 51% of their previous level, while an increase of one standard deviation of the second distance decreases them to ($e^{-0.19} =$) 82% of their previous level.

In order to check our assumption that relative knowledge flows across regions are not dramatically affected by the chosen period and therefore stable over time, we split the time interval of cited patents (1975-91) in two sub-periods (75-84 and 85-91) and we separately estimate the flows as revealed by citations from the 92-96 cohort to each of the two cohorts (column II and III in table 2). In spite of the time difference between the two cohorts the estimated coefficients are not dramatically different. Crossing the regional border and a linguistic border has less of an effect for older citations than for more recent. The passing of time could make knowledge relatively more available outside the region where it has been generated and outside the linguistic area. For other variables there is not a significant difference.

It is useful to represent in a diagram the decay of knowledge flows as we move out of the region where knowledge was generated. Figure 1 represents the decay of these flows starting from the initial amount of knowledge (100%=1) represented in the left extreme of the diagram. Using the estimates in Table 2 we calculate the percentage of knowledge flows as the region border is crossed (first drop), the next-region border is crossed (second drop), the country and the trade-block border are crossed (third and fourth drop) and intervals of 1000 Kilometers at a time are travelled. The figure, reporting the decay measured using estimates in specification I, II and III of Table 2, provides a graphical representation of the effects of geographical borders and distance on knowledge flows. We did not represent graphically the effect of linguistic borders and of technological and innovative distance.

Figure 1



The pattern obtained by using the whole 75-91 cited cohort (solid line) or each sub-cohort (dashed line for 75-84 and shaded line for 85-91) is rather similar. Looking at Figure 1 what stands out is the substantial drop in knowledge flows associated with crossing the regional border. Also significant are the effects of moving out of the next region and out of the country. The effect of pure distance, past the country border, is rather small. It appears that a large part of knowledge is highly localized, but is this degree of localization realistic? Let's establish this point through some further analysis and comparisons.

4.2 Comparisons

In order to have a better understanding of the sensitivity of knowledge flows to geographical barriers it is useful to compare the estimates obtained above, relative to overall knowledge, with two other types of estimates. First we consider how far reaching are knowledge flows in some particular sectors of innovation. Our idea of the "global scientific community" is probably affected by casual observation and anecdotal evidence relative to the computer and information technology sector. While knowledge in those sectors may be less sensitive to geographical barriers, it may still be highly localized for other sectors. We analyze, therefore, knowledge diffusion in the Computer sector as well as in two more "traditional" sectors such as Motors and Agriculture-Food[7]. Then we also compare knowledge flows to other types of flows that involve movements of goods or people (trade and migrations). Compared to these, more embodied, flows we expect knowledge to be less sensitive to geographical barriers and to reach farther.

[7]The patent classes in each of the categories used is described in Jaffe and Trajtenberg [14] pp. 452-454.

4.2.1 Computers, Motors and Agriculture-Food

The estimates of the previous section, showing a strong degree of localization of knowledge flows, seem in stark contrast with anecdotal evidence and casual observation suggesting perfect diffusion, via E-mail, web, phone and fax of any piece of knowledge generated in the advanced economies. Information technology should make knowledge flows perfectly portable in space and the community of innovators should be, especially in the nineties (to which the "receiving" cohort belongs), a global one. It is possible that part of the community of innovators is genuinely global, but this may not be the majority of them so that overall knowledge still has an important local component.

In order to inquire into this issue we consider three specific sectors: computers, motors and agriculture-food. As the information technology promotes globalization we think that researchers who study computers are most familiar with information technology. Moreover, thanks to a high degree of international standardization in computer language and computer procedures, knowledge in this sector should flow more freely across countries and space. Motors and Agriculture-food, on the other hand, are technologies more linked to the specificities of regional economies and probably less engaged into IT-promoted "globalization". Therefore we expect their knowledge flows to be less far reaching. All three sectors, though, are relevant innovators, with 13,235 patents granted to "Computer hardware and software" during the period 1992-1996, 11,026 patents granted to the "Motor and Engine" sector and 8,843 to the "Agriculture-Food" sector. Considering only patents and citations within each of these sectors, we estimate, for each of them separately, a specification like 5. Table 3 reports the estimated coefficients.

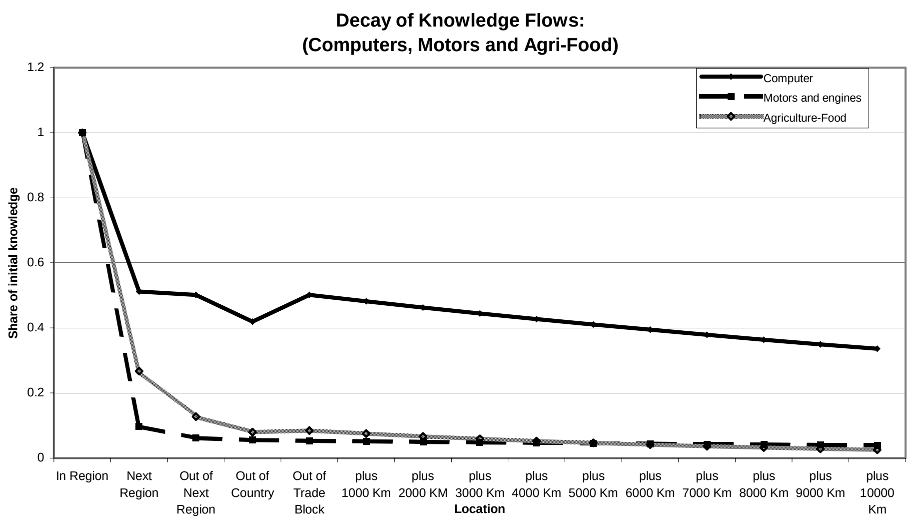
Table 3
Determinants of Knowledge Flows
(Computers, Motors and Agri-Food)

Specification	**I (Computers) Total citations**	**I (Motors) Total citations**	**III (Agri-Food) Total citations**
Crossing Region Border	-0.67*	-2.33*	-1.32*
	(0.27)	(0.29)	(0.30)
Crossing the "next" Region Border	-0.02	-0.46*	-0.74*
	(0.11)	(0.12)	(0.13)
Crossing the Country Border	-0.18*	-0.11	-0.47*
	(0.08)	(0.08)	(0.09)
Crossing the Trade Block Border	0.06	-0.04	0.06
	(0.06)	(0.06)	(0.06)
1000 Km farther	-0.04*	-0.03*	-0.12*
	(0.01)	(0.009)	(0.01)
Crossing Linguistic Border	-0.21*	-0.15*	-0.06
	(0.06)	(0.06)	(0.06)
1 std deviation in innovation activity farther (Patent/Worker)	-0.69*	-0.64*	-0.51*
	(0.04)	(0.04)	(0.04)
Citing Region Fixed Effects	Yes	Yes	Yes
Citied Region Fixed Effects	Yes	Yes	Yes
Observations	14280	17689	17955
R^2	0.15	0.15	0.17

Notes: Citation frequencies are calculated omitting self-citations. To all region-couples is added 0.01 citations to avoid zeroes. Heteroskedasticity Robust Std errors in parenthesis. *= significant at 1% confidence level.

