DETERMINANTS OF KNOWLEDGE FLOWS AND THEIR EFFECT ON INNOVATION

Giovanni Peri*

Abstract—Knowledge flows within and across countries may have important consequences for both productivity and innovation. We use data on 1.5 million patents and 4.5 million citations to estimate knowledge flows at the frontier of technology across 147 subnational regions during 1975–1996 within the frame of a gravity-like equation. We estimate that only 20% of average knowledge is learned outside the average region of origin, and only 9% is learned outside the country of origin. However, knowledge in the computer sector flows substantially farther, as does knowledge generated by technological leaders. In comparison with trade flows, we see that knowledge flows reach much farther. External accessible R&D gained through these flows has a strong positive effect on innovative activity for a panel of 113 European and North American regions over 22 years.

I. Introduction

THE large number of recent studies on innovation and R&D spillovers have yet to produce a consensus on how geography and technology can influence knowledge flows and on their effects on productivity and subsequent innovation. Often different approaches and research methods dealing with the same question have produced different estimates. In particular, two branches of the literature have progressed on separate avenues in their analysis of knowledge flows and have rarely reconciled their quantitative findings, in part because they lacked a common frame of analysis. One branch of the literature utilizes firm-level data, considering in great detail only a few sectors within a country, and develops the analysis of spillovers by focusing on *technological space.* We refer to this branch as the *micro-productivity* literature. The other branch examines technological flows and spillovers across large aggregate units such as countries or country sectors, emphasizing the geographic dimension of these flows. We call this branch the *trade-growth* literature. Interest in international knowledge flows was mainly generated by the theoretical analysis initiated by the new-growth and the new-trade literature. The idea of knowledge flows and R&D spillovers as key determinants of growth and international trade was first developed in seminal papers such as Krugman (1979), Grossman and Helpman (1991), and Rivera-Batiz and Romer (1991).

Received for publication May 9, 2003. Revision accepted for publication June 1, 2004.

* University of California, Davis; and NBER.

I thank Daron Acemoglu and two anonymous referees for extremely helpful comments and constructive criticism. Paul Beaudry, Richard Baldwin, Lee Branstetter, Marianne Feldman, Bronwyn Hall, Gordon Hanson, John Helliwell, Marc-Andreas Muendler, Maurice Obstfeld, Henry Overman, James Rauch, and Manuel Trajtenberg also provided useful suggestions. Paola Franceschi and Ahmed Rahman provided extremely competent advice in editing the paper. Shireen Al-Azzawi provided outstanding research assistance. I thank the Institute of Governmental Affairs (IGA) and the Institute on Global Conflict and Cooperation (IGCC) for funding this project. Errors are mine.

The contribution of this paper is to frame the issue of knowledge flows in a simple empirical specification, compatible with both the micro-productivity and the trade-growth traditions. We use a very large and detailed data set of cross-patent citations to learn about the relative direction and intensity of knowledge flows at the frontier of innovation across 147 subnational regions covering western Europe and North America. Further, we estimate the effect of these flows on innovation as revealed by regional patenting activity. The rest of the paper is organized as follows. Section II briefly reviews the relevant literature on knowledge flows. Section III describes the framework of our analysis. Section IV presents the data and discusses specification and measurement issues. In particular we discuss the use of patents as measures of innovation and of patent citations as indications of knowledge flows. Section V presents the estimates of aggregate knowledge flows across 147 European and North American regions. We qualify our results by looking at different sectors, different specifications, and different sources of knowledge flows. Section VI uses the estimates of knowledge flows across regions to calculate the impact of accessible external R&D on innovative output. Section VII concludes the paper.

II. Literature Review

In this brief review of the literature it is useful to distinguish between knowledge flows and their subsequent effects. *Knowledge flows* occur whenever an idea generated by a certain institution is learned by another institution. These flows denote a process of learning from someone else's ideas, effectively building a stock of accessible (or "borrowed") research and development (R&D) (Griliches, 1992). The *effects* of these flows, on the other hand, are measured by the impact of this "accessible R&D" on actual production or innovation. The present paper decomposes these two steps: we first analyze the propagation of knowledge through learning, and then we estimate its effect on innovation.

Knowledge flows defined as "learning" have been extensively analyzed in the micro-productivity literature following the seminal work of Zvi Griliches (1992). Several pieces of subsequent empirical research sharpened our understanding of the process of knowledge diffusion,[1] but, in actuality, they were simply the continuation and refinements of a well-established empirical tradition which analyzed R&D

[1] This short survey is meant to give a sense of the large body of existing work. Excellent surveys of the literature exist, such as Griliches (1992), Mohnen (1996), and Branstetter (1998).

spillovers.[2] A simple and widely used approach assumes that knowledge flows exist only between firms within the same "technological group"; no flows take place across different groups. This approach was used, for instance, by Bernstein and Nadiri (1989a, 1989b) for the U.S. high-tech industries, by Bernstein and Mohnen (1998) for U.S. and Japanese firms, and by Bernstein and Yan (1997) for Canadian and Japanese firms. More sophisticated measures of knowledge flows define technological distance as a bilateral concept and allow for different intensities of flows between each pair of firms. For example, Jaffe (1986) calculates the flow of knowledge between firm i and firm j as the uncentered correlation coefficient between the vectors of their specialization in technological sectors. Branstetter (2001) uses a similar methodology to analyze the impact of domestic and foreign R&D spillovers for U.S. and Japanese firms. Other studies attempting to proxy knowledge flows between firms or sectors include Wolf and Nadiri (1987), which uses input-output matrices; Terlecky (1980), which uses flows of intermediate capital goods; and Scherer (1984), which constructs a matrix of origin versus use of patents.

Only quite recently have knowledge flows been estimated using patent citations. This method stands out because it is the only one using a discernible trail left by learning. Patent citations actually document learning flows between the citing and the cited institution. Using such data, Jaffe, Trajtenberg, and Henderson (1992) test whether distance matters or not for knowledge flows within the United States. Jaffe and Trajtenberg (2002, chapters 8, 9), Adams (2002), and Jozefowicz (2002) compare knowledge flows originating in universities, federal labs, and firms, and Maruseth and Verspagen (2002) analyzes knowledge flows across European regions. Following these studies, we argue that patent citations provide interesting information tracking knowledge flows. With some caveats, citations provide the "trails in the sand" left by the act of learning, and they can be used to assess the direction and intensity of knowledge flows.

