Knowledge Flows and Productivity

Giovanni Peri*
University of California, Davis

La diffusione di conoscenze scientifico-tecnologiche a livello nazionale e internazionale è una determinante fondamentale del progresso economico. In questo articolo presentiamo una rassegna della letteratura sulla diffusione di conoscenza e proponiamo un metodo per stimare tale diffusione usando citazioni tra brevetti. Citazioni tra brevetti rivelano diffusione di idee alla frontiera tecnologica. Usiamo dati di 147 regioni in Europa e Nord America durante il periodo 1975-1996. La nostra analisi rivela che la distanza geografica e tecnologica costituiscono un'importante barriera alla diffusione di conoscenza. Inoltre tali flussi di conoscenza paiono avere un effetto positivo, ma piccolo, sulla produttività delle regioni.

National and international flows of knowledge are fundamental determinants of technological progress. In this article we review the existing literature on knowledge flows and we propose a method for estimating them, based on patent citations. Citations are links between inventions that reveal a learning process at the technological frontier. We use data for the period 1975-1996 for 147 subnational regions in Europe and North America. We find that geographical distance and technological differences constitute major barriers to knowledge flows. We also show that these flows may have positive, but small, effect on total factor productivity. [JEL Code: O47, R11, F02].

1. - Introduction

The concepts of knowledge flows and knowledge spillovers have

---

* *gperi@ucdavis.edu*. I thank Marianne Feldman, Roger Gordon, Bronwyn Hall, Gordon Hanson, Ann Harrison, John Helliwell, Henry Overman, Manuel Trajtenberg, MichelleWhite for their helpful suggestions. Shireen Al Azzawi provided excellent Research Assistance. Errors are mine.

pervasively populated the theoretical literature on economic growth for the last ten years. Innovation is the engine of growth, has been argued, and spillovers from existing knowledge are the "renewable" fuel for this engine. While it is clear how to characterize the source and the effect of these spillovers within the frame of specific theoretical models such as Aghion and Howitt (1992), Grossman and Helpman (1991) or Romer (1990), it is much less clear how to measure them in a precise but robust way using the data. First and foremost it is not clear that there are data that allow us to measure such an elusive concept as knowledge flows, or their effect on productivity, without adding several other assumptions.

Knowledge spillovers hinge on the diffusion of ideas, a process that leaves no track in the data. Economists have assumed, in order to estimate these spillovers, that diffusion of ideas depends on proximity in space, in technological specialization or in economic development. These are plausible assumptions but each of them is potentially controversial and should be tested. In particular we do not have a good measure of the quantitative dependence of knowledge diffusion on distance or technological proximity. Alternatively, several studies have decided to proxy diffusion of ideas with diffusion of goods and have used data on trade or foreign direct investment to track spillovers of research and development (e.g. Coe and Helpman, 1995). Interestingly, the theoretical tradegrowth literature has emphasized in several influential studies the importance of analyzing international knowledge flows as channels of productivity diffusion and growth, differentiating them from trade flows. Influential authors have argued that knowledge (rather trade) flows, are responsible for development and growth. Rivera Batiz and Romer (1991) show that «... trade in goods has no effect on the long-run rate of growth "while" ...allowing flows of ideas (i.e. knowledge flows) results in a permanently higher growth rate. — They go on stating that — [f]lows of ideas deserve attention comparable to that devoted to flows of goods». Grossman and Helpman (1991) in Chapter 9 of their very influential book *Innovation and Growth* point out that «[T]he growth effect of knowledge spillovers and those of commodity trade are conceptually distinct — and they develop models that show

how — "the most important benefit to a country from participating in the international economy might be the access that such integration affords to the knowledge base in the world at large"». Feenstra (1996) argues that convergence in growth rates across countries takes place only if «... Trade occurs simultaneously with international diffusion of knowledge» while if no diffusion of knowledge occurs, trade could actually generate divergence. Moreover, scientific and technological knowledge has been recognized for a long time as an important factor of production on par with labor and capital (Solow, 1956) and its growth regarded as the propellant of economic growth (Solow, 1957). Stimulated by these theoretical speculations one would think that the empirical trade-growth literature has made an effort to develop better measures of international knowledge flows, explicitly differentiating them from trade flows, and explicitly analyzing their effect on productivity growth. This has not happened in a significant way, yet. Certainly knowledge flows are hard to define, observe and measure so that our understanding of knowledge flows is still in its infancy if compared to the analysis of trade flows.

The goal of this paper is to survey different methods and data used in the literature to measure knowledge flows and their effects. We also present and develop a new and interesting method that uses data on patented innovation and citations across patents to measure the strength of knowledge flows across subnational regions covering Western Europe and North America. Finally, we use our estimated flows to calculate their impact on total factor productivity (*TFP*) across these regions. So our procedure and our article develops two subsequent steps. First we establish the presence of knowledge flows tracking "learning" of ideas across regions with the use of patent to patent citations. Second we establish the impact of flow-weighted R&D on the productivity of regions. Previous existing work (with the notable exception of Caballero and Jaffe, 1993) merges or assimilates these two phases. Either some mechanical assumption is made on the availability of ideas across space (as in Coe and Helpman, 1995 and Keller, 2002*a*) so that only a productivity equation is estimated, or knowledge flows are measured without worrying about their effect on productivity (Jaffe *et* Al., 1992; Jaffe and Trajtenberg, 2002). Our data on patent citations extend from 1975 to 1996 and allow us to track diffusion over a fairly long period of time. We find that knowledge flows depend on several factors and they are well described, like trade flows, using a gravity-like equation. Most importantly geographical and technological proximity and crossing a national border affect the intensity of diffusion of ideas. On the other hand we find that flow-weighted R&D from other regions has only small effects in determining regional productivity.

The rest of the paper is organized as follows: Section 2 introduces an easy framework to think about knowledge flows and their effects on productivity. Section 3 reviews the existing literature on knowledge flows and spillovers, Section 4 describes the measures of total factor productivity that we calculate for the regions. Section 5 presents the estimates of the intensity of knowledge flows obtained using data on patent to patent citations and section 6 presents the estimates of regional productivity and some preliminary estimates of the impact of knowledge flows on productivity. Section 7 concludes the paper.

2. - Basic Framework

Consider the measure Q_{it} as an index of the technological development of economic unit i at time t. Frequently in the literature some measure of total factor productivity has been used to capture Q_{it}. Total factor productivity determines how much output could be generated keeping the quantity and quality of labor and capital inputs constant. The units chosen are sub-national regions. Assuming that R&D activity is the main source of technological development, Q_{it} is produced as follows:

$$Q_{it} = (A_{it})^{\gamma} \, (A^{a}_{it})^{\mu} \tag{1}$$

A_{it} is the stock of past accumulated R&D resources invested yearly in region i (we indicate them as $R\&D_{it}$). A^{a}_{it} is the stock of past accumulated R&D resources invested in regions other than i

and "accessible" (hence the a superscript) to region i at time t. The objective of our analysis is to construct a measure of the two stocks A_{it} and A_{it}^{a} for European and North-American regions and to estimate their impact, captured by γ and μ, on the regional technological level. The accumulation of A_{it} is simply described as $\Delta A_{it} = R\&D_{it} - \delta A_{it}$ where the depreciation rate of R&D capital is equal to δ. We apply such "perpetual inventory method" to calculate the value of such stock.

Our main focus, however, is on the construction of A_{it}^{a} and on the estimate of μ. In the presence of complete and immediate diffusion of R&D results from any region of origin into any other region the total external stock of R&D (or the "pool" of R&D, as defined by Griliches, 1992) available in i would be $A_{it}^{a}=\sum_{j\neq i} A_{jt}$. However, considering less than perfect diffusion across regions, total accessible R&D in region i would be given by $A_{it}^{a}=\sum_{j\neq i} \phi_{ij}A_{jt}$. In this expression $\phi_{ji} \in [0, 1]$ is the percentage of R&D results generated in region j by time t and accessible to region i. This notation implies that if a certain percentage of R&D results from region j are available to region i it is as if region i had an indirect access to R&D resources of region j. Substituting this last expression for A_{it}^{a} into equation *(1)*, taking logs on both sides and re-arranging we have the following equation:

$$\ln\left(Q_{it}\right) = \gamma \ln\left(A_{it}\right) + \mu \ln\left(\sum_{j\neq i} \phi_{ji} A_{jt}\right) \tag{2}$$

Equation *(2)* says that the log of productivity in region i, $\ln(Q_{it})$, depends on the log level of the own stock of regional R&D $\ln(A_{it})$ and on the log level of the stock of external accessible R&D $\ln(\sum_{j\neq i} \phi_{ji}A_{jt})$. The above parametrization allows us to draw a clear distinction between knowledge flows and their effects. The parameters ϕ_{ji} capture the intensity of knowledge flows which could depend on several bilateral characteristics of the regions, their technological differences, their distance and so on. The parameter μ, on the other hand, captures the impact of "accessible external research" on technology also called R&D spillovers. These two parameters are conceptually and empirically very different and separating them would be important for our understanding of the knowledge-productivity link as well as for our ability of prescribing policy implications. For instance, finding a small effect of research in region j on productivity of region i could be due to little knowledge flows between the two regions (small ϕ_{ji}) or to a small impact of accessible external knowledge on productivity in region i (small μ). In the first case removing hurdles of communication between the two regions would result in higher innovative output of region i, in the second case it would not. We can use this simple frame for a selective, but still useful, review of the literature. We organize the review distinguishing between studies in the micro-productivity tradition that focus on firms and studies in the trade-growth tradition that focus on international R&D spillovers.