Consider the first two coefficients that capture the effect of moving out of the region and out of its neighbors. We notice the strikingly different behavior of knowledge in the computer sectors from the other two more traditional sectors. For the motor and agriculture-food sectors the sum of these two effects is larger than for the general case, leaving only 6% and 12% (respectively) of the initial knowledge out of the neighbor regions. For the computer sector fully 50% of the initial knowledge flows all the way outside of the trade block. Further travel in distance seems to harm especially the knowledge flow in the Agriculture-Food sector, that drops by 50% out of the country and decreases by 12% each thousand kilometers travelled (climatic reasons may play a role). Interestingly, for the Computer sector the linguistic border seems to be more harmful than for the two more traditional sectors. This could be due to the fact that the computer community is global, but it prevalently speaks English. The community of motor-engine inventors, and agricultural-food inventors, on the other hand, seems much more local and regionally/nationally based. Representing in Figure 2 the decay of knowledge flows as geographical proximity decreases, the extremely different pattern of Computer vs. Motor and Agriculture-Food sectors clearly stands out.

Figure 2



The overall knowledge flows are more similar to those measured in the "traditional" sectors (Motors and Agriculture-Food) than to those measured in the computer sector. Let us point out, interestingly, that the high geographical localization of knowledge flows does not correspond to higher localization of innovation. Computer innovation is much more concentrated than innovation in Motors or in Agriculture. Between 1992 and 1996 four regions[8] accounted for 55% of the innovation in the Computer sector and California,

[8] California, Texas, Massachussets and New York

the top innovator, accounted for full 25% of the patenting in Computer. In the same period we need to add innovation from the top ten regions to account for 55% of innovation in motors and we need the top 16 regions to make 55% of innovation in Agriculture and food. Moreover, while all the top innovators in Computer are US states, four European regions enter the top ten list for Motors and two the top ten list for Agriculture and Food. Motors and Agriculture-Food, in spite of having more localized knowledge flows exhibit much lower concentration in innovative activity than the Computer sector does. Were innovative agglomerations driven mainly by positive knowledge spillovers the opposite would have happened.

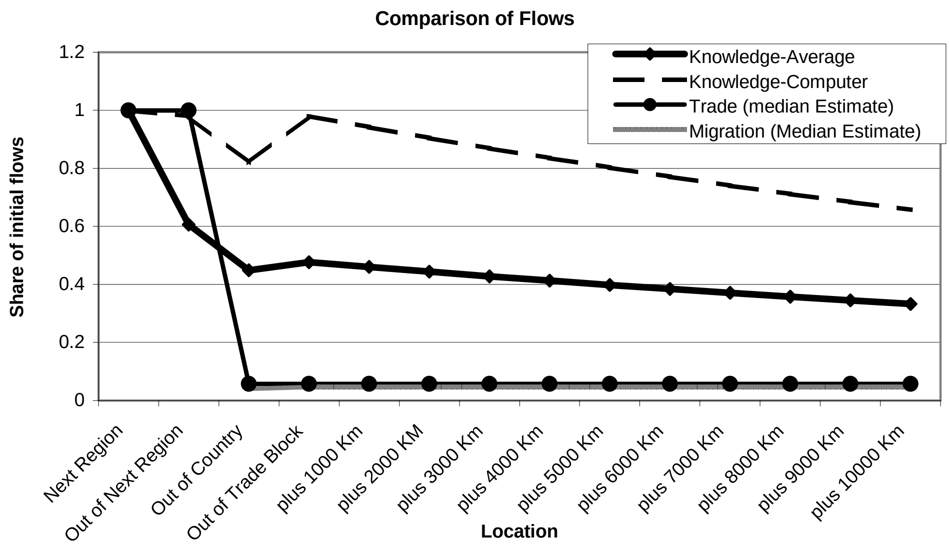
4.2.2 Trade and Migration

While overall knowledge flows are less far reaching than those in the highly globalized computer sector, how do these flows compare to some "material" flows such as trade and migration? There are, by now, several estimates of the effect of crossing a country border (U.S. to Canada) on trade flows and at least a few estimates of the effect of crossing a country border on migration flows. We concentrate on country-border effect as a useful benchmark: a region-border effect has not been estimated (to our knowledge) for trade and migration due to the unavailability of data for within-region trade and migration. In particular, it is useful to compare the country-border effect on trade and migration with the estimated border-effects for knowledge flows. In order to have a measure of knowledge flows comparable to flows of trade and migrants we omit within region flows (i.e. citations to the same region patents) and consider only citations made to patents generated in different regions.

For trade flows we use the estimates of "home bias" in trade from Helliwell [10] Chapter 1, that range between 2.5 and 3.2, including in the range the original estimates of 3.1 by McCallum [18]. These estimates imply a decline of 92-96% (leaving 8-4% of the previous level) of trade flows as the national border is crossed. For migration flows, Helliwell [10] Chapter 5 reports a much wider range of estimated home bias, with coefficients between 2 and 4.5, implying again a drop in migration flows at the border between 87 and 99%. Taking the median estimates for the above coefficients we report the "border effects" as drops in flows of trade and migration in Figure 3. On the same figure we report also the estimated decay in overall knowledge flows and in knowledge flows of the Computer sector. In Figure 3 we use the same estimates used to generate Figure 1 and 2, but we standardize to one (initial level) the amount of knowledge available just outside the region of origin (rather than the knowledge within it). The differences between the decay of trade-migration flows and knowledge flows are striking. The drop in trade (or migration) when crossing a country-border alone is 94-95%. Such effect is much larger than the cumulated effect on knowledge flows from crossing the next-region's , the country's, the trade-block's borders and travelling several thousands kilometers. For instance, more than 50% of the out-of-region knowledge flows reach beyond the trade-block border and that percentage is about 98% for computer-related knowledge flows.

Similarly, we can estimate the effect of distance on knowledge flows using as independent variable $\ln(d_1)_{r,s}$ to be able to compare it to the effect of distance on trade that is always estimated in logs. The estimate (not reported in the tables) of the coefficient on $\ln(d_1)_{r,s}$ is -0.09, fifteen times smaller than the effect of $\ln(distance)$ on trade, estimated around -1.4. Put in such perspective knowledge flows are much more "global" than trade and migration.

Figure 3



Flows of goods and people are much more affected by geographic barriers than knowledge flows are. This is particularly true if we consider that the estimates of the border effect for trade and migration were obtained using the Canada-US border (probably one of the most "permeable" among national borders) while the effect of knowledge flows were estimated for all Western Europe-North American countries. Such comparison reassures us about the plausibility of our estimates: knowledge has an important local component, but travels much farther than goods or people, moreover knowledge used in the information technology sector (computer innovation) is highly "globalized".