The trade-growth literature, on the other hand, has been hesitant to incorporate information from the data on patent citations, or to devote much attention to technological space in its analysis of international knowledge flows. One notable exception is the line of analysis pursued by Eaton and Kortum (1996). They use a structural model of trade and growth across countries, utilizing data on cross-country patenting, to identify knowledge flows. In particular, information on the share of inventions originating from country i and patented in country j is used to estimate the flow of knowledge between the two countries. A number of alternative approaches have been preferred by this literature. Following Coe and Helpman (1995), several articles treat trade flows as proxies for knowledge flows (for instance, Coe, Helpman, & Hoffmeister, 1997; Keller, 2002a, Madden, Savage, & Bloxham, 2001); other articles consider foreign direct investments as adequate proxies for knowledge flows.[3] Finally, rather than consider flows, some papers have inferred R&D spillovers from cross-country or cross-region productivity correlations (Conley & Ligon, 2002; Keller, 2002b).

Interestingly, the theoretical side of the trade-growth literature makes an emphatic distinction between trade flows and knowledge flows, arguing that the second, rather than the first, are responsible for development and growth [see, for instance, Grossman and Helpman (1991, chapter 9), Rivera-Batiz and Romer (1991)]. In spite of such theoretical attention, the empirical trade-growth literature has not made an effort to develop measures of international disembodied knowledge flows other than trade and FDI flows. The present paper takes a step in that direction using patent citation data as an alternative proxy of flows of knowledge across regions.

III. Basic Framework of Analysis

Let Q_{it} be an index of the innovative output of region i at time t. Region i is a subnational unit within country c. Assuming that the stock of R&D is the main input in the innovation activity of region i, then the production of Q_{it} can be expressed by the following log linear production function:

$$Q_{it} = X_{ct}(A_{it})^{\gamma}(A_{it}^{a})^{\mu}. \tag{1}$$

Here X_{ct} are institutional and policy factors specific to a country c and possibly evolving over time t; A_{it} is the stock of R&D, accumulated from past and current R&D investments in region i, denoted $R\&D_{it}$; and A_{it}^{a} is the stock of R&D accumulated 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} and to estimate the elasticities γ and μ.

Equation (1) can be seen as the production function of innovation. The accumulation of A_{it} is described as $\Delta A_{it} = R\&D_{it} - \delta A_{it}$, where δ is the depreciation rate of the R&D stock. We apply the perpetual inventory method to calculate the value of A_{it}. Our main focus, however, is on the construction of A_{it}^{a} and on the estimation of μ.

If research developments in one area were completely and immediately diffusible to all other areas, we could consider the external R&D stock accessible to region i simply as $A_{it}^{a} = \sum_{j\neq i} A_{jt}$. However, considering that diffusion of research results across regions may be less than perfect, the external accessible R&D stock in region i is given by $A_{it}^{a} = \sum_{j\neq i} \phi_{ij} A_{jt}$, where $\phi_{ij} \in [0, 1]$ is the percentage of R&D stock generated in region j that is accessible to region i. Substituting this expression for A_{it}^{a} into equation (1) and taking logs, we have the following equation:

[2] Bresnahan (1986), Mansfield et al. (1977), Scherer (1984), Schmookler (1951), Terlecky (1980), and Wolf and Nadiri (1987) are some notable examples of earlier studies.

[3] Blomstrom and Kokko (1998) review the main contributions of this literature.

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

According to equation (2), the log level of innovative output in region *i,* $\ln Q_{it}$, depends on a set of country-time effects, $\ln X_{ct}$; on the the stock of regional R&D, $\ln A_{it}$; and on the external accessible stock of R&D, $\ln(\Sigma_{j \neq i}\ \phi_{ij}\ A_{jt})$. If the stock of external accessible R&D has a positive impact on innovative output (that is, if $\mu > 0$), knowledge flows have positive effects.

However, in order to calculate the stock of external accessible R&D, $A_{it}{}^{a}$, we need a measure of ϕ_{ij} for each regional couple. To do this we use relative patent citation frequencies across regions, which implies the following assumption: if region *i* learns a certain share of knowledge generated in region *j,* this is equivalent to having access to that share of R&D resources from region *j.* Of course this constitutes an *indirect* access to knowledge, gained through the learning of results rather than through the direct use of the resources. Therefore such input is entered separately from the own stock of R&D, A_{it}. The parameters ϕ_{ij} capture the intensity of knowledge flows and can be interpreted as the share of the research results of region *j* learned by region *i.* These flows depend on several bilateral characteristics of the regions, their technological differences, their location, and anything else that can affect the cost and the value of learning from region *j* for scientists residing in region *i.* The parameter μ captures the the effect of accessible external research on production.

IV. Specification, Measurement, and Data

Our empirical analysis has two parts. We first estimate the parameters ϕ_{ij} using data on patent citation frequencies between regions. We then use these estimated values along with data on regional R&D and the number of patents granted to each region to estimate the elasticities μ and γ in equation (2). In this section we describe the data, explain our empirical procedure, and discuss some of our assumptions and caveats.

A. *Knowledge Flows and Patent Citations*

We indicate as $\phi_{ij}(\tau)$ the probability that a nonobsolete idea generated in region *j* at time t_0 is learned in region *i* by time $t_1 = t_0 + \tau$. Such probability will depend on characteristics of the regional couple (*i, j*), and on the time elapsed since invention, τ. We approximate $\phi_{ij}(\tau)$ with the share of ideas generated in region *j* that has been learned in region *i* within τ years of invention.[4] Similarly to Jaffe and Trajtenberg (2002, chapters 6, 7) and Caballero and Jaffe (1993), we represent the share $\phi_{ij}(\tau)$ as follows:

[4] This share would converge to the probability for large numbers of ideas.

$$\phi_{ij}(\tau) = e^{f(i,j)}(1 - e^{-\beta\tau}). \quad (3)$$

The factor $1 - e^{-\beta\tau}$ captures the notion that the likelihood of research results in region *j* becoming available in region *i* grows larger as time τ passes. It thus represents the cumulative probability function of region *i* learning the idea within τ years since invention. The factor $e^{f(i,j)}$ indicates that the intensity of learning between sending region *j* and receiving region *i* may depend on a large set of bilateral regional characteristics acting as potential *resistance factors.*

The main simplifying assumption embedded in equation (3) is that the effects of these resistance factors *f(i,j)* and the effects of time τ interact in a multiplicative way. This implies that, as time passes, more ideas originating from region *j* are learned in all other regions, but such an increase is proportional for any pair of receiving regions. In our empirical analysis we experiment with different time intervals between generated and learned ideas, from 2 to 10 years. In order to characterize the diffusion of knowledge, we fix the same interval of time τ for all regions, collect the constant terms, and explicitly express the function *f(i,j)* as dependent on a host of geographic and technological characteristics in the following manner:

$$\begin{aligned}\phi_{ij} = Ce^{f(i,j)} = \exp\ [&a + b_1(out\_region)_{ij} \\ &+ b_2(out\_next)_{ij} + b_3(out\_country)_{ij} \\ &+ b_4(out\_lang)_{ij} + b_5(out\_trbl)_{ij} \\ &+ b_6(dist)_{ij} + \underline{\gamma}(\underline{Tech.Controls})_{ij}].\end{aligned} \quad (4)$$