3. - Review of the Literature

3.1 *The Micro-Productivity Literature*

A little more than a decade ago Griliches (1992) made the point and set the agenda of "the search for R&D spillovers". Several pieces of empirical research followed that seminal paper and improved our understanding of the process of knowledge diffusion and of R&D spillovers. In actuality, the micro-productivity studies were simply the continuation and the refinements of an empirical tradition that had analyzed R&D spillovers for a long time[1]. We consider here mostly work produced during the last ten years[2].

A first simple method used to proxy knowledge flows across

[1] BRESNAHAN T.F. (1986); GRILICHES Z. - LICHTENBERG F. (1984); MANSFIELD E. *et* AL. (1977); SCHERER F.M. (1984); TERLECKY N. (1980); WOLF E.N. - NADIRI M.I. (1987) are some notable examples of earlier studies.

[2] Far from being a complete survey the present overview of the literature is meant to give a sense of the large body of work existing on this topic. Excellent surveys of the literature in the proper sense exist (notably GRILICHES Z., 1992); (MOHNEN P., 1996); (BRANSTETTER L., 1998); (CINCERA M., 2001).

firms assumed that only firms within the same "technological group" (for instance two or three digit SIC sector) experience knowledge flows between each other. Using the notation developed in the previous section we can summarize this assumption as restricting ϕ_{ji} to be equal to 1 for firms in the same group while ϕ_{ji} = 0 for firms in different groups. This approach was used, for instance, by Bernstein and Nadiri (1989*a*), (1989*b*) for the US high-tech industries and Bernstien and Mohen (1998) for US and Japan. Similar to this "discrete" form of weighting are also those methods that use geographical information to establish location of a firm within or outside a certain area. These studies impose ϕ_{ji} = 1 for firms in the same county, region or within a certain radius of distance and ϕ_{ji} = 0 outside that (see for instance Anselin *et* Al. 1997).

More sophisticated measures of knowledge flows define technological distance as a truly bilateral concept and allow ϕ_{ji} to be different for each pair of firms. Jaffe (1986) describes each firm as the vector of shares of R&D (or innovative activity) of the firm in each sector. The flow ϕ_{ji} is calculated as the uncentered correlation coefficient between the vector of firm i and the vector of firm j. Perfect coincidence in the sectors' shares results in a correlation of 1 between firm i and j while perfect complementarity generates a value of 0 of the correlation coefficient. Using a similar methodology Branstetter (2001) analyzes the impact of domestic and foreign R&D spillovers for US and Japanese firms. Still trying to proxy ϕ_{ji} with some technological distance other authors have used "flows" connecting firms (or sectors) i and j. Among these Wolf and Nadiri (1987) used input-output matrices, Terlecky (1980) used flows of intermediate capital goods and Scherer (1984) constructed a matrix of origin-use of patents. Recently Kaiser (2002) has tried to establish some comparisons among the above described methods. Once ϕ_{ji} have been used to construct A^a_{it} most of the articles analyze the impact of the accessible stock of knowledge on total factor productivity or on the innovation output of firms. There is a wide range of estimates but most of the studies find an elasticity to external accessible R&D (μ) between 50% and 100% of the elasticity to own R&D (γ).

Finally and notably in the most recent years, thanks to the availability of new data, from the US patent office and also from the European patent office, the parameters ϕ_{ji} have been estimated using patent citations. This method stands out because it is the only one in which "signs" of the presence of learning flows are actually observed in the data. Patent citations provide evidence on learning flows without making any *a-priori* assumption on their determinants (such as technological or geographical proximity). Using these data Jaffe *et* Al. (1992) test that distance matters for knowledge flows within the US, Jaffe and Trajtenberg (2002, Chapter 8 and 9), Adams (2002) and Jozefowicz (2002) compare knowledge flows originating in Universities, Federal Labs and firms, Globerman *et* Al. (2000) analyze knowledge flows for Swedish firms, Maruseth and Verspagen (2002) analyze knowledge flows across European regions and Jaffe and Trajtenberg (2002, Chapter 7) analyze knowledge flows across countries. While certainly more accurate and superior in estimating knowledge flows (ϕ_{ji}) those studies do not attempt to assess the impact of these flows on productivity.

3.2 *The Trade-Growth Literature*

A large part of the tradition in the trade-growth literature followed the practice of Coe and Helpman (1995) and measured ϕ_{ji} using trade (imports or exports) shares between country i and j. Several "improvements" to that paper followed. Keller (1998) raised some doubts on the methodology of the Coe and Helpman (1995) study, Edmond (2001) and Kao *et* Al. (1999) applied panel cointegration techniques to the analysis. Frantzen (2002) added human capital and some estimation improvements. Coe *et* Al. (1997) extended the analysis of R&D spillovers to seventy-seven developing countries and Madden *et* Al. (2001) to six Asian countries. Most of these studies confirmed the original findings of strong R&D externalities (μ as large as γ) especially from developed to developing economies. A natural extension to the use of trade is to use flows of foreign direct investments to proxy for knowledge flows. FDI's have long been consider as a mean of technological transfer and imply movement of capital and knowhow. Several studies such as Braconier and Sjoholm (1998) find that FDI's facilitate spillovers (within sectors across countries but not across sectors). Blomstrom and Kokko (1998) review the main contributions of this literature. A distinctive line of analysis pursued by Eaton and Kortum (1996) adopts a more complete and structural model of trade and growth across countries. They identify ϕ_{ji} using flows of cross-country patenting. In particular the share of inventions originated in country i and patented in country *j* is used to estimate ϕ_{ji}.

Finally, some recent studies on international R&D spillovers often do not use any information on flows in order to estimate ϕ_{ji} but they estimate it simultaneously with μ by exploiting the correlation structure of data on R&D, productivity and growth. Conley and Ligon (2002) analyze the correlation across long-term growth rates and find that it positively depends on "economic distance" while Keller (2002*a*) identifies ϕ_{ji} and μ contemporaneously by estimating the effect on *TFP* of domestic R&D and R&D from the G5 countries. Identification relies on the specified functional form and on the dependence of ϕ_{ji} on geographical distance. Again, the overall message from this literature is that μ is positive, its estimates however, vary widely.

4. - Regional Productivity

We assume a Cobb-Douglas production function at the regional level with constant return to scale. This implies that the elasticity of income to labor and to capital equal the respective shares of income going to each of the two factors. Calling α the share of income going to labor and $(1 - \alpha)$ the share going to capital, we assume, following the literature, that their value is equal to 0.66 and 0.33. As we use regional data we may rely on the within country variation to define total factor productivity, disregarding the (probably large) differences in total factor productivity across countries due to institutional and political features of the economies. The cross-regional differences in *TFP* can be more appropriately interpreted as differences in the level of technology used (adopted) there. The disadvantage is that we do not have data on physical capital at the regional level. We may nevertheless recover the contribution of regional *TFP* to production per worker in one of the following two ways. We either assume that physical capital moves across regions in order to equate its real return within a country (narrower definition) or we impose the same capital labor-*ratio* across regions of the same country (broader definition). The first method produces a lower bound of the region-specific *TFP* differentials, as it attributes part of the regional production per worker differences to capital intensity. The second method produces an upper bound of these differentials as it attributes the whole region-specific differentials in output per worker to differences in regional *TFP*. We perform both decompositions. To be precise, the first one distinguishes among the following three contributions to differentials in output per worker: 1) Differences in regional capital intensity (capital-labor *ratios*); 2) Differences in country-specific total factor productivity; 3) Differences in region-specific *TFP*. The second decomposition, on the other hand, distinguishes among the following contributions: 1) Differences in country-specific *TFP*; 2) Differences in country-specific capital-output *ratios*; 3) Differences in region-specific *TFP*.