4.3 "Innovators" and "Receivers"

We might have been excessively "generous" in considering each citation as the sign of the flow of an idea. Relevant ideas may be generated only by few institutions at the frontier of technological development, while other regions are receivers of these ideas and merely apply small variations or adjustments to them. In this section we assume that the regions with largest innovative capacity in each country (i.e. most patented innovations in 1975-1991) are the large international sources of ideas, while other regions are mere receivers. Citations of ideas from these large innovators are considered as flows of truly relevant ideas, citations from other regions are ”minor” references and are neglected. We consider, therefore, citations from all 141 regions to the twenty regions that are top innovators in each of the 19 countries considered (for the US we choose the two largest, California and New York, that are also the largest innovators overall in our dataset). If the ideas flowing out of these large innovators are more relevant for future use they may also be more far reaching. Knowledge out of these twenty regions (denoted with an asterisk in the ”list of Regions” in the Appendix) may depict a different image of diffusion around the world. The estimates obtained from this exercise and presented in Table 4 show that it is not the case. Diffusion of knowledge as estimated using flows from selected ”innovators” to ”receivers” looks extremely similar to diffusion estimated using all regions.

Table 4
Determinants of Knowledge Flows
from large Innovators to other regions
(All Sectors)

Specification	**(All sectors) using citations to top-regions only**
Crossing Region Border	-1.82*
	(0.56)
Crossing next-Region Border	-0.44*
	(0.19)
Crossing Country Border	-0.26*
	(0.13)
Crossing Trade-Block Border	-0.08
	(0.09)
Crossing Linguistic Border	-0.22*
	(0.10)
1000 Km farther (γ_1)	-0.01
	(0.01)
1 std deviation of technological distance farther (γ_2)	-0.45* (0.08)
1 std deviation in innovation activity farther (γ_3)	-0.07 (0.04)
Citing Region Fixed Effects	Yes
Citied Region Fixed Effects	Yes
Observations	2820
R^2	0.74

Notes: Citation frequencies are calculated omitting self-citations. To all region-couples is added 0.01 citations to avoid zeroes. Heteroskedasticity Robust Std errors in parenthesis. *= significant at 1% confidence level.

The coefficients in Table 4 are quite similar to those estimated in column I, Table 2. The standard errors are larger in Table 4, as the number of observations has been reduced to one seventh of the initial sample. Moving out of the region is now estimated to cause knowledge to drop to 16% of its initial level (the previous estimate was 22%). Moving out of the next region causes a further drop to 64% (previous estimate was 60%) of the previous level reducing knowledge to 10% of the initial level. Finally, once out of the country knowledge is around 8% of the initial level. In the previous estimate it was 9.4%. Also the linguistic border effect is similar to what was estimated in Table 2 (0.22 vs. 0.31). Only the effect of geographical distance and of innovative distance is strongly reduced from the previous estimates. For innovative distance this is probably due to the substantial reduction in the variability of that measure. As we are only considering intense innovators as sources of citations their innovative distance from the other regions is very close to being simply a citing region fixed effect . Technological distance has still a relevant impact on flows.

Looking in general to the results, the estimates obtained using only few regions as innovators do not differ much from those that use all regions. Knowledge generated by the large innovators and revealed by citations does not exhibit a degree of localization different from that of knowledge generated by other regional sources.

5 Innovation Function

5.1 Regions' Balanced Growth Path

Estimation of Equation (8) is the second task of the paper. That relation relies on the assumption that the process of knowledge creation was around its Balanced Growth Path (BGP) during the considered 1975-1996 period. In this section we test two conditions that ensure Equation (8) to be well specified. They are:

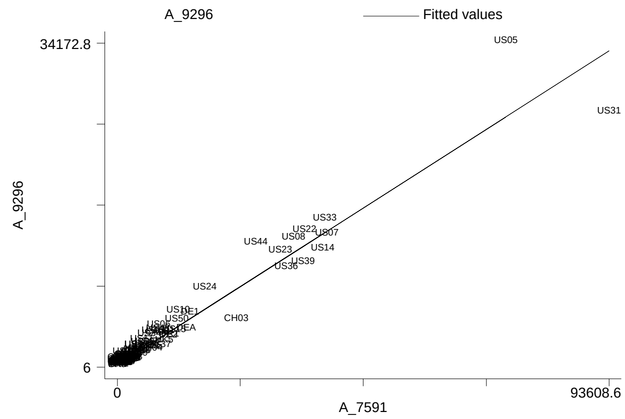
1. $(Pat_{9296} * \widehat{\beta})_r = \gamma(Pat_{7591} * \widehat{\beta})_r + e_r$
2. The deviations from the BGP relation written above, e_r, have zero average, and are not correlated to $L_r^{R\&D}$ or h_r.

The first condition says that new ideas generated in each region r during the 92-96 period must be proportional, in BGP, to the ideas generated in the 75-91 period[9] up to a random disturbance e_r. This is proven in the appendix.

The second condition ensures that the deviations from BGP, e_r, are zero mean and not correlated with other determinants of innovation so that the errors u_{rs} in equation (8) are uncorrelated with the explanatory variables.

[9] Recall that A_{r96}-$A_{r92} = (Pat_{9296} * \beta)_r$ and A_{r91}-$A_{r75} = (Pat_{7591} * \beta)_r$.

Figure 4
Proportionality of new ideas in Regional BGP:
(92-96) period versus (75-91) period



Before performing a formal test of condition 1 it is useful to look at Figure 4 . The BGP relation implies that there is a linear relation with zero intercept between the amount of new ideas generated in region r during the period 75-91 ($A\_7591 = (Pat_{7591} * \widehat{\beta})_r$) and those generated during the period 92-96 ($A\_9296 = (Pat_{9296} * \widehat{\beta})_r$). Figure 4 plots these two variables against each other. The visual impression that we get just from looking at the picture is that there exists a very tight linear relation between the two variables and that the intercept of the regression line is just about zero. Table 5, confirms that a linear relationship between the two variables explain 95% of the variance of $(Pat_{9296} * \widehat{\beta})_r$ and confirms that the intercept of the linear relation is not statistically different from 0 (column I, with all observations and Column III without two outliers, identifiable in Figure 4 as California and New Jersey). Also, no significant concavity or convexity, captured by a quadratic and a cubic term, exists in the relation (Column II and Column IV).

Table 5: Balanced Growth Path relationship

Dep. Variable $(A\_{9296})_r=(Pat\_{9296})_r * \hat{\beta}_r /1000$

	I	II	III	IV
	All Observations (141 obs)		**Without top 2 outliers (139 obs)**	
Constant	-0.18	0.33	0.15	-0.11
	(0.14)	(0.17)	(0.08)	(0.07)
A_{7592}	0.35^*	0.22^*	0.36^*	0.37^*
	(0.03)	(0.10)	(0.02)	(0.07)
$(A\_{7592})^2$		0.006		0.001
		(0.005)		(0.008)
$(A\_{7592})^3$		0.00006		0.00004
		(0.00004)		(0.0001)
R^2	0.95	0.96	0.96	0.96

Moreover regressing the estimated residuals e_r from the regressions in Table 4 (not reported) we find no significant correlation with $L_r^{R\&D}$ and h_r. Neither in level nor in logs does there exist any significant correlation between the residuals and the two explanatory variables. Our data do not reject the assumption that knowledge creation in European and North-American regions has been on average on BGP with a common long run growth rate of A_r. Deviations from this BGP have been rather small, and random. In particular they were uncorrelated with regional R&D. As a consequence equation (8) is correctly specified.