Equation (4) states that the (time-invariant) relative intensity of knowledge flows from region *j* to region *i* depends on several bilateral regional characteristics. Six geographic characteristics are considered, and resistance factors depending on technological characteristics are bundled in the vector labeled *Tech.Controls.* The variable $(out\_region)_{ij}$ is a dummy which equals 0 if $i = j$ and 1 otherwise; this indicates whether a learned idea has crossed at least one regional border. The variable $(out\_next)_{ij}$ is equal to 0 if $i = j$ or if region *i* and *j* share a border and 1 otherwise; this indicates whether a learned idea has crossed at least two regional borders. The variable $(out\_country)_{ij}$ is 0 if the two regions belong to the same country and 1 otherwise; this indicates whether a learned idea has crossed a national border. The variable $(out\_lang)_{ij}$ is 0 if the same language is spoken in the two regions and 1 otherwise; this indicates whether a learned idea has crossed a linguistic border. The variable $(out\_trbl)_{ij}$ is 0 if the two regions belong to the same trade bloc and 1 otherwise; it indicates whether a learned idea has crossed a trade-bloc border. Finally $(dist)_{ij}$ is simply the geographic distance between region *i* and region *j.* Estimates of the parameters b_1–b_6 and $\underline{\gamma}$ would thus provide a detailed characterization of how geographic and technological characteristics affect the flows of ideas across regions.

Though we do not observe ϕ_{ij} directly, we do observe patents and citations between patents. Following an established tradition, we utilize patent statistics to proxy the generation of innovative ideas. Though not perfect, the correspondence between patents and new ideas has been extensively employed in economic analysis, and does seem reasonable from both theoretical and empirical points of view. Moreover, some of the potential distortions in this correspondence are largely mitigated by our choice of regional units and by our controls. First, according to the standards of patentability defined by the U.S. Patent Office, a patentable idea should be original, nonobvious, and exploitable for economic profit. This is precisely what we consider to be a "new idea." Second, many applied economists have drawn from the large pool of patent data, and used it as a convenient measure of "new ideas" [see Griliches (1990) for a survey]. Similarly, theoretical economists [such as Romer (1990) and Grossman and Helpman (1991)] have equated one idea to one patent in their models.

In practice, however, two sources of "noise" prevent a perfect correspondence between patents and ideas. The first is that the propensity to patent a new idea may vary across regions. The second is that patents may have dissimilar contents of ideas, with some patents containing many ideas and other relatively few (see, for instance, Jaffe & Trajtenberg, 2002, chapter 2). Relative to firm-level studies, our analysis much diminishes this second problem. Because we rely on a very large number of patents in each region (almost 10,000 per region on average), differences in the content of ideas for individual patents are likely to be averaged out in large aggregates. Addressing the first issue, we allow the propensity to patent to differ across regions, denoting it as $1/\beta_j$ (not observable), so that the relation between the number of ideas generated in region j, Υ_j, and the count of patents granted to region j, P_j, is $\Upsilon_j = \beta_j P_j$. We designate a patent's region of origin as the region of residence of its first inventor. This method, as documented by Jaffe et al. (1992), allows us to attribute each patent to the region where the idea was actually developed. The regions considered in our analysis correspond to subnational areas with territorial unity as well as some administrative autonomy.

Whereas patents proxy new ideas, citations between patents proxy the diffusion of these ideas through learning. All patent applicants in the United States are required to identify the "prior art" used in developing the patent, which they do by including citations to previous patents. A citation informs us that the researcher knew about an existing idea and that that idea had some relevance in the research process. Jaffe et al. (1992) argue that such citations establish a "learning" link and that they are limited to those patents that had strict relevance to the development of the new ideas. Patent reviewers, in fact, may drop some citations if they judge them irrelevant. By the same token, inventors do not want to proliferate citations, as that would excessively restrict their claims on the use of the patent and reduce their potential profits. Therefore, unlike the incentives for writers of academic articles, there is an economic incentive for inventors not to overcite. What introduces noise for our use of citations is the fact that *reviewers* may add citations to the patent which do not necessarily reveal ideas known to the author. Thus we assume that reviewers simply add noise to the information contained in patent citations.

We use the extremely large number of citation links available in our data (approximately 4.5 million in total, implying an average of approximately 200 citations for each regional couple) to estimate the relative flows of knowledge from each region j to any other region i. Defining c_{ij} as the count of citations from patents in region i to patents in region j, and Φ_{ij} as the actual flow of ideas from region j to region i, we assume the following relationship between citations and knowledge flows:

$$c_{ij} = \psi_i \, \Phi_{ij} \, e^{\varepsilon_{ij}}. \tag{5}$$

Here ψ_i is a citing-region fixed effect that allows the average number of citations per patent to differ across (citing) regions, Φ_{ij} is the effective number of ideas generated in region j and learned in region i, and $e^{\varepsilon_{ij}}$ is a randomly distributed disturbance, where ε_{ij} is a zero-mean random noise.

From the relationship between patents and ideas and from equation (5) it follows that

$$\phi_{ij} = \frac{\Phi_{ij}}{\Upsilon_j} = \frac{c_{ij}}{\psi_i \beta_j P_j e^{\varepsilon_{ij}}} = Ce^{f(i,j)}. \tag{6}$$

The first equality is a definition: ϕ_{ij} equals the number of ideas learned in i (Φ_{ij}) relative to the total number ideas produced in j (Υ_j). The second equality is obtained by substituting the definitions of Φ_{ij} and Υ_j. The last equality comes from the first part of equation (4). Substituting equation (4) into (6) and rearranging, we obtain the following estimable specification:

$$\begin{aligned} c_{ij} = \exp [\rho_i &+ \vartheta_j + b_1(out\_region)_{ij} + b_2(out\_next)_{ij} \\ &+ b_3(out\_country)_{ij} + b_4(out\_lang)_{ij} \\ &+ b_5(out\_trbl)_{ij} + b_6(dist)_{ij} \\ &+ \underline{\gamma}(\underline{Tech.Controls})_{ij} + \varepsilon_{ij}]. \end{aligned} \tag{7}$$

This equation has an easy interpretation along with some features that appeal both to the micro-productivity and trade-growth literatures. The dependent variable is the count of citation links calculated for region i as citing region and region j as cited region. As mentioned above, such a measure proxies for the flow of ideas from region j to region i. It depends on citing-region fixed effects $\varphi_i = \ln \psi_i$, and on cited-region fixed effects $\vartheta_j = \ln(\beta_j P_j)$. The first set of effects controls for different propensities to cite across regions; the second set controls for the different numbers of patents (P_j) and different propensities to patent (β_j) across regions. In general, the fixed effects control for any citing- and cited-region-specific characteristics. Let me emphasize that, as a consequence, the information used to identify relative flows of ideas is completely orthogonal to the regional propensity to patent new ideas, the regional propensity to cite ideas, and the total amount of regional patenting. The identifying variation is thus given by the frequency of citations from the average patent in region j to the average patent in region i, relative to the frequency of citations between patents of region j itself.[5] Once we control for fixed effects and we allow for random errors ε_{ij}, we can estimate the parameters b_1–b_6 and $\underline{\gamma}$.