4.1 *Narrow TFP Definition*

Let's assume that teach region *i* produces a perfectly tradable good, *Y*, according to the following production function:

$$Y_i = Q_c Q_i (L_i)^{\alpha} (K_i)^{1-\alpha} \tag{3}$$

where Y_i is total output of region *i*, L_i is total labor input, K_i total capital input, Q_c is the average *TFP* of country *c* to which the region belongs and Q_i is a region-specific *TFP* factor such that if $Q_i > 1$, region i has larger total factor productivity than the country as a whole while if $Q_i < 1$ then the region has lower *TFP* than the country as a whole. Also we assume that

$$\sum_{j \in c} \ln\left(Q_i\right) = 0$$

which implies that the average log *TFP* for regions of a country is equal to the aggregate country's log *TFP*. α is the elasticity of output to labor that equals the income share of labor. In order to decompose total income in the contribution of each component we assume that within country c the return paid to capital is the same in each region i due to the mobility of capital. This implies:

$$\frac{\partial Y_i}{\partial K_i} = \left(1-\alpha\right) Q_c Q_i \left(L_i\right)^{\alpha} \left(K_i\right)^{-\alpha} = r_c \tag{4}$$

Marginal productivity of capital at the country level is also equal to r_c. Denoting with $k_c = (K_c/L_c)$ the capital-labor *ratio* for the country. Equation *(4)* implies that we can write total capital in region i as:

$$K_i = k_c Q_i^{\frac{1}{\alpha}} L_i \tag{5}$$

Substituting equation *(5)* into the production function and rearranging we obtain the following expression for output per capita in region i:

$$\frac{Y_i}{L_i} = Q_c K_c^{1-\alpha} Q_i^{\frac{1}{\alpha}} \tag{6}$$

Defining output per worker in region i, as $y_i = Y_i/L_i$ and taking log of both sides of expression *(6)* we obtain a variation of the classic growth accounting equation:

$$\ln y_i - \left(1-\alpha\right) \ln k_c = \ln Q_c + \frac{1}{\alpha} \ln Q_i \tag{7}$$

The above notation is helpful to understand how to construct each single component of the equation. The left hand side is obtained as log GDP per worker in the region minus the share of capital times the national capital-labor *ratio*. To obtain the right hand side we can regress the left hand side on a set of country dummies (standardizing one to 0) and the residuals will measure the region-specific *TFP* term: $1/\alpha \ln Q_i$. Once we calculate these terms we can re-write the above equation in the following way, that captures exactly the contribution to regional productivity given by regional capital intensity, national *TFP* and region-specific *TFP*:

$$(8) \qquad \ln y_i = \underbrace{\left[(1-\alpha)\ln k_c + \frac{1-\alpha}{\alpha}\ln Q_i\right]}_{\text{Regional capital contribution}} + \underbrace{\ln Q_c}_{\text{Country } TFP} + \underbrace{\ln Q_i}_{\text{Narrow regional } TFP}$$

We call this measure of regional *TFP* the "narrow" *TFP* definition.

4.2 *Broad TFP Definition*

The second method assumes simply that $k_c = k_i$, i.e. that the capital labor *ratio* in each region is equal to the average capital labor-ratio of a country. Under this condition, we simply divide both sides of equation *(3)* by L_i, we substitute the condition above and take logs. The decomposition of output per worker results as follows:

$$(9) \qquad \ln y_i = \underbrace{\left[(1-\alpha)\ln k_c\right]}_{\text{Country capital contribution}} + \underbrace{\ln Q_c}_{\text{Country } TFP} + \underbrace{\ln Q_i}_{\text{Broad regional } TFP}$$

Note the similarity of the above decomposition with equation *(8)*. Now, as k_c is assumed constant within countries, the differentials left after controlling for country-specific factors, are attributed completely to regional *TFP* differentials. Compared to the previous method, it is as if the direct effect of *TFP* and its indirect effect, caused by larger physical capital in the region, are merged together. As *TFP* differences are the cause for the existence of both terms in the decomposition *(8)*, it makes sense to consider also this second method. We call this measure of regional *TFP* the "broad" *TFP* definition. In the implementation of both of these accounting equations we standardize to one the average output per worker, capital per worker and the average national *TFP* in the US.

5. - Knowledge Flows and Patent Citations

5.1 *Identification and Data*

We indicate the probability that a non-obsolete[3] idea generated in region i at time t_0 is learned in region j by time $t_1 = t_0 + \tau$ as $\phi_{ij}(\tau)$. This notation emphasizes the fact that such probability depends on characteristics of the couple of regions i, j, and on τ, the time elapsed between the invention and the act of learning. If there is a large number of ideas created in a region then, for the law of large numbers, $\phi_{ij}(\tau)$ is the range of ideas learned in region j out of those generated in region i within interval τ since their invention. Inspired by what is done in the "micro-productivity" literature, in particular by Jaffe and Trajtenberg (2002, chapter 6 and 7) and by Caballero and Jaffe (1993), we model the share $\phi_{ij}(\tau)$ as follows:

$$\phi_{ij}(\tau) = xe^{f(i,j)} \left(1 - e^{-\beta\tau}\right) \tag{10}$$

The term $1–e^{-\beta\tau}$ captures the fact that ideas generated in region i become available in larger share to any other location j as time passes. If the event of learning an idea happens with a

[3] Obsloescence of knowledge is incorporated in the depreciation used in calculating the stock of R&D of each region.

constant probability over time then this term captures the cumulative density of probability of learning the idea within τ years. However, the term $e^{f(i,j)}$ depends on a whole set of bilateral characteristics, indicated that the intensity of learning depends on the sending region i as well as on the receiving region j. The main simplifying assumption embedded in *(10)* is that the effect of bilateral characteristics $f(i, j)$ and the effect of time τ interact in a multiplicative way in determining knowledge diffusion. This means that, as time passes, more ideas that originated in region i are learned in any region, including itself, but such an increase is proportionally the same for any region so that the relative absorption of ideas originated from region i is constant over time. In our empirical analysis we experiment with different time intervals between generated and learned ideas τ = 2, 4, 6 and 10.

We fix the same interval of time τ for all regions so that bilateral flows can be written as ϕ_{ij}. On the other hand we can proxy the relative intensity of knowledge flows across regions with the relative citation frequencies. There is a close relationship between learning of ideas (knowledge flows) and patent citations. Patent applicants are required to identify the "prior art" used in order to produce their innovative ideas. They do so by including citations to previous patents that had some relevance in developing the idea. These citations inform us that the researcher knew about the cited idea and that such idea had some relevance in the research process leading to the new discovery. For our purposes if we had only the citations included by the authors of the patent we would have the best information available to establish the existence of knowledge flows[4]. What intro-

[4] JAFFE *et* AL. (1992) argue that citations not only establish a "learning" relation but also that they are limited to those ideas that had strict relevance to the development of the current innovation. This is because the inventors do not want to include irrelevant citations which would be dropped by the reviewer or would excessively restrict their claims on the use of the patent. For our purposes this restriction is not crucial. We only care that patents establish a learning relation between citing and cited idea. The fact that the cited idea was strictly relevant or not is not so crucial to us as we do not assume (but we estimate later) that knowledge flows have positive effect in generating new ideas.

duces noise for our use of citations is the fact that reviewers added citations to the patent. These added citations do not necessarily reveal ideas known to the author. Jaffe *et* Al. (1992) argue that the reviewers are expert in the area and they do a systematic search in the field so that these "added" citations should not have any (or much less of a) geographical pattern. We assume that they simply add noise to the relation between knowledge flows and patent citations. A survey study (Jaffe and Trajtenberg, 2002, chapter 12) confirms that while citations are not a perfect measure of the inventors' learned knowledge they contain a large amount of information about it. We rely on the extremely large amount of citation couples used (about 4.5 millions in total implying an average of about two hundreds citations for each regional couple) to reduce the random noise. Defining as c_{ij} the count of citations from patent in region i to patents originated in region j and as P_i and P_j the number of patents (ideas) originated in region i and j respectively we can proxy the unobservable relative flows of knowledge ϕ_{ij} with the observable relative citation frequencies $(c_{ij}/P_j)/(c_{ii}/P_i)$. In assigning a patent to a region, we choose the region of residence of its first inventor. This method, as documented by Jaffe *et* Al. (1992), allows us to locate each patent in the region where the idea was actually developed by its inventor(*s*) rather than in the region where the paperwork for the filing procedure was prepared (headquarters of the assignee company). The regions considered in our analysis are sub-national territorial units in eighteen countries in Europe and North America. They correspond to areas with some territorial unity and identity as well as administrative and policy autonomy. They are fifty federal states plus DC, Puerto Rico, Guam and Virgin Islands for the US, ten federal provinces plus Yukon and Northwestern Territories for Canada and the so-called "NUTS 1[5]" regions within each of sixteen European countries (EU15 plus Switzerland) for a total of 147 regions covering the whole Western European and North American continents. Our measure of relative knowledge flows imposes the standardization of f_{ii} = 1.

[5] Nomenclature Units Territorial Statistics, level 1.