5.2 Basic Specification

We have established, so far, that knowledge flows across regions depend on several bilateral characteristics. Proving the existence of these flows though, does not prove the existence of externalities (and in particular of positive externalities) of knowledge on innovation. Available knowledge originating in other regions may very well bring, together with contribution for new ideas, increased standard of novelty, a reduction in the "yet unexplored" innovation possibilities. It may push companies or the government to employ less able people in R&D. These effects may very well generate a zero net effect or even a negative net effect on the productivity of researchers in innovation. Therefore no clear prior exists on the sign and magnitude of ε_I in equation (8). Conversely, we expect ε_L and ε_h to be non negative, as an increase of researchers and of their resources should not decrease their findings of new ideas.

We estimate the innovation function in (8). First we construct the stock of available ideas in each region r, A_r^{AVA} as $\sum\limits_{s\in S} \exp\left(\widehat{\gamma}_1(d_1)_{r,s} + \widehat{\gamma}_2(d_2)_{r,s} + \gamma_3(d_3)_{r,s} + \widehat{f}(\underline{Char_{r,s}})\right)\widehat{\beta}_s P_s$. The term in parenthesis is the estimate of $\phi(r,s)$ using relation (2). We use the parameter estimates from Table 2 column I in order to calculate this value. In estimating the equation we include country fixed effects C_i, as the propensity to patent in the US may differ across countries, due to different costs, and we control also for the previously estimated average intensity of ideas per patent in regions, $\ln(\widehat{\beta}_r)$. The exact specification that we estimate is as follows:

$$\begin{aligned} \ln(P_r) \quad &= \quad C_i + c\ln(\widehat{\beta}_r) + \varepsilon_L \ln(L_r^{R\&D}) + \varepsilon_h \ln(h_r) + \\ &\quad\; +\varepsilon_I \ln A_r^{AVA} + u_{r,s} \end{aligned} \tag{10}$$

Table 6 reports the estimates of the coefficients of equation (10). We estimate the equation using instrumental variables technique and we report heteroskedasticity-robust standard errors. As the variable A_r^{AVA} is measured potentially with error and is endogenous to the innovation process we instrument it with $R\&D_r^{AVA}$

calculated as follows:

$$R\&D_r^{AVA} = \sum_{s \in S} \exp\left(\widehat{\gamma}_1 (d_1)_{r,s} + \widehat{\gamma}_2 (d_2)^{r,s} + \widehat{\gamma}_3 (d_3)_{r,s} + \widehat{f}(\underline{Char_{r,s}})\right) (L^{R\&D} * h)_s$$

In this set of estimates we maintain that total R&D spending ($L^{R\&D} * h$) is exogenous. We will deal with potential endogeneity problem of R&D resources in section 5.4. We use the flow-weighted amount of R&D regional spending as instrument for the flow-weighted amount of available ideas .

Table 6: The innovation function

Dep. Variable $\ln(Pat_{9296})$

A^{AVA} is measured using the regression estimates of phi(r,s) to construct flows (weights)

Spec.	**I (All obs.)**	**II (All obs.)**	**III No Outliers**	**IV No Outliers**	**V Europe Only**	**VI North America Only**
$\ln(L^{R\&D})$	0.83*		0.92*			
	(0.11)		(0.12)			
ln(h)	0.62*		0.58*			
	(0.16)		(0.16)			
$\ln(L^{R\&D}h)$	p-value[1]:	0.73*	p-value[1]:	0.73*	0.68*	0.74*
	0.40	(0.06)	0.18	(0.08)	(0.10)	(0.10)
$\ln(A^{AVA})$	0.25	0.32	0.01	0.11	0.12	0.43
	(0.21)	(0.27)	(0.19)	(0.26)	(0.32)	(0.42)
$\ln(\beta_r)$	-0.53*	-0.48*		-0.63*	-0.77*	-0.17*
	(0.13)	(0.18)		(0.12)	(0.15)	(0.08)
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	141	141	127	127	79	62
R^2	0.92	0.92	0.89	0.89	0.92	0.92

Notes: Method of estimation: IV with robust std. error. Instrument for $\ln(A^{AVA})$: $\ln(R\&D^{AVA})$

Outliers: Regions with more 50,000 patents or less than 50 Patents in 1975-1991.

1: The p-value is relative to the test of equality of the two coefficients: ln(R&D)= ln(h)

The first Column of Table 6 presents the estimates for the basic specification, using data for all regions and including, as separate inputs, R&D employment, $L_r^{R\&D}$, and spending per R&D employee, h_r. While these two inputs are both very important in generating innovation, with an elasticity of 0.83 and 0.62 respectively, the elasticity of the stock of available knowledge is not significantly different from 0. While the estimates of the coefficient of $\ln(A_r^{AVA})$ are not very precise, its point estimate (0.25) is much lower than the elasticity of R&D resources. A formal test does not reject the hypothesis that the coefficients on $\ln(L_r^{R\&D})$ and on $\ln(h_r)$ are equal. Therefore we estimate specification in column II, using this restriction and controlling for total R&D spending ($L_r^{R\&D} * h_r$) as an input of production. This restriction improves the precision of the estimates of innovation elasticity to R&D. Elasticity of new ideas to total R&D spending is 0.73. Again the effect of $\ln(A_r^{AVA})$ on innovation is not significantly different from 0, but both the point estimate (0.32) and standard error (0.27) are larger than in column I.

In order to check the robustness of our estimates, and to improve, potentially, the precision of the estimates of the ε_I coefficient we re-estimate our equation omitting outlier regions. These regions are those that patented disproportionately more (California, New York and New Jersey all with more than 50,000 inventions in the 1975-91 period) and those that patented disproportionately less (several regions in Greece, Southern Italy, Spain and East Germany with less than 20 patents during 1975-91) than the rest. As we see from column III and IV of our specification, the elasticity of innovation to R&D resources remains close to 0.7, and the effect of available knowledge is even closer to 0 (0.01-0.11). In this case we can reject at 5% confidence level, using a double-sided test, that the coefficient on $\ln A_r^{AVA}$ is equal to the coefficient on $\ln(L_r^{R\&D})$ in specification III, or to that on $\ln(L_r^{R\&D} h_r)$ in specification IV.

Finally, in order to see if the behavior of innovation is different for North America and Europe, we re-estimate the equations on regions of each area separately (column V and VI). Besides a general increase in standard errors, results are similar to those obtained for the two areas together. Elasticity of innovation to R&D still close to 0.7 for both areas and impact of available knowledge much smaller. Only the impact of $\ln(A_r^{AVA})$ is somewhat larger for North America than for the whole sample, but the precision of this estimate worsens significantly.