The regression (7) is familiar in the micro-productivity literature and is often estimated using either a nonlinear least squares regression (for example, Jaffe & Trajtenberg, 2002, chapter 7) or, because citations are count data, a negative-binomial regression (Branstetter, 2000). On the other hand if we take logs on both sides of equation (7), we obtain a linear regression which is reminiscent of a *gravity equation* very popular and heavily used in the recent trade literature.[6] Typically, the trade literature would estimate such an equation using OLS and simply omit the regional couples with no trade between them.

B. *Own R&D, Accessible External R&D, and Innovation*

In section VI we estimate the effect of a region's stock and the external accessible stock of R&D on the region's innovative output. We construct $\hat{\phi}_{ij}$, the estimated share of the knowledge flowing from j to i, by substituting the estimated parameters b_1–b_6 and $\underline{\gamma}$ from the regression (7) into equation (4). Such weights, plus measures of A_{jt}, are used to construct the estimated stock of accessible external R&D for each region i: $A^a_{it} = \Sigma_{j\neq i}\, \hat{\phi}_{ji}\, A_{jt}$. We adopt the standardization $a = 0$ in equation (4) so that, by construction, $\hat{\phi}_{ii} = 1$. This means that, by definition, the results A_{it} of research generated in region i are fully accessible to region i itself.

Equation (2) is estimated using citation-weighted patent counts P_{it} as measures of Q_{it}. The coefficients $\hat{\phi}_{ji}$ are estimated using information orthogonal to the amount of patenting and to the propensities to patent and to cite in each region. They are based only on relative frequencies. The construction of the variable A^a_{it}, therefore, does not introduce any mechanical correlation with the dependent variable P_{it}.

The stock A_{it} in each region for the period 1975–1996 is constructed using the perpetual inventory method. We initialize R&D stocks for the year 1975[7] and use the recursive formula $A_{it} = (1 - \delta)\, A_{i\ t-1} + (R\&D)_{it}$ to calculate the stock in the following years. The value chosen for δ, the depreciation of R&D capital, is 10%, a calibration value preferred by most of the literature (see Keller, 2002b). Finally we control for country-time fixed effects D_{ct}.[8] This implies that time-varying institutional or policy differences across countries, or country-specific propensities to patent in the United States, or indeed any other factor changing with country and time, does not affect our estimates of γ and μ. The variations that identify the coefficients are strictly the differences (in patenting and R&D) across regions within countries. Including a zero-mean random disturbance u_{it}, the estimated equation is

$$\ln P_{it} = D_{ct} + \gamma \ln A_{it} + \mu \ln\left(\sum_{j\neq i} \hat{\phi}_{ji}\, A_{jt}\right) + u_{it}\,. \tag{8}$$

C. *Description of the Data*

Patent and citation data originate from the NBER Patent and Citation data set, which is publicly available and described in detail by Hall, Jaffe, and Trajtenberg (2001). This data set contains all the patents granted by the U.S. patent office and, since 1975, all citations made by each patent of other patents. We choose the sample of patents granted between 1975 and 1996 whose inventor is a resident of one of the 147 subnational regions within one of eighteen countries listed in appendix A (all in Europe and North America). Due to the very exacting manual effort required in locating the residence of each inventor within a region and in gathering regional R&D data, we limited our study to Europe and North America. The only important innovating country thus left out is Japan. However, its exclusion should not affect our estimates much. Due to its remoteness, none of its regions share any borders, languages, or trade agreements with any of the regions considered here. Thus we can think of knowledge generated in that country simply as having a fixed common effect on European and North American regions. Our final sample contains approximately 1.5 million patents and approximately 4.5 million pairs of citations.

Table 1 reports some summary statistics at the regional level. Panel A shows the average and standard deviation for the number of patents granted each year to residents of the 147 regions, as well as for R&D spending, patent citations, and geographic distance between regions. Panels B and C show the identity and some characteristics of the most and least innovative regions in our sample. The top innovator is California, which was granted more than 6,000 patents per year. High in the ranking are also some German, French, and British regions. The bottom of the list is occupied by Greek, Spanish, and East German regions, each with one or less than one patent granted each year (on average).

Data on R&D for the period 1975–1996 are not available for all regions. From national statistical agencies we obtain

[5] Excluding self-citations.

[6] For a review of the main estimates obtained using gravity equations see Feenstra (2003, chapter 5).

[7] The initial value of R&D stock is set at $A_{i1975} = (R\&D)_{i1975}/(\delta + g_i)$, where δ is the depreciation rate of R&D capital and g_i is the average growth rate of R&D spending in the country to which region i belongs for the period 1975–1980.

[8] These effects are equivalent to the terms $\ln X_{ct}$ in equation (2).

TABLE 1.—DESCRIPTIVE STATISTICS RELATIVE TO 147 REGIONS IN EUROPE AND NORTH AMERICA

Panel A: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Average number of patents granted yearly, 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-citations	171	1,147	0	99,137
Geographical distance (thousands of kilometers)	4.44	3.22	0	13.70

Panel B: Representative High-Patenting Regions

Region	Country	Yearly granted patents average (1975–1996)	R&D spending (% GDP, 1975–1996)
California (overall rank: 1)	USA	6,434	3.86
New York (overall rank: 2)	USA	3,856	2.00
New Jersey (overall rank: 3)	USA	2,978	3.59
Nordrhein-Westfalen (overall rank: 10)	GER	1,507	1.86
Baden Württemberg (overall rank: 11)	GER	1,423	2.93
Ile de France (overall rank: 16)	FRA	1,104	3.51
Southwest U.K. (overall rank: 17)	UK	976	3.45

Panel C: Representative Low-Patenting Regions

	Country	Yearly granted patents average (1975–1996)	R&D spending (% GDP, 1975–1996)
Sachsen-Anhalt	GER	1.00	1.50
Mecklenburg-Vorpommern	GER	0.91	1.14
Prince Edward Island	CAN	0.86	0.71
Centro España	SPA	0.64	0.44
Kentriki Ellada	GRE	0.41	0.27
Kriti	GRE	0.27	0.53

Notes: Citation frequencies are calculated omitting self-citations, that is, citations between patents whose first authors belong to the same company or institution.

the share of total national R&D in the business sector that is performed in each region of nine important countries.[9] We then use the ANBERD data on intramural business enterprise R&D, measured in constant 1990 U.S. dollars, and allocate the national aggregates according to the regional shares. Missing years were filled by interpolation. This method allows us to obtain a balanced panel for regions in all the main countries, namely the United States, Canada, Germany, France, Italy, the United Kingdom, Spain, and the Netherlands. These countries include the 113 regions from which virtually all the major innovations arise. Table 1 shows that the important innovators (top regions) spend between 2% and 4% of their GDP on business R&D, and the least active regions spend less than 1% on R&D.