We explicitly allow the function $e^{f(i,j)}$ to depend on bilateral characteristics and we estimate the following regression:

$$(11) \qquad (c_{ij}/P_j)/(c_{ii}/P_i) = \exp[\vartheta_i + \varphi_j + b_1(out\text{–}region)_{ij} + \\ + b_2(out\ next)_{ij} + b_3(out\ country)_{ij} + b_4(out\ lang)_{ij} + \\ + b_5(out\ trbl)_{ij} + b_6(dist)_{ij} + \gamma(Controls)_{ij} + \varepsilon_{ij}]$$

Equation *(11)* states that the (time-invariant) relative intensity of knowledge flows from region i to region j depends on an exponential function of several bilateral regional characteristics. We include citing region fixed effects φ_i as well as cited region fixed effects ϑ_i. These effects control for any region-specific characteristic affecting creation and reception of knowledge. We also explicitly consider six geographic characteristics which we want to analyze in detail, while the others, concerning technological and productive characteristics of the regions are bundled in the vector of *Controls* and will be considered explicitly in the empirical sections. The bilateral characteristics considered here as determinants of the intensity of learning from i to j are mostly dummies. $(out\_region)_{ij}$ is a dummy which equals zero if $i = j$ and one otherwise and indicates whether ideas crossed one regional border. $(out\_next)_{ij}$ is equal to zero if $i = j$ or if region i and j share a border and 1 otherwise, it indicates whether ideas crossed two regional borders. $(out\_country)_{ij}$ is zero if the two regions belong to the same country and zero otherwise, it indicates whether ideas passed a national border. $(out\_lang)_{ij}$ is zero if the two regions speak the same language and 1 otherwise. It indicates whether ideas passed a linguistic border. $(out\_trbl)_{ij}$ is one if the two regions belong to the same trade block and one otherwise. It indicates whether ideas passed a trading-block border. Finally $(dist)_{ij}$ is simply the geographical distance between region i and region j. Estimates of the parameters b_1-b_6 and of γ would provide a detailed characterization of how geographic, technological and productive characteristics affect the flows of ideas across regions. The term ε_{ij} is a zero-mean random error.

Regression *(11)* is familiar to the micro-productivity literature and is often estimated using a non linear least squares regression (e.g. Jaffe and Trajtenberg, 2002, chapter 7) or, more frequently, due to the count-data nature of citations, using a negative binomial regression (Branstetter, 2001) or, given the mass of observation at 0, using a Tobit regression (e.g. Maruseth and Verspagen, 2002).

On the other hand if we take logs on both sides of *(11)* we obtain a linear regression. Such regression is reminiscent of one that is heavily used in the trade-growth literature, mainly to analyze trade and is known as "gravity equation". In such equation a flow (of knowledge in this case) between region i and region j is regressed on "sending regions" and "receiving regions" characteristics and on a measure of distance between them as well as some other bilateral characteristics (such as belonging to the same country or sharing a border). Our specification is the most general form of a gravity equation as we control very generally for any sending and receiving regional fixed effect and we estimate parameters relative to the crossing of several geographical borders and relative to traveling geographic and technological distances. Typically, the trade literature estimates such equation using a linear regression and omits (as logs are taken on both sides) the couple of regions for which a zero trade link is present. In Section 5 we estimate several variations of equation *(11)* using all the methods mentioned above. Luckily different estimation methods give very similar coefficient estimates.

The patent and citation data used are from the *NBER Patent and Citation Dataset*, which is publicly available and described in detail in Jaffe and Trajtenberg (2002, chapter 13). This dataset contains all the patents granted by the US patent office and, since 1975, all the citations made by each patent. It includes information on the technological class of the patent and several data on the applicants and inventors. We choose the sample of patents granted between 1975 and 1996 whose inventor is resident of one of the eighteen countries considered and listed in the appendix of Peri (2003). From the address of the first inventor we assigned patents to sub-national regions. Our final sample contains about 1.5 million patents and about 4.5 million citation couples, distributed across 147 regions. We use all the bilateral relationships among the 147 regions (total of 21,609 pairs some of which with 0 citations) when we estimate the "gravity-like" equation *(11)*. Table 1 reports some summary statistics at the regional level. It shows average and standard deviation for the number of patents granted each year to residents of the 147 regions. The average region had 426 patents granted per year (clearly large variation over time is hidden in this table) but very large disparities across regions exist. The least innovative region was granted a patent every four years (0.27 yearly) and the most innovative was granted 6,434 patents per year. The top innovator, with a very large lead on the second region, is California, that was granted more than 6,000 patents per year. High in the ranking are also some German, French and British regions (mostly the regions corresponding to large cities such as London, Paris and Hannover). They all have one thousand or more patents granted each year. The bottom of the list is taken by Greek, Spanish and East German regions that are granted one or less than one patent per year.

TABLE 1

DESCRIPTIVE STATISTICS RELATIVE TO 147 REGIONS IN EUROPE AND NORTH AMERICA

Variable	Mean	Std. deviation	Min	Max
Number of average yearly granted patents 1975-1996	426	830	0.27	6.434
Share of GDP spent in R&D average 1975-1996 %	1.77	1.23	0.27	7.69
Number of total region to region citations without self, *cij*	171	1147	0	99.137
Geographical distance (in thousands of km)	4.44	3.22	0	13.70

5.2 *Estimation Results*

5.2.1 Aggregate Flows, Geographical Determinants

We present in this Section the results of estimating the basic specification *(11)*. These estimates provide a measure of aggregate knowledge flows which could depend also on the sector- composition of regional ideas. We devote the following two sections to a detailed treatment of technological distance and of differences across technological categories. Specification I in Table 2 is the baseline regression for this section. We estimate equation *(11)* taking logs of both sides and using OLS with 147 citing-region and 147 cited-region fixed effects and we report the heteroskedasticity robust standard errors. The dependent variable is the log of the count of citation links, omitting self-citations[6], between patents of region i and patents of region j generated within the first 10 years since the cited patent is granted. We are confident that this time-span is long enough to capture the most relevant part of knowledge diffusion. If an idea has not been learned in ten years it is likely that it will not be very useful for innovation. However, we analyze flows also after 2, 6 years and after the longest available period in our sample (more on this below). As some regional couples have no citations, we simply drop those observations. This is why of the 21,609 possible couples (147 by 147) the first specification is only estimated on 15,839. The equation and the estimation method are akin to what the trade literature calls a "generalized" gravity equation used for trade flows. We choose this as basic specification for its simplicity and for the comparability of the coefficients to those obtained by the trade literature. Each coefficient captures the drop in knowledge flows as we move out of the region of origin and as we pass several borders. For instance the first coefficient

[6] Self-citations are citations between patents assigned to the same institution. Those citations denote, arguably, knowledge flows, but probably should not be included in the analysis of pure R&D externalities. Companies may reward their inventors for citing each other and for knowing about each other work. We estimated specifications including self-citations and the only difference is that the coeffcient on "Crossing Region Border" is increased by roughly 10-15%.

says that in moving out of the region of origin average knowledge flows drop to ($e^{-1.91}$) = 0.15 of their initial level. Another way of saying it is that 85% of knowledge generated in the average region is not learned outside it but remains local. The second coefficient says that only ($e^{-0.43}$) = 65% of the 15% of knowledge flowing out of the regional border passes the next regional border. Only 9.75% (=15%*65%) of the initial knowledge, that is, flows outside the regions that share a border with the original one. Another 20% (= 1 – $e^{-0.19}$) is lost passing the country-border leaving about 8% of the

TABLE 2

GEOGRAPHICAL DETERMINANTS OF AVERAGE KNOWLEDGE FLOWS[a]

Specification	I Baseline: within 10 years	II Negative binomial within 10 years	III Tobit within 10 years	IV Same 3-digit within
Crossing region border	-1.91* (0.07)	-2.05* (0.04)	-1.98* (0.06)	-1.91* (0.02)
Crossing next-region border	-0.43* (0.02)	-0.45* (0.02)	-0.45* (0.03)	-0.44* (0.03)
Crossing country border	-0.19* (0.02)	-0.18* (0.01)	-0.19* (0.02)	-0.19* (0.02)
Crossing trade-block border	0.05 (0.03)	0.06* (0.03)	0.05 (0.03)	0.05 (0.03)
Crossing linguistic border	-0.19* (0.01)	-0.20* (0.01)	-0.19* (0.02)	-0.17* (0.02)
1000 Km further	-0.05* (0.002)	-0.05* (0.001)	-0.05* (0.002)	-0.05* (0.002)
Citing region fixed effects	yes	yes	yes	yes
Cited region fixed effects	yes	yes	yes	yes
Observations	15,378	21,609	21,609	14,395
Log-likelihood		-65,584.92		
R^2	0.91	na	na	0.89
Original total number of citations	2,864,298	2,864,298	2,864,298	1,589,958

[a] Citations are calculated omitting self-citations, i.e. citation within the same institution. Heteroskedasticity-robust standard errors in parenthesis. *= significant at 1% confidence level.

TABLE 2 *(cont)*

GEOGRAPHICAL DETERMINANTS OF AVERAGE KNOWLEDGE FLOWS[a]

Specification	V Within 2 years	VI Negative binomial within 2 years	VII Within 6 years	VIII All couples
Crossing region border	1.80* (0.05)	-2.05* (0.05)	-1.91* (0.07)	-1.90* (0.07)
Crossing next-Region border	-0.37* (0.03)	-0.40* (0.03)	-0.42* (0.02)	-0.43* (0.02)
Crossing country border	-0.21* (0.02)	-0.18* (0.02)	0.20* (0.02)	-0.20* (0.02)
Crossing trade-block border	0.06 (0.03)	0.05 (0.03)	0.06 (0.03)	0.04 (0.03)
Crossing linguistic border	-0.11* (0.02)	-0.20* (0.02)	-0.18* (0.01)	-0.17* (0.01)
1000 Km farther	-0.05* (0.002)	-0.05* (0.002)	-0.05* (0.002)	-0.05* (0.002)
Citing region fixed effects	yes	yes	yes	yes
Cited region fixed effects	yes	yes	yes	yes
Observations	12,807	21,609	14,019	15,839
Log-likelihood				
R^2	0.86	na	0.89	0.92
Original total number of citations	528,829	528,829	1,977,435	4,710,215

initial knowledge. Crossing a trade block border has basically no effect, while passing a linguistic border further reduces the flow by 19%. On top of these effects, geographical distance adds a 5% decrease for each 1,000 kilometers traveled. Each coefficient is very precisely estimated, they are all very significantly negative (except for the effect of crossing a trade block border that is essentially zero) and extremely robust across specifications. The estimated drop in learning as consequence of geography is quite substantial. On the other hand, by far the most drastic drop takes place as we move out of the region itself, proving the very large local component of learning.