5.3 Simple measure of Knowledge Flows

The reader may be worried that the estimates of $\ln A_r^{AVA}$ rely too much on the modelling of the knowledge flows across regions. Parameters entering the calculation of $\widehat{\phi}(r,s)$ are estimated using the model of knowledge diffusion represented by expression (2) and therefore susceptible to any criticism addressed to that model. In order to dissipate these worries we use here a more direct measure of $\phi(r,s)$. Considering its definition in expression (3) and simply using the correspondence between ideas and patents and between their flows and citation. We construct the index as

$$\widetilde{\phi}(r,s) = \frac{citations(s,r)/(Patents * Ave.cit.)_s}{\sum_s (citations(s,r)/(Patents * Ave.cit_s))}. \tag{11}$$

The numerator captures citations between region r and s standardized by the number of patents, $Patents$, in the cited region s multiplied by the average number of citations received by patents of region s. The term $Ave.cit.$ adjusts the number of patents generated in each region to account for their importance measured as average number of citations received by the region's patents. The standardization at the denominator guarantees that the index is between 0 and 1, that it does not depend on average citation propensity of region r and that the sum over "sending" regions is equal to 1. As all quantities on the right hand side are observable we can directly measure $\widetilde{\phi}(r,s)$. The values obtained are used to construct A_r^{AVA} as $\sum\limits_{s\in S}\widetilde{\phi}(r,s)\Delta A_s$. We can, then, re-estimate equation (10) using this new measure of A_r^{AVA} and using as dependent variable the citation-weighted patents of region r (rather than using $\ln(\widehat{\beta}_r)$) to control for the region-specific importance of patents. Similarly we use $\sum_{s\in S}\widetilde{\phi}(r,s)(L_s^{R\&D}h_s)$ to calculate $R\&D_r^{AVA}$ that is used as instrument. Table 7 reports the estimates for the same specifications as in Table 6.

Table 7: The innovation function

Dep. Variable: ln(citation-weighted Patents)

A^{AVA} is measured using the direct measures of phi(r,s) to construct flows (weights)

Specification	**I (All obs.)**	**II (All obs.)**	**III No Outliers**	**IV No Outliers**	**V Europe Only**	**VI North America Only**
$\ln(L^{R\&D})$	0.66*		0.74*			
	(0.10)		(0.11)			
ln(h)	0.66*		0.63			
	(0.19)		(0.18)			
$\ln(L^{R\&D}h)$	p-value[1]:	0.60*	p-value[1]:	0.70*	0.59*	0.57*
	0.98	(0.07)	0.68	(0.04)	(0.08)	(0.08)
$\ln(A^{AVA})$	0.10	0.20	-0.05	-0.09	0.10	0.05
	(0.20)	(0.23)	(0.27)	(0.28)	(0.24)	(0.37)
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	141	141	127	127	79	62
R^2	0.60	0.58	0.61	0.60	0.66	0.73

Notes: : Method of estimation: IV with robust std error. Instrument for $\ln(A^{AVA})$: $\ln(R\&D^{AVA})$

Outliers: Regions with more 50,000 patents or less than 50 Patents in 1975-1991.

1: The p-value is relative to the test of equality of the two coefficients: ln(R&D)= ln(h)

The elasticity of Innovation to total R&D resources is still in the vicinity of 0.6-0.7. The restriction of equal coefficients on R&D employees and R&D spending is still not rejected by the data. Those coefficient are nearly identical in column I. The effect of available knowledge is still not significantly different from 0 in any specification. In this set of estimates, the point estimates of knowledge externalities is even lower than before, in some cases negative. In four specifications out of six we can reject at the 5% level that the coefficient on $\ln(A_r^{AVA})$ is equal to the coefficient on $\ln(h_r)$ or on $\ln(L_r^{R\&D}h_r)$. Interestingly, even in this case, the specifications that omit outliers (III and IV) produce the lower point estimates of knowledge externalities (-0.05 and -0.09). While measurement error for the variable A_r^{AVA} could generate attenuation bias, the use of IV method and the fact that these estimates are close to the previous ones reassures us.

5.4 Instrumental Variables: Regional Market Potential

Our estimates of the effect of R&D on innovation could be biased by the presence of some unobservable regional factors that attract R&D while also increasing its returns in terms of innovation. One of these factors could be past accumulated knowledge that may also generate endogeneity: R&D goes where past ideas have been generated because available past ideas increase research productivity, past ideas in turn are correlated with new ideas generated. In order to address this issue we need an instrument for R&D which is correlated with its regional distribution but not with its regional productivity in generating innovation. It is reasonable to assume that R&D distribution across regions is affected by the return of doing research in each region. If the expected return from inventions is larger in some regions we would expect more research done there. Expected return from innovation is given by the value of a patent, and, due to transport costs, a patent is more valuable where there is a larger local market for innovations. If the same invention is more valuable where there is a larger local market but the local market has no impact on making R&D resources more productive then we can use market potential as an instrument for R&D[10].

Different regions have different market potentials for innovations, depending on their location and their connections with the other markets. Interestingly we have a revealed measure of the potential market for patented goods, at least in each country. This measure is the extent to which patent protection is pursued by residents of a country in other countries. If a patent is protected only on the domestic market it is because the inventors believe that small profits would come by trading the good abroad and therefore it is not worth seeking protection there. On the contrary a patent that is protected in a broad collection of countries reveals the intention of its inventors to protect their potential profits in foreign markets. We assume that seeking international patenting in other countries reveals what areas the inventor considers as potential markets for the new good.

We use data available from WIPO (World Intellectual Property Organization) on patents granted in each of our 19 countries to inventors residing in each of them for the period 1993-1996. Identifying the patent protection in a country as the sign that an invention is targeting that country as a market, allows us to estimate the effect of country characteristics and distance on the potential demand for innovations coming from each country. This allows us to calculate the market potential for innovation in each of our regions. We use this measure of market potential as instrument for regional R&D.

Here we briefly describe how we use the data on cross-country patenting to infer the market potential for innovation in each region. Let's denote with π_{ij} the share of world patents granted in country j to residents of country i. This is a good proxy of the market for new goods that inventions from country i have in country j (in relative terms). π_{ij} is the demand coming from country j for innovations generated in country i.

It is useful to think of the market π_{ij} as depending on country j and country i characteristics and on bilateral characteristics affecting the relative intensity of patenting from a country into another. The bilateral characteristics that we use are geographical distance ($dist_{ij}$) and a same-country dummy ($SameC_{ij}$). π_{ij} can be decomposed as follows:

$$\pi_{ij} = (\Pi_{i,all})(\Pi_{all,j})(e^{-\delta dist_{ij}})(e^{fSameC_{ij}})\varepsilon_{ij} \tag{12}$$

[10]This type of instrument for R&D is also used in Bottazzi and Peri [2] .