V. Estimates of Knowledge Flows

A. *Aggregate Flows: Geographic and Technological Determinants*

We present in this section the estimated coefficients from equation (7). Specification I in table 2 is the baseline regression and shows the effects of geographic and technological resistance factors on aggregate knowledge flows. We estimate equation (7) by maximum likelihood, using a negative-binomial specification.[10] We also estimated all specifications using OLS [after taking logs of both sides of equation (7)], obtaining similar coefficient estimates. For the sake of brevity we report only the negative-binomial estimates. The dependent variable in specification I of table 2 is the count of citation links, omitting self-citations,[11] between patents of region i and patents of region j generated within the first 10 years after the cited patent was granted. Such a time span should be long enough for us to capture the most relevant portion of nonobsolete knowledge diffusion. Nonetheless, we also analyze flows within 2 years in specification III, flows within 6 years in specification IV, and all citation pairs within the entire sample in specification V. Finally, in specification VI we fix a large cohort of originating patents, 1975–1985, and allow the citing patents to cover the longest period available, 1975–1996.

The geographic resistance factors included in each specification are presented in section IV A. Each coefficient

[9] The detailed sources for national R&D data are described in appendix C of Peri (2003).

[10] Using this method, we can include all the regional couples with zero citations and, by assuming a generalized Poisson data-generating process, we allow for the fact that citations are count data.

[11] 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 R&D spillovers. We also estimated specifications including self-citations, and the only difference is that the coefficient on crossing region border is increased by roughly 10%–15%.

TABLE 2.—BASIC SPECIFICATION: GEOGRAPHIC AND TECHNOLOGICAL DETERMINANTS OF AVERAGE KNOWLEDGE FLOWS

Specification: Flow	I Within 10 years	II Within 10 years	III Within 2 years	IV Within 6 years	V All Couples	VI Citing: 75–96 Cited: 75–85
Crossing region border	−1.57*	−1.50*	−1.45*	−1.57*	−1.53*	−1.38*
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Crossing next-region border	−0.33*	−0.32*	−0.27*	−0.31*	−0.34*	−0.32*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Crossing country border	−0.19*	−0.19*	−0.19*	−0.19*	−0.20*	−0.18*
	(0.02)	(0.15)	(0.02)	(0.02)	(0.02)	(0.02)
Crossing trade-block border	0.04	0.04	0.04	0.03	0.03	0.03
	(0.025)	(0.025)	(0.03)	(0.025)	(0.025)	(0.02)
Crossing linguistic border	−0.19*	−0.20*	−0.16*	−0.19*	−0.18*	−0.16*
	(0.01)	(0.12)	(0.02)	(0.01)	(0.01)	(0.01)
1,000 km farther	−0.03*	−0.03*	−0.04*	−0.03*	−0.03*	−0.02*
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Difference in technological specialization[a]	−2.91*	−2.85*	−3.02*	−2.98*	−2.87*	−2.82*
	(0.05)	(0.04)	(0.07)	(0.06)	(0.05)	(0.06)
Difference in technological advancement[b]	−0.75*		−0.22*	−0.52*	−0.79*	−1.00*
	(0.10)		(0.05)	(0.12)	(0.10)	(0.11)
Technological advantage of receiving region (citing)		−0.62*				
		(0.10)				
Technological disadvantage of receiving region (citing)		−0.81*				
		(0.10)				
Citing-region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cited-region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,609	21,609	21,609	21,609	21,609	21,609
Log likelihood	−55,952.4	−55,855.9	−33,815.1	−50,758.6	−59,778.4	−52,892.7

Notes: Citations are calculated omitting self-citations, that is, citations within the same institution. Method of estimation: maximum likelihood on a negative-binomial specification. Asymptotic heteroskedasticity-robust standard errors in parentheses. An asterisk indicates significance at 1% level.

[a] Index $(SpecDis)_{ij}$ *defined in section V A.*

[b] Difference in logged average real R&D spending per worker between the receiving and the originating region (1991–1996).

Specification I: Dependent variable: log of citations between patents with citing and cited patents less than 10 years apart during 1975–1996.
Specification II: Dependent variable: log of citations between patents with citing and cited patents less than 10 years apart during 1975–1996.
Specification III: Dependent variable: log of citations between patents with citing and cited patents less than 2 years apart during 1975–1996.
Specification IV: Dependent variable: log of citations between patents with citing and cited patents less than 6 years apart during 1975–1996.
Specification V: Dependent variable: log of citations between patents, including all the citing and cited patents, during 1975–1996.
Specification VI: Dependent variable: log of citations between patents including citing patents 1975–1996, cited patents 1975–1985.

captures the reduction of knowledge flows as each designated border is crossed. To convert each value to a percentage drop we need to use the exponential formula. For instance, the first coefficient of specification I in table 2 implies that on crossing the first regional border, knowledge diminishes to 21% ($= e^{-1.57}$) of its initial level. The second coefficient implies that only 72% ($= e^{-0.33}$) of the 21% of this initial knowledge passes a second regional border. Therefore only 15% ($= 21\% \times 72\%$) of initial knowledge flows across and beyond two regional borders. A further 19% ($= 1 - e^{-0.19}$) is lost passing the country border, leaving approximately 12% of the initial knowledge. Crossing a trade-bloc border has essentially no effect (the estimated coefficient is not significantly different from 0), whereas passing a linguistic border reduces knowledge by a further 19%. On top of these effects, geographic distance reduces flows by 3% for each thousand kilometers traveled. Overall, the drop in learning resulting from geographic resistance factors is substantial.

Estimates across the specifications (I through VI) in table 2 are remarkably stable. Whether 2, 6, or 10 years elapses, the degree of relative geographic localization of knowledge remains rather stable. Even when we allow for the citing patents to span the whole sample and we track the diffusion of knowledge from a large cohort of cited patents (specification VI), localization decreases only slightly. For example, 25% of ideas generated between 1975 and 1985 were learned out of their region of origin by 1995, as opposed to 21% estimated when considering only the first 10 years after invention. Similarly small differences between specifications I and VI arise when considering the negative effect of crossing a linguistic border (−16% rather than −19%) or the overall effect of distance (−2% per 1,000 km, rather than −3%). In general, given that the bulk of citation is received during the first 10 years after the invention, adding longer lags does not significantly change the geographic pattern of diffusion.

As regions with similar levels of technological specialization and sophistication may be located near each other, failing to control for technological differences across regions may result in overestimating the effect of geography. Therefore we include in each specification of table 2 two proxies of regional technological differences. The first proxy[12] captures the difference in technological specialization between two regions. Specifically, all patents granted to a region (call it region i) are grouped into 36 technological classes as defined by the international patent classification.[13] The share of patents granted to region i in each technological

[12] We follow Jaffe (1986).