In order to gain confidence that the count-data nature of citations and the relatively large number of zeroes do not distort our linear estimates we use in column II and III the techniques that handle these issues explicitly. In column II we report estimates of equation *(11)* in levels and using a negative binomial regression. The advantage of this method is that we include the zeroes and that, by assuming a generalized Poisson process as generating the data, we account for the fact that citations are "count data". The method used to estimate this model is maximum likelihood. Column III uses a Tobit regression. In particular, as there is a large mass of data at 0, we assume that log flows have a linear dependence on their geographical determinants but, for observation smaller than 0, we observe the variable truncated at 0. This specification is estimated using maximum likelihood. Column II and III of Table 2 show that these two methods of estimation deliver coefficient estimates almost identical to the simple log linear regression. In particular all coefficients are literally identical up to a 2% difference except for the first one (regional border effect) that is slightly higher in absolute value when estimated with the negative binomial regression (–2.1) or with the Tobit (–1.98). Even for this coefficient, however, both estimates are within two standard deviations of the linear one, and quantitatively they make very little difference[7]. We perform negative binomial estimates of our coefficients throughout the paper and when they are significantly different from OLS, due to the treatment of the zero observations, we report and prefer them. However it is normally the case that these estimates are rather similar to the OLS basic specification in which case we report only the OLS estimates.

Column IV in Table 2 investigates whether flows within a sector of technological innovation are more or less localized than flows across them. In this specification we select only citation links within the same 3-digit class (in the *International Patent Classification* code). These classes are rather specific, and

[7] The coeffcient estimate using the Negative binomial would imply 13% and the Tobit 14% of regional knowledge learned outside the region. This as opposed to 15% estimated using the linear regression.

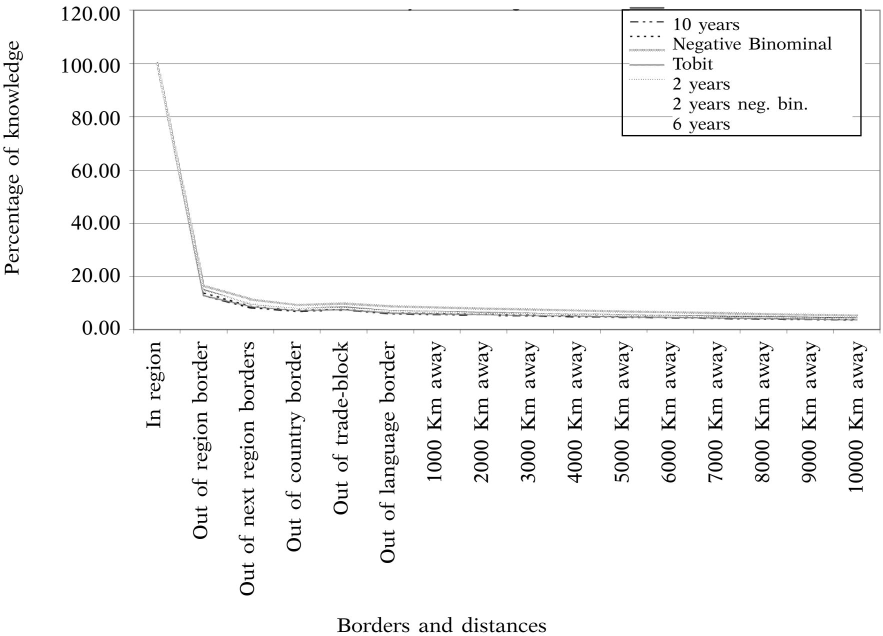
there are about 400 of them[8]. We may think that diffusion of knowledge within a narrow field is farther reaching than diffusion across fields. Estimates of column IV are very similar to the baseline, providing evidence that diffusion of knowledge across fields does not exhibit significantly different localization pattern from diffusion within fields. Such feature was already pointed out by Jaffe, Trajtenberg and Henderson (see Jaffe and Trajtenberg, 2002 p. 175) when they found that within class citations do not have more tendency to be co-located than across-class citations.

Column V to VIII explore the robustness of our estimates when we allow shorter or longer interval of time between the citing and the cited patents. Column V and VI include citations within the first 2 years, column VII within the first 6 years, and column VIII all citations couples in the 1975-1996 period so that ideas generated early in the period include learning up to 20 years from their invention. Column VII and VIII show estimates basically identical to column I. Only for the 2-year interval there is some difference, which is probably driven by the larger number of zeroes omitted in the OLS regression, as the negative binomial regression is extremely similar to the 10-year one. In any case even the OLS estimates do not exhibit any important difference with the basic 10-year case, and certainly not stronger localization for the 2-year interval. Knowledge flows maintain their relative spatial distribution as time elapses, so that while knowledge of an invention becomes more available over time it does not become relatively more available far away than it is in the region. The pattern of regional diffusion within 2 years is pretty much representative of the overall pattern allowing for longer delays. The way we chose to model space and time diffusion keeping them multiplicatively interactive seems reasonably good to analyze our data. This is very fortunate as we can focus here on geographic diffusion without risking to have a very different analysis depending on the lag that we consider.

[8] Some examples of these classes are "robots" or "distillation: apparatus" or "batteries".

GRAPH 1

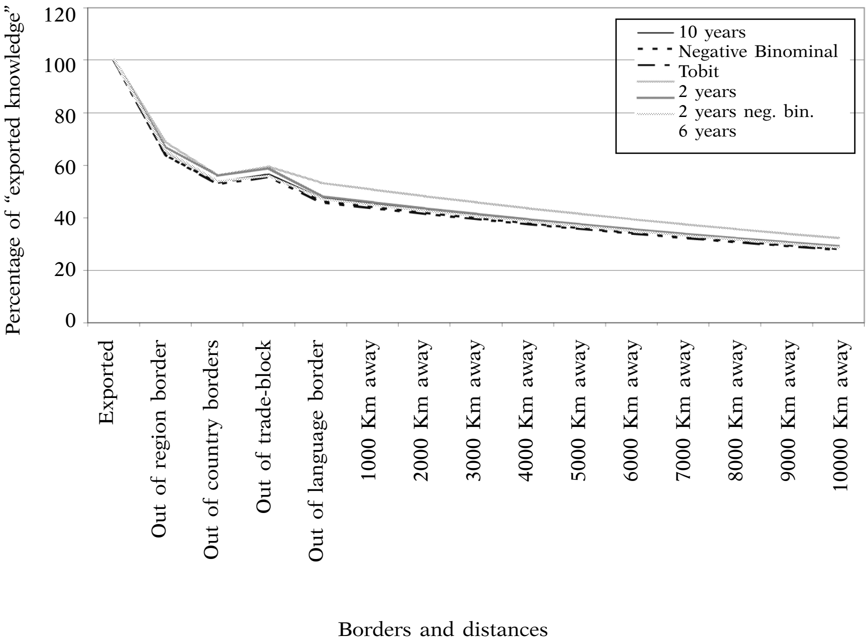
DECAY OF KNOWLEDGE FLOWS



Before moving to further specifications it is useful to summarize the results of this section using a couple of pictures. Graphs 1 and 2 represent the estimated decay of knowledge flows as one moves from a region, out of it, out of its neighbor, out of the country, out of the linguistic area, out of the trade block and travels by steps of 1000 Km. In Graph 1 the total amount of knowledge generated in a region is standardized to 100. As we move from left to right the lines show the fall in knowledge flows as we pass borders and as we travel farther and farther. Six lines are reported and they correspond to the values obtained using estimates in column I, II, III, V, VI and VII of Table 2 respectively. What is clear is the predominance of the first drop (when moving out of the region) re-

GRAPH 2

DECAY OF "EXPORTED" KNOWLEDGE FLOWS



lative to all others and the extreme similarity of rate of decay estimated using any specification. In order to have a better visual sense of the further decay out of the region, in Graph 2 we simply consider only what we call "Exported Knowledge", i.e. knowledge flows once the own regional border has been crossed. We standardize that level to 100 and we track the decay from there on. Again we report six lines corresponding to the estimates I, II, III, V, VI and VII in Table 2. We can still appreciate the extreme similarity in patterns across different estimates. Now we see that a very significant percentage of the exported knowledge (about one half of it) flows all the way out of the trade block.