$\Pi_{i,all}$ is the share of total patents in the world granted to inventors that are residents of country i, $\Pi_{all,j}$ is the share of of total patents in the world granted by the patent office of country j. $e^{-\delta dist_{ij}}$ is an exponential function in the geographical distance between i and j, $e^{fSameC_{ij}}$ is the effect of patenting in the same country where the inventor is resident and ε_{ij} is a positive multiplicative random factor distributed as a lognormal that captures all the other unobserved bilateral determinants of π_{ij}. As π_{ij}, $dist_{ij}$ and $SameC_{ij}$ are observable we may use a simple regression to perform the above decomposition. Taking logs of both sides of equation (12) we have

$$\ln(\pi_{ij}) = C_i + C_j - \delta(dist_{ij}) + f(SameC_{ij}) + u_{ij}, \tag{13}$$

where $C_i$and C_j are country specific effects and capture the origin and destination country effects $(\ln(\Pi_{i,all}), \ln(\Pi_{all,j}))$. $dist_{ij}$ is the geographical distance between the two regions and $SameC_{ij}$ is a dummy which is equal to one if the inventor's country and the granting country are the same. $u_{ij} = \ln(\varepsilon_{ij})$ is a zero mean, normally distributed error. Once we have used OLS to estimate equation (13) we can construct the predicted share $\pi_{ri,dj}$ which measures the innovation generated in region r_i of country i and demanded in region d_j of country j. The parameters C_j, δ and f depend on where inventions are patented and how the distance and national borders affect this marketing decision. They allow us to capture the market potential for innovation. Let's call $shmar_{dj}$ the share of region d_j in the market for innovation of country j. Also we denote with a hat$\widehat{}$ the OLS estimates of our parameters. The predicted potential demand coming from region d_j for innovation invented in r_i would be

$$\widehat{\pi}_{ri,dj} = (shmar_{dj}\widehat{\Pi}_{all,j})\exp(-\widehat{\delta}dist_{ri,dj} + \widehat{f}SameC_{r_i d_j}). \tag{14}$$

In our implementation we measure $shmar_{dj}$ as the share of GDP of country j produced in region d_j Therefore the total market potential for innovation produced in a generic region r of country i is

$$Pot1_r = \sum_{d_j}\sum_{j}\widehat{\pi}_{r,dj}. \tag{15}$$

Alternatively, using a region's gross product as proxy for the demand for new goods (rather than the value $shmar_{dj}\widehat{\Pi}_{all,j}$ derived from international patenting) we can define another measure of demand from region d_j for innovation invented in region r of country i : $\widehat{p}_{r,dj} = \exp(-\widehat{\delta}dist_{r,dj} + \widehat{f}SameC_{r,dj})Y_{d_j}$. The total market potential for region r of country i would be

$$Pot2_r = \sum_{d_j}\sum_{j}\widehat{p}_{r,dj}. \tag{16}$$

Both constructs use the geographic position of each region and the inter-country pattern of demand for innovations, revealed by international patenting, to evaluate the main determinant of a patent's value, which is the market potential that the average patent has in region r of country i. We estimate equation (10) using $Pot1_r$ or $Pot2_r$ and their transform $Potn_{AVA} = \sum_{s \in S} \widehat{\phi}(r,s) Potn_s$ as instruments for total $R\&D$ spending and for A_r^{AVA}. Table 8 reports the results of the instrumental variables estimations using different measures for A_r^{AVA} and different instruments.

Table 8: Instrumenting R&D with Mkt potential
Dep. Variable: ln(Citation-weighted Patents)

Specification:	**I**	**II**	**III**	**IV**	**V**	**VI**
$\ln(R\&D*h)_r$	0.84*	0.85*	0.76	0.91*	0.92*	0.75*
	(0.21)	(0.23)	(0.17)	(0.25)	(0.30)	(0.18)
$\ln(A^{AVA})$		-0.09	0.16		-0.21	0.32
		(0.42)	(0.51)		(0.54)	(0.53)
Country Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	141	141	141	141	141	141
R^2	0.86	0.75	0.77	0.73	0.72	0.77

Notes: Method of estimation: IV with robust std errors. Column I, II and III instrumenting $\ln(R\&D*h)_r$ and $A^{AVA}{}_r$ with ln(Pot1) and $\ln(Pot1_{AVA})$. Column IV, V and VI instrumenting $\ln(R\&D*h)_r$ and $A^{AVA}{}_r$ with ln(Pot2) and $\ln(Pot2_{AVA})$.

$Pot1_r$ and $Pot1_r^{AVA}$ explain only about 10% of the variation of total $R\&D$ spending and of A_r^{AVA}. As a consequence, the IV estimates (specification I,II and III) have larger std. errors than the previous estimates. Similarly large errors are obtained when we use $Pot2_r$ and $Pot2_r^{AVA}$ (column IV, V and VI) as instruments. Column I and IV in Table 8 include only R&D resources as input. The point estimates of the elasticity of innovation to R&D (0.8-0.9) do not change much from the specification in Table 6 and 7 that assumed exogenous R&D. Point estimates are somewhat larger but standard errors increase (by a factor of three) so that no significant difference is revealed. No evidence of a meaningful endogeneity bias is present. When we include the term $\ln(A_r^{AVA})$ either calculated using $\widehat{\phi}(r,s)$ (column II and V) or calculated using $\widetilde{\phi}(r,s)$ (column III and VI) no evidence of significant positive impact of this variable is found. Point estimates are negative in 2 cases and only in one case are they larger than 0.2. Unluckily the large standard errors prevent us from making strong inferences based on these estimates. On average, though, these estimates do not suggest either positive and significant knowledge externalities or a positive and significant bias of the OLS estimates of $\ln(A_r^{AVA})$.

5.5 Impact of Knowledge from "Innovators"

As we did for knowledge diffusion, we consider here if knowledge generated only in the main "innovating regions" has a significant impact on innovation of other regions. Regions that are sources of most international innovation, may be the true stimulators in knowledge creation. Adding other regions as source of spillovers may simply add noise to our estimates. Also, considering the effect of knowledge going from these "innovators" to "receiving" regions helps us correcting the potential R&D endogeneity. In this case we can safely assume that R&D done in the innovating regions is exogenous for the receivers. This exercise, although applied to the innovation process rather than to productivity, is similar to what Keller [17] does by analyzing the impact of R&D in the five largest world innovators on productivity of other nine countries. No expectations of similar results should arise, though, as we are considering knowledge externalities in innovation while he is concerned with any technological externality in production. The relevant knowledge available to region r, denoted in this case as $A_r^{AVA,20}$, is calculated as $\sum_{\widetilde{s=1}}^{20} \widetilde{\phi}_{20}(r,s)\Delta A_s$. The summation is taken over the twenty top-country innovators. $\widetilde{\phi}_{20}(r,s)$ is calculated as $\widetilde{\phi}(r,s)$ and standardized so that the sum across the twenty top-innovators is equal to one. ΔA_s is measured as the citation-weighted patents generated in each of the innovating regions (s), in 1975-1991. The results of the estimation of the innovation function are reported in Table 9.

Table 9: The innovation function
Dep. Variable: ln(Citation-weighted Patents)
$A^{AVA,20}$ is measured using only 20 top innovators

Specification	I (including all "receivers")	II (including all "receivers")	III Europe Only	IV North America Only
$\ln(L^{R\&D})$	0.64*			
	(0.11)			
ln(h)	0.68*			
	(0.19)			
$\ln(L^{R\&D}h)$	p-value:	0.65*	0.61*	0.71*
	0.88	(0.05)	(0.11)	(0.05)
$\ln(A^{AVA,20})$	-0.24	-0.23	-0.27	-0.37
	(0.25)	(0.25)	(0.31)	(0.39)
Country Dummies	Yes	Yes	Yes	Yes
Observations	122	122	64	58
R^2	0.74	0.74	0.58	0.72

Notes: : Method of estimation: IV with robust std errors. Instrument for $\ln(A^{AVA,20})$: $\ln(R\&D^{AVA,20})$

1: The p-value is relative to the test of equality of the two coefficients: ln(R&D)= ln(h).