[13] Classes are reported in Appendix B.

class s (= 1, 2, . . . , 36) is then arranged into a vector, $\underline{Sh}_i = (sh_{i1}, sh_{i2}, \ldots, sh_{i36})$. The uncentered correlation coefficient between the vectors of regions i and j, calculated as $(SpecCorr)_{ij} = (\underline{Sh}_i'\underline{Sh}_j)/[\Sigma_s (sh_{is})^2 \Sigma_s (sh_{js})^2]^{1/2}$, is a measure of the similarity in technological specialization. Its value ranges between 0 and 1; the closer it is to 1, the larger is the overlap in technological classes of specialization. Thus we use $(SpecDis)_{ij} = 1 - (SpecCorr)_{ij}$ to proxy for the differences in technological specialization between region i and region j. The second index we construct captures differences in technological development between two regions. It measures the difference in logged average real spending in R&D per worker (1991–1996) between the region that receives and the region that originates the knowledge flow.[14]

Examining both indices, we find that specialization in different technological fields tends to impede knowledge flows, whereas greater technological distance between two regions tends to encourage knowledge flows from more advanced regions. Using estimates in column I, two regions with technological specializations in completely different fields [$(SpecDis)_{ij} = 1$] have knowledge flows only 5% (= $e^{-2.91}$) as large as two regions with identical technological specializations [$(SpecDis)_{ij} = 0$]. Moreover, a region receives 8% (= $e^{-0.1\times0.75}$) higher knowledge flows from a region 10% more intensive in R&D than from a region as R&D intensive as itself. In specification II we separate the potential impact of differences in technological advancement between cases when this difference is positive (receiving region has technological advantage) and when it is negative (receiving region has technological disadvantage). The estimates show, however, that the impact of a change in R&D intensity is similar in the two cases. If the receiving region is more R&D-intensive than the originating region, an increase of 10% in that intensity reduces by 6.1% the flows received; if it is less R&D-intensive, the same increase reduces flows by 7.8%.

Unlike geographical distance, technological distance appears to matter more for knowledge flows as time passes. The estimate of this coefficient grows larger (in absolute value) as we progress from 2-year lags to the longest lags (column VI). Thus, as time passes, knowledge from technologically advanced regions is learned proportionally more by less advanced ones.

Several other robustness checks were performed. Most of them are available in Peri (2003). Here we mention only two of them. We estimate knowledge flows separately for two sub*periods* (1975–1986 and 1986–1996) as well as for two sub*samples* (Europe and North America). In either case our estimated effects of the geographic and technological resistance factors variables remain virtually the same.

[14] We also used the difference in output per worker as index of difference in technological advancement. We prefer R&D spending per worker because when we included both indices only the one based on R&D per worker had a significant effect on knowledge flows.

At this point it is useful to summarize some of our results graphically. Figure 1 represents our estimated effects of geographic resistance factors on knowledge flows. It shows the estimated decay of knowledge flows moving out of the originating region, out of its neighbor, out of the country, out of the linguistic area, out of the trade bloc, and out by steps of 1,000 km. The total knowledge generated in a region is standardized to 100. Five decay functions are reported, which correspond to the estimated parameters in specification I, III, IV, V, and VI of table 2. Visual inspection of figure 1 confirms the significance of the drop in knowledge flows when moving out of the region. We can also appreciate from figure 1 that decay functions produced using different estimates are rather close to each other.