5.2.2 Aggregate Flows, Technological Determinants

The estimates of the previous section provide a very interesting characterization of the effect of geography and borders on average knowledge flows, once we have controlled for citing and cited region effects. However, some important bilateral determinants of knowledge flows are missing. In particular, some measure of distance in technological space capturing the difference in technological fields of specialization should certainly be included. As we noted above the trade-growth literature has focused on aggregate flows and productivity and has not paid much attention to the relevance of technological space. However there is a huge body of evidence from the micro-productivity literature analyzing this issue. In particular, as regions with similar level of technological specialization could be located close to each other, failing to control for "technological distance" may result in overestimating the effect of geography. Table 3 shows the estimation of the basic specifications with a measure of technological distance added as control.

The index of technological distance is constructed following Jaffe (1986) and had a very large use in the micro-productivity literature. Specifically all patents granted to a region (call it region *i*) are grouped into 36 technological classes. These classes constitute specific areas of research, defined following the international patent classification and reported in the appendix of Peri (2003). The shares of regional patents (1975-1996) generated by region i in each technological class s is calculated. A vector of shares Sh_i = $(sh_{i1}, sh_{i2}...sh_{i36})$ is then associated to each region. The uncentered correlation coefficient (or angular distance) between the vector of region i and j, calculated as $(TecCorr)_{ij} = (Sh'_i Sh_j)/ [\Sigma_s(sh_{is})^2 \Sigma_s(sh_{js})^2]^{1/2}$ is a measure of "similarity" in technological space. Its value is between 0 and 1 and it is closer to one the larger is the "overlap" in technological classes of specialization. For perfect overlap the index is 1, for no overlap at all the index is 0. We use $(TecDis)_{ij} = 1-(TecCorr)_{ij}$ as a control in Table 3.

The estimates of the effect of this variable is statistically and economically extremely significant. The OLS estimates for both

TABLE 3

ROBUSTNESS CHECKS[a]

Specification	I OLS within 10 years	II Neg. Bin. within 10 years	III OLS within 2 years	IV Neg. Bin. within 2 years
Crossing region border	-1.34* (0.06)	-1.50* (0.06)	-1.30* (0.05)	-1.45* (0.06)
Crossing next-region border	-0.32* (0.02)	-0.32* (0.02)	-0.29* (0.02)	-0.27* (0.02)
Crossing country border	-0.22* (0.02)	-0.20* (0.02)	-0.24* (0.02)	-0.19* (0.02)
Crossing trade-block border	0.05 (0.03)	0.04 (0.02)	0.05 (0.03)	0.04 (0.03)
Crossing linguistic border	-0.16* (0.02)	-0.18* (0.01)	-0.15* (0.02)	-0.17* (0.02)
1000 Km farther	-0.05* (0.002)	-0.04* (0.002)	-0.05* (0.002)	-0.05* (0.002)
Technological distance	-2.27* (0.06)	-2.86* (0.04)	-2.01* (0.07)	-3.10* (0.06)
Citing region fixed effects	yes	yes	yes	yes
Cited region fixed effects	yes	yes	yes	yes
Observations	15,361	21,609	14,065	21,609
Log likelihood		-56,555.16		-36,753.69
R^2	0.92	na	0.87	na

[a] Citations are calculated omitting self-citations, i.e. citation within the same institution. Heteroskedasticity-robust standard errors in parenthesis. *= significant at 1% confidence level.

the ten and two year delay specifications (Column I and III) produce similar results. The flow between two regions specialized in totally different areas is 87-90%[9] lower than the flow between two regions with identical technological specialization. As the standard deviation of $(TecDis)_{ij}$ is 0.17 increasing the difference in specialization by one standard deviation reduces learning by 31-33%. Even more dramatically, the negative Binomial estimates imply a decay of knowledge flows between 95 and 97% going from iden-

[9] $0.87 = (1 - e^{-2.01})$, $0.90 = (1 - e^{-2.27})$.

tical to completely different specialization. Moreover the inclusion of proxies of technological differences reduces the geographical effects. Particularly the effect of crossing the own region border and the next region border are reduced, respectively from 1.8-1.9 to 1.3-1.5 and from 0.4 to 0.3. About twenty percent of the previously estimated attrition in learning when moving out of the originating region and attributed to geographical factors is, in reality, the result of technological distance.

5.2.3 Sectors, Periods and Continents

While our focus is on aggregate knowledge flows, as technological specialization plays an important role in determining what flows a region gets, we analyze in greater detail here, the geographical behavior of flows dividing them in large technological sectors. Moreover, as we assumed stable behavior of these region to region flows over the years and across the two analyzed continents (Europe and North America) we explore here to what extent the data support such assumptions.

Table 4 reports the estimates of distance and crossing borders on knowledge flows within each of six sectors. As we are only analyzing flows within a sector we omit the controls for differences in technological specialization. We choose only patents and citations within each sector (within ten years from the originating patent) and we perform OLS estimation including citing and cited region effects and we report heteroskedasticity-robust standard errors. The sector estimates are reported in column I to VI. Interestingly, the negative effect of the first two dummies (crossing regional border and crossing next region border) on knowledge flows grows in absolute value moving from computers to "other sectors". The computer sector exhibits by far the largest geographical diffusion of knowledge. Close to 40% of computer-related knowledge generated in a region is learned outside of it and 25% of it flows all the way out of the country and linguistic area. In contrast the mechanical sector seems much more localized with only 18% of knowledge flowing out of the originating region and

Table 4

KNOWLEDGE FLOWS FOR SIX TECHNOLOGICAL SECTORS[a]

Specification	I Computers 10 years	II Drugs 10 years	III Electronics 10 years	IV Chemical 10 years	V Mechanical 10 years	VI Others 10 years
Crossing	-1.00*	-1.43*	-1.50*	-1.61*	-1.67*	-1.82*
region border	(0.07)	(0.06)	(0.07)	(0.06)	(0.05)	(0.06)
Crossing	-0.17*	-0.10*	-0.25*	-0.33*	-0.38*	-0.44*
next-region border	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Crossing	-0.16*	-0.21*	-0.21*	-0.12*	-0.13*	-0.20*
country border	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.02)
Crossing	0.06	0.01	0.06	0.05	0.04	0.06
trade-block border	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Crossing	-0.07*	-0.04*	-0.07*	-0.12*	-0.08*	-0.12*
linguistic border	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
1000 Km farther	-0.04*	-0.04*	-0.04*	-0.04*	-0.05*	-0.06*
	(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.02)
Citing region fixed effects	yes	yes	yes	yes	yes	yes
Cited region fixed effects	yes	yes	yes	yes	yes	yes
Observations	7,173	8,662	9,573	10,446	11,231	11,842
R^2	0.80	0.79	0.83	0.81	0.83	0.84
Original number of citations	243,563	243,902	333,637	342,572	356,614	486,513

[a] Citations are calculated omitting self-citations, i.e. citation within the same institution. Heteroskedasticity-robust standard errors in parenthesis. *= significant at 1% confidence level.

a slim 7% making it out of the country and linguistic area. Table 4 provides a representation of the "degree of globalization" of each sector. If we think that the sector "others" contains technological classes such as "agriculture", "apparel", "furniture" and "heating" we find that knowledge in "newer" technological fields, such as computers or biotech (contained in the category Drugs) reaches further than knowledge in more "traditional" technologies, such as mechanical or chemical. Interestingly almost all of the geographical hurdles seem to cause a stronger attrition as we move from computer to "others". Of the "exported knowledge", i.e. of that share of knowledge learned outside the region of origin, fully 50% of computer-related knowledge reaches regions as far as 10,000 Kilometers out of the country and linguistic area. To the contrary for knowledge in "other" sectors only 25% of the "exported knowledge" reaches 10,000 Km of distance outside the country and language area. While in the remainder of the work we analyze aggregate flows of knowledge the above discussed results make us aware that the technological composition of knowledge affects the geographical reach of its flows.