The first column of Table 9 presents the estimates including all "receiving" regions, i.e. those regions that are not top innovators in the nineteen countries. The estimates of elasticity of innovation to employment $(L^{R\&D})$ and to resources per worker in R&D (h) are very similar to the estimates obtained in Table 7, using all regions. The effect of spillovers from the top innovators only, is still not significantly different from 0. In all specifications of Table 9 the point estimate on knowledge externalities is negative and in all cases the hypothesis of equal coefficients on knowledge available and on total R&D spending is rejected at the 1% confidence level. Column II imposes the restriction of equal coefficients on R&D employees and resources per R&D employee and estimates an elasticity of innovation to total R&D spending of 0.65. The separate estimates for European and North American regions (column III and IV) do not reveal any relevant difference in productivity of R&D resources and knowledge available in generating innovation. The elasticity of innovation to R&D is still in the 0.6-0.7 range, while available knowledge has no significant effect.

To sum up, while no estimate, singularly taken, makes us certain that available knowledge has no effect on innovation, the overall picture, emerging from several specifications, does not provide any evidence in favor of positive and significant knowledge spillovers in innovation.

6 Conclusions

The present study analyzes the process of knowledge diffusion and knowledge externalities as evidenced aggregating data on patent creation and patent citations. While diffusion of ideas is needed in order to have externalities of knowledge, there is no reason to believe that simply measuring the intensity and scope of this knowledge diffusion provides us with a measure of knowledge externalities. In order to have positive externalities, in fact, we need that existing ideas affect positively the productivity of scientists in generating new ideas. Some doubts that knowledge has positive externalities on innovation have been raised by the literature that analyzes innovation over time (Griliches [8]). There seems to be evidence that productivity of R&D in innovation has decreased over time. Diffusion of new ideas, in fact, brings not only "new inspiration" to researchers but also increased standard for innovation, and it reduces the unexplored territories of human knowledge. These effects may offset the positive spillovers. Our study finds that there are very important positive and negative determinants of knowledge diffusion: regions farther away from each other, in different countries, specialized in different sectors and speaking different languages exhibit much lower flows of knowledge than close, similar regions in the same country. Nevertheless these flows do not bring significantly positive knowledge externalities when we consider the effect of available knowledge on the innovation generating function. While we think that this study confirms the presence of decreasing returns in innovation, already suspected from time-series data analysis, we do not address here the impact of knowledge externalities on production. Knowledge spillovers could be good for production as they are one component (but not the only one) of technological spillovers. As other works (e.g. Coe and Helpman [5], Keller [17]) find evidence of positive technological externalities on production, it would be interesting to measure the impact of knowledge externalities on production in particular. We leave this line of research to further development of our work.

References

[1] Aghion Philippe and Peter Howitt (1992) "A Model of Growth through creative Destruction" *Econometrica* 60: 323-51.

[2] Bottazzi Laura and Giovanni Peri (2002) "Innovation and Spillovers in Regions: Evidence from European Patent Data" forthcoming in *European Economic Review*.

[3] Caballero R. and A. Jaffe (1993) "How High are the Giants' Shoulders: An Empirical Assessments of Knowledge Spillovers and Creative Destruction in a Model of Economic Growth" *NBER Macroeconomic Annuals* vol. 8 pp.15-72 MIT Press, Cambridge MA.

[4] Ciccone Antonio (2002) "Agglomeration effects in Europe" *European Economic Review* (46)2 pp. 213-227

[5] Coe David and Elhanan Helpman (1995) "International R&D Spillovers" European Economic Review, 39 pp. 859-887.

[6] Griliches Zvi Eds. (1984) " R&D Patents and Productivity" University of Chicago Press, Chicago.

[7] Griliches Zvi (1990) "Patent Statistics as Economic Indicators: A Survey" *Journal of Economic Literature*, 28(4), pp.1661-1707.

[8] Griliches Zvi (1994) "Productivity, R&D and the Data constraint" *American Economic Review*, 84(1) pp. 1-23.

[9] Grossman G. and E. Helpman (1991) "Endogenous Product Cycles" Economic Journal, v. 101 pp: 1214-1229.

[10] John Helliwell (1988) "How Much do National Borders Matter?" (Brookings Institution Press, Washington D.C.)

[11] Bronwyn H. Hall, Adam B. Jaffe and Manuel Trajtenberg, "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," National Bureau of Economic Research Working Paper No. 8498 (October 2001)

[12] 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.

[13] Jaffe Adam, Manuel Trajtenberg and Rebecca Henderson (1992) "Geographic Localization of Knowledge Spillovers as evidenced by Patent Citations" *Quarterly Journal of Economics* 108(3) pp. 577-98

[14] Jaffe A. and M. Trajtenberg (2002) "Patents, Citations and Innovations" (MIT Press Cambridge Ma)

[15] Jones C. (1995a) "Time Series tests of Endogenous Growth Models" *Quarterly Journal of Economics* 110 (May 1995).

[16] Jones C. (1995b) "R&D Based Models of Economic Growth" *Journal of Political Economy* 103: pp.739-784.

[17] Keller Wolfgang (2002)"Geographic Localization of International Technology Diffusion" forthcoming in *American Economic Review.*

[18] John McCallum (1995) "National Borders Matter: Canada-US Regional Trade Patterns" *American Economic Review*, Vol 85 (3) pp. 615-623.

[19] Romer P.M., 1990 "Endogenous Technological Change" *Journal of Political Economy* 98(5), 71-102

A List of Regions

Austria: OSTOSTERREICH*, SUDOSTERREICH, WESTOSTERREICH

Belgium: BRUXELLES, VLAAMS GEWEST*, REGIONE WALLONNE

Canada (Provinces): NEW FOUNDLAND, PRINCE EDWARDS ISLAND, NOVA SCOTIA, NEW BRUNSWICH, QUEBECK, ONTARIO*, MANITOBA, SASKATCHEWAN, ALBERTA, BRITISH COLUMBIA.

Denmark: DENMARK*

Finland: FINLAND*

France: ILE DE FRANCE*, BASSIN PARISIENNE, NORD-PAS DE CALAIS, ESTE, OUESTE, SUD-OUEST, CENTRE-EST, MEDITERRANEE.

Germany: BADEN-WURTENBERG, BAYERN, BERLIN, BRANDENBURG, BREMEN, HAMBURG, HESSEN, MECKLENBURG-VORPOMMEM, NIEDERSACHSEN, NORDRHINE-WESTFALIA*, RHEINLAND-PFALZ, SAARLAND, SACHSEN, DESSA, SCHLESWIG-HOLSTEIN, TURINGEN.

Greece: VORAIA ELLADA, KENTRIKI ELLADA, ATTIKI*, NISIA AIGAIOU, KRITI.