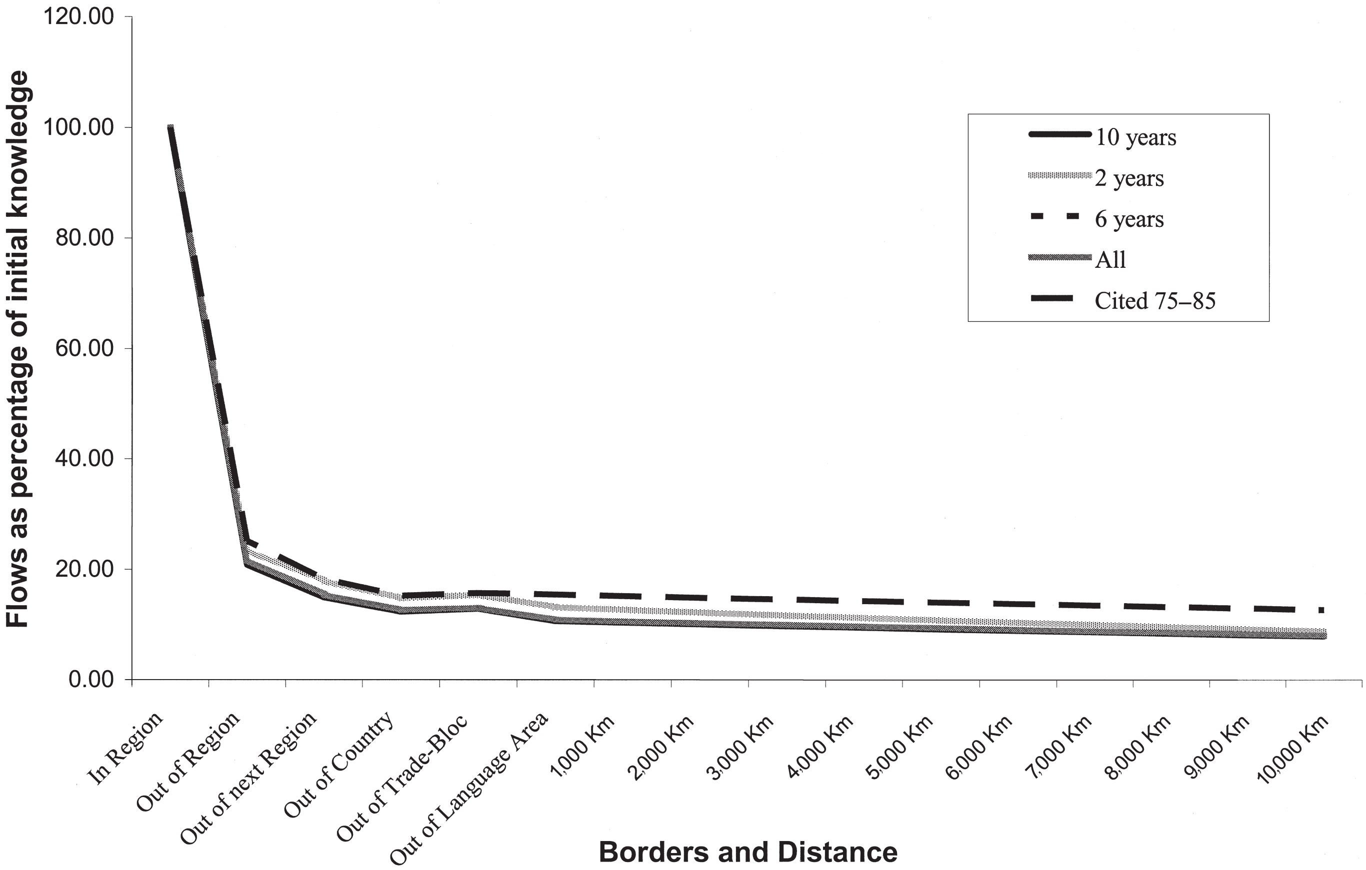
B. *Flows within Sectors*

Table 3 reports the estimated effects of resistance factors on knowledge flows within six sectors. We use a time window of 10 years from the originating patent, and perform the usual maximum likelihood estimation on a negative-binomial specification. The differences in technological specializations across regions are calculated using specializations in subsectors within each considered sector. Differences in technological advancement are computed using total R&D intensity, as in table 2. Figure 2 shows the decay functions for each technological sector using the coefficients in table 3. The computer sector exhibits by far the most extensive geographic diffusion of knowledge: its decay function is significantly above those of all other sectors. Information technology, mostly computer-related, seems an exception in its range of diffusion and looks much like a global technology. Close to 40% of computer-related knowledge generated in a region is learned outside of it, and 25% of this knowledge flows all the way out of its country and linguistic area of origin. The other sectors' knowledge is much more localized, although not uniformly so: ideas in the electronics and drugs sectors spread farther than ideas in the mechanical or chemical sector. Specifically, more than 25% of new ideas in electronics are learned outside the region of origin, versus less than 19% and 10%, respectively, for ideas in the chemical and the mechanical sector. The relatively wider diffusion of ideas from sectors related to information technologies (computer and electronics) seems an illustration both of the importance of these sectors and of the global scope of research in this area.

C. *Flows from Leading Regions*

As suggested by the positive effect of technological advancement on outgoing flows of knowledge, the regions leading world research originate ideas that are more likely to flow to other regions. It is reasonable to think that technological leaders not only generate larger flows toward less advanced regions (captured by the distance in technological advancement), but also generate flows with wider

FIGURE 1.—DECAY OF KNOWLEDGE FLOWS DUE TO GEOGRAPHICAL BARRIERS



geographic scope. Our data show that research across regions is rather concentrated. The top 20 regions (out of 147) perform 60% of total R&D in the sample (which is approximately 50% of the world R&D). The technological leaders, therefore, may act as learning sources for other regions more than an average region would.

To explore this aspect further, we focus on the top twenty regions (for total R&D spending), and we consider only the knowledge flows originating in those regions. Of the top twenty regions, eleven are in the United States, four in Germany, one in Canada, one in France, one in the United Kingdom, one in the Netherlands, and one in Italy. Table 5 shows the estimated effects of resistance factors on knowledge flows, considering only the top twenty regions as sources of learning (regions whose patents are cited).[15] Specifications I to III use different time intervals, from 10 years (column I), to 6 years (column II) to 2 years (column III). Specification IV uses the ideas originated in 1975–1985 and their diffusion, measured using all citations until 1996. Consistently and robustly across specifications these estimates show much less geographic localization than the average knowledge flows estimated in table 2. Even considering the most conservative estimate (column III), 56% of knowledge originating from leading regions crosses over at least one regional border, compared with only 20% for the knowledge originating in the average region. Similarly, 35% flows out of the linguistic border, versus 10% for average knowledge, and 25% flows all the way to 10,000 km of distance, versus only 7.5% for average knowledge. Technological differences in specialization and advancement still play important roles in knowledge diffusion. The greater quality and relevance of knowledge generated by technological leaders likely grants such knowledge large diffusion. Figure 3 demonstrates the visual comparison between the decay functions of knowledge flows from technological leaders (estimates I and III in table 4) and average flows (estimate I in table 2). The visual impression confirms the strikingly broader reach of knowledge generated by technological leaders relative to the average region. Finally, the decay function of knowledge calculated using long lags (shaded line) is discernibly above the function calculated using two-year lags (solid line).

D. *Comparisons with Existing Estimates*

Our estimates, which imply a high degree of localization of knowledge flows, raise the question: are such estimates

[15] We performed the same exercise using the top 15 and the top 25 regions, and we obtained very similar results.

TABLE 3.—DETERMINANTS OF KNOWLEDGE FLOWS FOR SIX TECHNOLOGICAL CLASSES

Specification: Flow	I Computers	II Electronics	III Drugs	IV Mechanical	V Chemical	VI Others
Crossing region border	−1.00*	−1.38*	−1.57*	−1.57*	−1.68*	−1.38*
	(0.10)	(0.11)	(0.08)	(0.11)	(0.08)	(0.09)
Crossing next-region border	−0.21*	−0.24*	−0.08*	−0.34*	−0.31*	−0.39*
	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Crossing country border	−0.14*	−0.17*	−0.20*	−0.10*	−0.19*	−0.20*
	(0.04)	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)
Crossing trade-block border	0.05	0.06	0.01	0.03	0.04	0.04
	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)
Crossing linguistic border	−0.11*	−0.14*	−0.11*	−0.17*	−0.15*	−0.14*
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
1000 km farther	−0.03*	−0.03*	−0.03*	−0.05*	−0.03*	−0.05*
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)
Difference in technological specialization within the sector[a]	−2.20*	−2.70*	−1.17*	−3.29*	−2.67*	−3.46*
	(0.14)	(0.12)	(0.06)	(0.14)	(0.11)	(0.12)
Difference in technological advancement[b]	−1.00*	−1.10*	−0.88*	−0.80*	−0.20*	−0.20
	(0.05)	(0.05)	(0.09)	(0.07)	(0.04)	(0.17)
Citing-region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Cited-region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,609	21,609	21,609	21,609	21,609	21,609
Log likelihood	−19,150.3	−26,503.6	−23,898.7	−32,042.4	−30,591.7	−35,340.7
Original number of citations	243,563	333,637	342,572	243,902	356,614	486,513

Notes: Citations are calculated omitting self-citations (citations within the same institution). Method of estimation: maximum likelihood on a negative-binomial specification. Asymptotic, heteroskedasticity-robust standard errors in parentheses. An asterisk indicates significance at 1% level.
[a] Index $(SpecDis)_{ij}$ defined in section V A for subsectors within technological class.
[b] Difference in logged average real R&D spending per worker between the receiving and the originating region (1991–1996).
Specification I: Dependent variable: log of citations between patents in computer class with citing and cited patents less than 10 years apart.
Specification II: Dependent variable: log of citations between patents in drugs class with citing and cited patents less than 10 years apart.
Specification III: Dependent variable: log of citations between patents in electronics class with citing and cited patents less than 10 years apart.
Specification IV: Dependent variable: log of citations between patents in chemical class with citing and cited patents less than 10 years apart.
Specification V: Dependent variable: log of citations between patents in mechanical class with citing and cited patents less than 10 years apart.
Specification VI: Dependent variable: log of citations between patents in other classes with citing and cited patents less than 10 years apart.

reasonable? In particular one may wonder how localized these knowledge flows are, relative to trade flows. As knowledge diffusion does not necessarily require the movements of goods or people, its effective reach should be beyond that of trade. Yet, to our knowledge, no one in the literature has performed an empirical comparison of the geographic scope of knowledge and trade flows.