As for the geographical reach of knowledge flows across different decades or in different continents (Europe versus North America) Table 5 provides us some reassurance that the assumption of stability of coefficients is reasonably good. All estimates use maximum likelihood negative binomial method because the handling of zeroes seems to make some difference in this case. Column I and II of Table 5 show estimates for knowledge flows (within 2 years) for the 1975-1986 period and for the 1986-1996 period, respectively. The only coeffcients that are somewhat different across the two decades are the effect of crossing the own regional border (–1.33 versus –1.45) and the effect of crossing the country border (–0.12 versus –0.20). Let's remind the reader that the estimates in each sub-period, using fewer observations than the overall estimates, are less precise. Given that differences are not very significant and are in the direction of slightly larger localization of knowledge flows in the later period (while we would expect, if anything the opposite), we interpret the differences as due to noise and we confirm our assumption of basically identical effects. Column III and IV of Table 5 report the estimates of the effects of geographical characteristics on knowledge flows of the computer sector, also splitting the period between 1975-1986 and 1986-1996. The computer sector has been the one whose share of innovation has increased most in this period. The reader may be worried that if the geographical reach of knowledge flows for this sector has changed, this could affect the perspectives of knowledge flows and their future behavior. Although some small differences exist, there is no clear pattern of stronger localization in the earlier period. Even for this sector our simplifying assumption of similar geographical diffusion before and after 1986 seems reasonable. Finally column V and VI compare the impact of geographical characteristics on knowledge flows in Europe and

Table 5

KNOWLEDGE FLOWS IN DIFFERENT DECADES/CONTINENTS[a]

Specification	I Average flows 2 years lag 1975-1986 negative binomial	II Average flows 2 years lag 1986-1996 negative binomial	III Computers 2 years lag 1975-1986 negative binomial 10 years	IV Computers 2 years lag 1986-1996 negative binomial	V Average sector 10 years lag Europe negative binomial	VI Average sector North America 10 years negative binomial
Crossing	-1.33*	-1.49*	-0.85*	-0.94*	-1.50*	-1.30*
region border	(0.10)	(0.10)	(0.09)	(0.08)	(0.10)	(0.13)
Crossing	-0.28*	-0.26*	-0.18*	-0.07*	-0.26*	-0.23*
next-region border	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)
Crossing	-0.12*	-0.20*	-0.06	-0.14*	-0.30*	-0.41*
country border	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
Crossing trade	-0.07	0.05	0.04	0.04		
block rorder	(0.04)	(0.03)	(0.04)	(0.04)		
Crossing	-0.20*	-0.19*	-0.06	-0.05	-0.18*	-0.08
linguistic border	(0.02)	(0.02)	(0.04)	(0.04)	(0.02)	(0.07)
1000 Km farther	-0.05*	-0.05*	0.04*	-0.05*	-0.04*	-0.05*
	(0.003)	(0.002)	(0.005)	(0.005)	(0.01)	(0.01)
R&D difference	-0.12*	-0.09*			-0.04*	-0.09*
	(0.01)	(0.01)			(0.02)	(0.02)
Technological distance	-3.3*	-3.2*			-2.67*	-4.01*
	(0.11)	(0.10)			(0.06)	(0.14)
Citing region fixed effects	yes	yes	yes	yes	yes	yes
Cited region fixed effects	yes	yes	yes	yes	yes	yes
Observations	19,845	19,845	4062	4350	16,709.44	3,798
Log likelihood	-23,006.902	-27,615.223	-24,023.902	-23,615.253	-44,318.62	-13,798.08

[a] Citations are calculated omitting self-citations, i.e. citation within the same institution. Heteroskedasticity-robust standard errors in parenthesis. *= significant at 1% confidence level.

North America. As probably expected, there is a slightly stronger localization in Europe. Moving out of the region and its neighbors reduces learning flow by 83% of their initial value relative to a reduction of 79% for north American regions. However these differences are small, and it appears that linguistic borders play more of a role in Europe than in North America[10]. On the other

[10] Notice that the Language Effect for North America is very imprecisely estimated as it is based on two regions only (Quebeck and Puerto Rico) that do not speak english.

hand the effect of "technological distance" and of crossing a country border on learning flows appears larger for North American regions.

6. - Impact on Productivity

6.1 *Measures of Regional Productivity*

The source of data used to perform regional accounting are described in detail in the Appendix of Peri (2003). In general, data on gross regional product and total regional employment are from the region dataset for the EU countries, from the national statistical offices for Norway and Switzerland, from the Bureau of Economic Analysis for US states and from Cansim for Canadian Provinces. For each region we take the average value for the 1991-1996 period, in order to smooth any effect from business cycle fluctuations. The capital-labor ratio at the country level is taken from Hall and Jones (1999) and is relative to year 1988. The share of labor is assumed to be the same across regions and equal to 0.66. This Section documents that variation of region specific total factor productivity explains a substantial part of the differences in regional labor productivity. Differences across regions within countries are as important as differences across countries in explaining productivity differentials in Europe-North America. The cross-regional variation of $\ln(y_i)$, decomposed in a cross-country component and a cross-region, within-country component reveals that 48% of the total variance is explained by country-variation and 52% by regional variation.

We then perform the accounting exercise described in section 4 where we standardize the average output per worker, the average capital per worker and the average country *TFP* of the US to be equal to 1 (therefore their log is 0). We rank regions according to their output per worker (average current $ 1991-1996) from the most productive (Bruxelles) to the least (Portugal). The contribution of each component (capital-labor *ratio*, country *TFP* and region *TFP*) to the differences in output per worker for each region, using the "narrow" and the "broad" regional *TFP* definition can be requested from the author. In Table 6 *(a)-(b)* we simply report the decomposition into components of the differences in productivity between the Top and the bottom regions. On average between 17% and 48% of the differences in output of regions in the top10 group versus regions in the bottom 10 group are explained by regional *TFP* when we use the narrow definition. The percentage increases to 24-73% when we use the broad definition. Differences in regional factor productivity are very relevant in explaining differences in output per worker. Considering them as a measure of technological advancement across regions we now analyze the impact of R&D and of knowledge spillovers on them.

TABLE 6

CROSS-REGION ACCOUNTING ON OUTPUT PER WORKER
(1991-1996 AVERAGES, 141 REGIONS)

	$\Delta\ln(k_{regio})$	$\Delta\ln(TFP_{country})$	$\Delta\ln(TFP_{regio})$	$\Delta\ln(y)$	% explained by $\Delta\ln(TFP_{regio})$
	a) Narrow definition of regional *TFP*				
Top - Bottom	0.42	0.90	0.25	1.57	17
Top2 - Bottom2	0.38	0.52	0.43	1.34	32
Top5 - Bottom5	0.54	0.01	0.51	1.05	48
Top10- Bottom10	0.37	0.10	0.25	0.72	35

	$\Delta\ln(k_{country})$	$\Delta\ln(TFP_{country})$	$\Delta\ln(TFP_{regio})$	$\Delta\ln(y)$	% explained by $\Delta\ln(TFP_{regio})$
	b) Broad definition of regional *TFP*				
Top - Bottom	0.30	0.90	0.38	1.57	24
Top2 - Bottom2	0.17	0.52	0.66	1.34	49
Top5 - Bottom5	0.28	0.01	0.77	1.06	73
Top10- Bottom10	0.24	0.10	0.38	0.72	52

6.2 *Estimates of the Impact of Knowledge Flows*

Table 7 reports the estimates of equation *(2)*. We use the basic coefficient estimates from equation *(11)* from Table 2, Column I to calculate for each couple of regions, and then we construct for each region the "accessible" stock of R&D $A^a_{it} = \Sigma_{j \neq i} \phi_{ji} A_{jt}$. The estimates reported in Table 7 are those for the parameters γ and μ using a cross-section (average 1991-1996 *TFP*) of 141 regions. Column I uses the "narrow *TFP*" definition as dependent variable, while column II uses the "broad" definition. First of all, the elasticity of regional *TFP* to local R&D resources spent is between 0.067 and 0.10 and is estimated rather precisely in both specifications. The coefficient is magnified when we consider the effect of R&D on broadly defined regional *TFP* (column II) rather than on narrowly defined regional *TFP* (column I). These estimates can be compared with those in Coe and Helpman (1995) and with those in Keller (2002*b*). Both works estimate this elasticity using the time-series variation of R&D and total factor productivity at the country level. Coe and Helpman using the long-run cointegrating relation find a number between 0.078 and 0.097, very close to the effect we estimate on broad *TFP*. As Coe and Helpman (1995) use the long-run cointegrating relation among variables we think they are using the same information as we are. Keller (2002*b*), who estimates this elasticity using sector-country yearly data finds a

TABLE 7

IMPACT OF KNOWLEDGE FLOWS ON PRODUCTIVITY[a]

Dep. var: ln (Q_i)	I	II
ln(A_{it}), Own R&D	0.067*	0.10*
	(0.01)	(0.01)
ln(A^a_{it}), External Accessible R&D	0.03	0.05
	(0.03)	(0.04)
Observations	141	141
R^2	0.23	0.23

[a] Heteroskedasticity-robust standard errors in parenthesis. *= significant at 1% confidence level.

value somewhat lower between 0.04 and 0.055. This can be due to the shorter time horizon (yearly data) of his analysis. The effect of external available R&D, μ, is also estimated to be positive, and about 50% of the own R&D effect (0.03-0.05). However the standard error is large and we cannot conclusively rule out zero effect. There are several factors that may concur to finding a small effect of external R&D on productivity. First, diffusion of disembodied knowledge may be more important for innovation than for production. Peri (2003) shows that the impact of these flows is positive on innovation. Second, effects on production may also need transfers of embodied knowledge via trade, or via direct investments. Third, the diffusion of knowledge may help productivity but may also crowd out local knowledge having an overall ambiguous effect. Fourth, total factor productivity may be a partial and imperfect measure of technological advancement being based on residuals and not on positive observations. From our analysis external R&D appears to have a positive but small impact on productivity. Further research and better data on regional capital, however, may be needed in order to measure more precisely that effect.