Ireland: IRELAND*

Italy: NORD OVEST, LOMBARDIA*, NORD-EST, EMILIA ROMAGNA, CENTRO, LAZIO, ABRUZZO-MOLISE, CAMPANIA, SUD, SICILIA, SARDEGNA.

Luxemburg: LUXEMBURG*

Norway: NORWAY*

Portugal: PORTUGAL*

Spain: NOROESTE, NORESTE, COMUNIDAD DE MADRID, CENTRO, ESTE*, SUR, CANARIAS.

Sweden: SWEDEN*

Switzerland: REGIONE LEMANIQUE, ESPACE MITTELAND, NORTHWESTSCHWEITZ*, ZURICH, OSTCHWEITZ, ZENTRALSCWEITZ, TICINO.

United Kingdom: NORTH, YORKSHIRE AND THE HUMBER, EAST MIDLANDS, EAST ANGLIA, SOUTHEAST*, SOUTHWEST, WEST MIDLANDS, NORTH WEST, WALES, SCOTLAND, NORTHERN IRELAND.

USA (States): ALABAMA, ALASKA, ARIZONA, ARKANSAS, CALIFORNIA*, COLORADO, CONNECTICUT, DELAWARE, D.C., FLORIDA, GEORGIA, HAWAII, IDAHO, ILLINOIS, INDIANA, IOWA, KANSAS, KENTUCKY, LOUISIANA, MAINE, MARYLAND, MASSACHUSSETS, MICHIGAN, MINNESOTA, MISSISSIPI, MISSOURI, MONTANA, NEBRASKA, NEVADA, NEW HAMPSHIRE, NEW JERSEY, NEW MEXICO, NEW YORK*, NORTH CAROLINA, DAKOTA, OHIO, OKLAHOMA, OREGON, PENNSYLVANIA, RHODE ISLAND, SOUTH CAROLINA, SOUTH DAKOTA, TENNESSEE, TEXAS,

UTAH, VERMONT, VIRGINIA, WASHINGTON, WEST VIRGINIA, WISCONSIN, WYOMING.

*= Largest innovator(s) in the Country

B Data Appendix

B.1 R&D Expenditure Data (1992-1996)

- **Europe:**

  Main Source for Data on R&D1992-1996: Eurostat Regio Database

  (http://europa.eu.int/comm/eurostat)

As there were some missing values for some regions we interpolated existing values or we imputed regional values using the share of national R&D in the region from a previous year applied to the national Figure for the year. The following is the detailed description of the interpolated and imputed data:

Austria : linear interpolation for 1992, imputed for 1994,1995.

Belgium: linear interpolation for 1992 country total, imputed for 1993

Denmark: imputed for 1994-95.

Germany: imputed for 1992-94-95.

Greece: imputed for 1995.

Spain: imputed for 1995.

France: imputed for 1995.

Italy: imputed for 1992-94-95

The Netherlands: imputed for 92-95.

Portugal: linear interpolations for 93,94.

Sweden: linear interpolations for 92,94.

U.K. Inputed for 1992.

Switzerland used regional GDP shares to get regional values 92-95 where total R&D Expenditure had been obtained from the Swiss Statistical Office.

- **U.S.A.:**

Main Source: National Science Foundation/Division of Science Resources Studies, Survey of Industrial Research and Development: 1998.

Missing values for 1992 and 1994 were obtained through linear interpolation.

Other Interpolations due to 'NA':

1991:Colorado, Kansas and North Dakota, 1989 and 1991: Idaho, Missouri, Maine, Montana, New Hampshire, West Virginia and Vermont.

For Delaware the growth rate between 1993 and 1994 was applied to get 1992 value.

- **Canada:**

Main Source: The document Cat No. 88F0006XIB01001" Estimates of Canadian Research and Development Expenditures(GERD), Canada, 1989 to 2000, and by Province 1989 to 1998." obtained from www.statcan.org.

Exchange rates:

http://www.oanda.com/convert/fxhistory

B.2 R&D Employment Data (1992-1996)

- **Europe:**

Main Source for Data on R&D1992-1996: Eurostat Regio Database

(http://europa.eu.int/comm/eurostat)

Missing values were treated in the same way as done for R&D Spending:

Austria: linear interpolation for 1992, imputed for 1994-95.

Belgium: linear interpolation for 1992-93 country total, imputed 1992-93 regional values.

Denmark: linear interpolation for 1994.

Germany: linear interpolation for 1992, imputed regional values for 1992-94.

Greece: linear interpolation for 1992, imputed values for 1994-95.

Italy: Imputed values for 1992-93.

Portugal: Linear interpolations for 1993-94.

Sweden: Linear interpolations for 1992-94.

U.K.: Imputed regional values for 1992, used 92-93 growth rates to impute 1994-95 values.

Switzerland: used regional GDP shares to impute regional values 1992-95 where total R&D Employment had been obtained from the Swiss Statistical Office.

- **U.S.A.:**

The data on employment by state have been obtained using the share of Scientists and Engineers by state in total employment from the Census 1991, and applying them to the total employment in 1992-1996.

- **Canada:**

Values for provinces in 1992-94 were interpolated using the share of each province in 1995 multiplied by the total employment in R&D for that year. Value for 1995 was obtained from www.statcan.org.

B.3 Population and Employment Data (1992-1996)

- **Europe:**

Main Source: Eurostat Region CD 1999. Demographic Statistics section.
(London is missing for all years)
Norway: Statistics Norway at
http://www.ssb.no/english/subjects/02/nos_c607_en/tab/t-105.html
Switzerland: file Swiss_cantonal_income_90_99 from Swiss Statistics- Swiss Federal Statistical Office-
(http://www.statistik.admin.ch/eindex.htm)

- **U.S.A.:**

Bureau of Economic Analysis website, regional statistics section
http://www.bea.doc.gov/bea/regional/spi/

- **Canada:**

CANSIM database at www.statcan.org

C BGP implies Condition 1 Section 5.1

Assume that A_{rt} is the existing stock of ideas in region r and year t and that g is the common yearly growth rate of the stock of ideas in BGP for each region r. Then considering three years such that $t_0 < t_1 < t_2$ we have, for the generic region i:

$$A_{rt_2} - A_{rt_1} = [(1+g)^{t_2-t_1} - 1]A_{rt_1} \tag{17}$$

$$A_{rt_1} - A_{rt_0} = [(1+g)^{t_1-t_0} - 1]A_{rt_0} \tag{18a}$$

$$A_{rt_1} = (1+g)^{t_1-t_0} A_{rt_0} \tag{19}$$

Merging the three expressions and considering $t_0 = 75$, $t_1 = 92$, $t_2 = 96$ we obtain $(A_{r96} - A_{r92}) = \gamma\,(A_{r92} - A_{r75})$ where $\gamma = \{[(1+g)^4 - 1](1+g)^{17}\}/[(1+g)^{17} - 1]$. Measuring the change of stock of ideas between two years as the number of patent granted times the estimated intensity of ideas in each patent the equation written above yields: $(Pat_{9296} * \widehat{\beta})_r = \gamma(Pat_{7591} * \widehat{\beta})_r$. Adding a random disturbance we have the relation under 1 in Section 5.1.