7. - Conclusions

The present study analyzes the process of knowledge diffusion as evidenced by data on patent citations and the effect of these flows on *TFP*. While diffusion of ideas is needed in order to have externalities of knowledge there is no reason to believe that simply by measuring the intensity and scope of this knowledge diffusion, we have a measure of knowledge externalities. Diffusion of new ideas, in fact, brings not only "new inspiration" to researchers but also introduces higher standards of 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. These flows, likely to be important for innovation and knowledge diffusion, as shown in Peri (2003) do not seem to bring large significant effects on productivity, at least as it is measured by regional *TFP*.

Several reasons may generate this weak result. On one side, as stated above, positive and negative effect of knowledge diffusion may in part balance each other, on the other *TFP* may not be the best measure of technological progress and more direct measures of innovation and R&D output could be more intensely affected by this type of knowledge flows.

Our study, however, indicate a new procedure, new data and an original approach to capture knowledge flows at the technological frontier. We are convinced that pursuing these line of research further we may learn much about generation of ideas, technology and its diffusion.

BIBLIOGRAPHY

ADAMS J., «Comparative Localization of Academic and Industrial Spillovers», *Journal of Economic Geography,* vol. 2, no. 3, 2002, pp. 253-278.

AGHION P. - HOWITT P., «A Model of Growth Through Creative Destruction», *Econometrica,* vol. 60, 1992, pp. 323-51.

ANSELIN L. - VARGA A. - ACS Z., «Local Geographic Spillovers between University Research and High Technology Innovations», *Journal of Urban Economics*, vol. 42, no. 3, 1997, pp. 422-48.

BERNSTEIN J.I. - MOHEN P., «International R&D Spillovers between U.S. and Japanese R&D Intensive Sectors», *Journal of International Economics*, vol. 44, no. 2, 1998, pp. 315-38.

BERNSTEIN J.I. - NADIRI M.I., «Interindustry R&D Spillovers, Rate of Return and Production in High Tech Industries», *American Economic Review*, *Papers and Proceedings,* vol. 78, no. 2, 1998*a*, pp. 429-34.

— — - — —, «Research and Development and Intra-Industry Spillovers: An Empirical Application of Dynamic Duality», *Review of Economic Studies*, vol. 56, 1989*b*, pp. 249-69.

BLOMSTROM M. - KOKKO A., «Multinational Corporations and Spillovers», *Journal of Economic Surveys*, vol. 12, no. 3, 1998, pp. 247-77.

BRACONIER H. - SJOHOLM F., «National and International Spillovers from R&D: Comparing a Neoclassical and Endogenous Growth Approach», *Weltwirtschaftliches-Archiv*, vol. 134, no. 4, 1998, pp. 638-63.

BRANSTETTER L., «Looking for International Knowledge Spillovers: A Review of the Literature with Suggestion for New Approaches», *Annales d'Economie et de Statistique*, vol. 0, no. 49-50, 1998, pp. 517-40.

— —, «Are Knowledge Spillovers International or Intranational in Scope? Microeconometric Evidence from the U.S. and Japan», *Journal of International Economics*, vol. 53, n. 1, 2001, pp. 53-79.

BRESNAHAN T.F., «Measuring Spillovers from Technical Advance», *American Economic Review*, vol. 76, 1986, pp. 741-55.

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

CINCERA M., «International R&D Spillovers: A Survey», *Cahiers Economiques de Bruxelles,* vol. 169, 2001, pp. 3-31.

COE D. - HELPMAN E., «International R&D Spillovers», *European Economic Review*, vol. 39, 1995, pp. 859-87.

COE D. - HELPMAN E. - HOFFMEISTER A., «North-South R&D Spillovers», *Economic Journal,* vol. 107, n. 440, 1997, 139-49.

CONLEY T. - LIGON E., «Economic Distance and Cross-Country Spillovers», *Journal of Economic Growth,* vol. 7, no. 2, 2002, pp. 157-87.

EATON J. - KORTUM S., «Trade in Ideas: Productivity and Patenting in the OECD», *Journal of International Economics*, vol. 40, 1996, pp. 1-28.

EDMOND C., «Some Panel Cointegration Models of International R&D Spillovers», *Journal of Macroeconomics,* vol. 23, no. 2, 2001, pp. 241-60.

FEENSTRA R., «Trade and Uneven Growth», *Journal of Development Economics*, vol. 49, 1996, pp. 229-56.

FRANTZEN D., «Intersectoral and International R&D Knowledge Spillovers and Total Factor Productivity», *Scottish Journal of Political Economy,* vol. 49, no. 3, 2002, pp. 280-303.

GLOBERMAN S. - KOKKO A. - SJOHOLM F., «International Technology Diffusion: Evidence from Swedish Patent Data» *Kyklos*, vol. 53, no. 1, 2000, pp.17-38.

GRILICHES Z., «The Search for R&D Spillovers», *Scandinavian Journal of Economics*, vol. 94, no. 0, 1992, pp. 29-47.

GRILICHES Z. - LICHTENBERG F., «R&D and Productivity Growth at the Industry Level: Is There Still a Relationship?», in GRILICHES Z. (ed.), *R&D, Patents and Productivity*, Chicago, University of Chicago Press, 1984.

GROSSMAN G. - HELPMAN E., *Innovation and Growth in the Global Economy*, Cambridge, MIT Press, 1991.

HALL R. - JONES C., «Why Do Some Countries Produce So Much More Output Per Worker Than Others?», *Quarterly Journal of Economics*, vol. 114, no. 1, 1990, pp. 83-116.

JAFFE A., «Technological Opportunities and Spillovers of R&D: Evidence from Firm's Patents, Profits and Market Value», *American Economic Review*, vol. 76, 1986, pp. 984-1001.

JAFFE A. - TRAJTENBERG M. - HENDERSON R., «Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations», *Quarterly Journal of Economics,* vol. 108, no. 3, 1992, pp. 577-98.

JAFFE A. - TRAJTENBERG M., *Patents, Citations and Innovations*, Cambridge (MA), MIT Press, 2002.

JOZEFOWICZ J., «Estimating University Spillovers with Count Panel Data», *Pennsylvania Economic Review,* vol. 11, no. 1, 2002, pp. 58-75.

KAISER U., «Measuring Knowledge Spillovers in Manufacturing and Services: An Empirical Assessment of Alternative Approaches», *Research Policy,* vol. 31, no. 1, 2002, pp. 125-44.

KAO C. - CHIANG M.H. - CHEN B., «International R&D Spillovers and Application of Estimation and Inference in Panel Cointegration», *Oxford Bulletin of Economics and Statistics,* vol. 61, no. 0, 1999, pp. 691-709.

KELLER W., «Are International R&D Spillovers Trade-Related? Analyzing Spillovers among Randomly Matched Trade Partners», *European Economic Review,* vol. 42, no. 8, 1998, pp. 1469-81.

— —, «Trade and the Transmission of Technology», *Journal of Economic Growth*, vol. 7, 2002*a*, pp. 5-24.

— —, «Geographic Localization of International Technology Diffusion», *American Economic Review,* vol. 92, 2002*b*, pp. 120-42.

MADDEN G. - SAVAGE S. -BLOXHAM J.P., «Asian and OECD International R&D Spillovers», *Applied Economic Letters,* vol. 8, no. 7, 2001, pp. 431-5.

MANSFIELD E. - RAPPOPORT J. - ROMEO A. - WAGNER S. - BEARDSLEY J., «Social and Private Rates of Returns for Industrial Innovation», *Quarterly Journal of Economics*, vol. 77, 1977, pp. 221-40.

MARUSETH P.B. - VERSPAGEN B., «Knowledge Spillovers in Europe: A Patent Citation Analysis», *Scandinavian Journal of Economics*, vol. 104, no. 4, 2001, pp. 531-43.

MOHNEN P., «R&D Externalities and Productivity Growth», *STI Review OECD*, no. 18, 1996, pp. 39-66.

PERI G., «Knowledge Flows, R&D Spillovers and Innovation», Mannheim, ZEW, Discussion Paper, no. 03-40, 2003.

RIVERA-BATIZ L. - ROMER P.M., «Economic Integration and Endogenous Growth», *Quarterly Journal of Economics*, vol. 106, no. 2, 1991, pp. 227-44.

ROMER P.M., «Endogenous Technological Change», *Journal of Political Economy*, vol. 98, no. 5, 1990, pp. 71-102.

SCHERER F.M., «Using Linked Patents and R&D Data to Measure Interindustry Technology Flows», in GRILICHES Z. (ed.), *R&D, Patents and Productivity Editions*, Chicago, University of Chicago Press, 1984.

SOLOW R., «A Contribution to the Theory of Economic Growth», *Quarterly Journal of Economics*, vol. 70, 1956, pp. 65-94.

— —, «Technical Change and the Aggregate Production Function», *Review of Economics and Statistics*, vol. 39, 1957, pp. 312-20.

TERLECKY N., «Direct and Indirect Effects of Industrial Research and Development on the Productivity Growth of Industries», in KENDRIK J.B. - VACCARA B.N. (eds.), *New Development in Productivity and Measurement and Analysis*, Chicago, University of Chicago Press, 1980.

WOLF E.N. - NADIRI M.I., «Spillover Effects, Linkage Structure, Technical Progress and R&D», New York, NYU C.V. Starr Center, *Research Report,* vol. 87, no. 43, 1987.