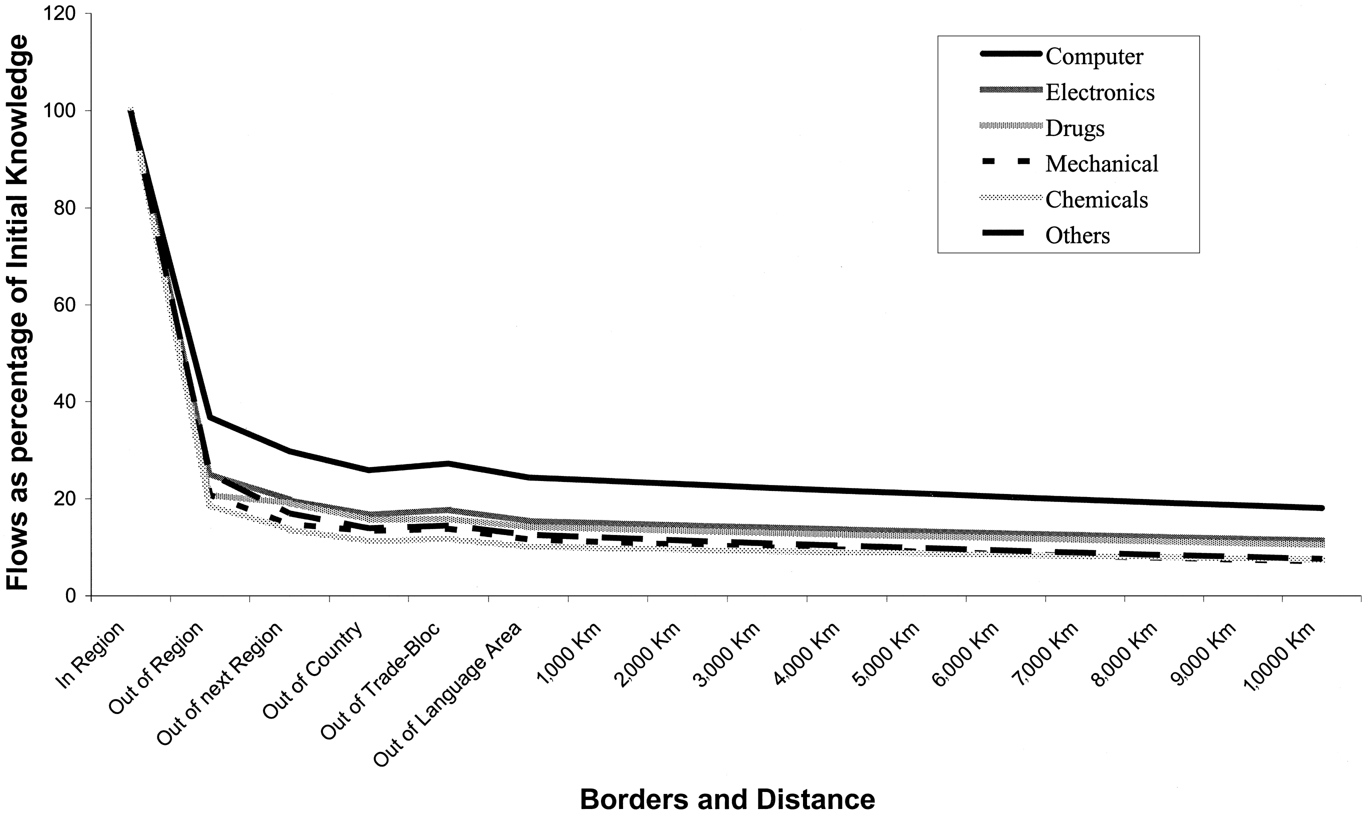
The trade literature has extensively studied the effect of geographic variables (mainly distance and borders) on total trade flows. As we have precise estimates of these effects on knowledge flows, we concentrate on these two effects and compare our estimates with those from standard trade gravity equations. To ensure maximum comparability with the existing trade estimates, we enter distance linearly (rather than exponentially) in equation (7), as is commonly done in trade specifications, and we include the whole set of citing- and cited-region fixed effects. We do not include other controls (such as proxies for technological distance), as they are normally not included in trade estimates. The specification is very similar to column I in table 2, except we enter distance in logs and we omit the technological variables. For the sake of brevity we do not report the full estimation.[16] The results of interest are as follows: the effect on knowledge flows of crossing a country's border is estimated to be −0.20 (standard error 0.01), and the effect of log(distance) is −0.19 (standard error 0.01). Estimates of border and distance effects on total trade from Anderson and Van Wincoop (2001) are −1.65 and −0.79, respectively, whereas Feenstra (2003) estimates them to be −1.55 and −1.25. Reasonably, borders and distance reduce trade flows 4 to 5 times more than they reduce knowledge flows.

Finally, our estimates can be also compared with some existing estimates of the geographic reach of knowledge flows based on patent citation data. The initial and influential work that assessed the degree of localization of knowledge flows using citations across patents was Jaffe et al. (1992). That paper used a much smaller sample, limited to the United States, and a very different method to estimate localization. However from the coefficients reported in their table III, we can recover some effects that can be compared with ours. Considering their sample without self-citations and with originating cohort 1980 [columns 4, 5, and 6 in the "Matching by State" panel in Jaffe et al. (1992)], they estimate a drop of citation flows moving out of the state[17] (corresponding to our "region" for the United States) of 50%–60%. Our most comparable estimates (column VI, table 2) give a drop of approximately 75% moving out of the region. For the country border effect, Jaffe et al. (1992)

[16] This and a longer discussion of knowledge and trade flows is in Peri (2003).

[17] We obtain this effect by comparing their matching fraction within SMSA with the matching fraction of the control group.

FIGURE 2.—DECAY OF KNOWLEDGE FLOWS BY TECHNOLOGICAL CLASS



estimate a drop by 12%–15% of citation flows, and our preferred estimate puts that drop at 18%–19% for 1975–1996 overall (table 2).

Maruseth and Verspaghen (2002) use citations between European-granted patents located in 112 European regions. Their estimated resistance effects of (log) distance, crossing a linguistic border, and crossing a country border are in the ranges 0.29–0.38, 0.20–0.28, and 1.53–1.56, respectively. The first two effects are rather similar to our estimates; however, our estimate of the country-border effect is significantly smaller than theirs (ranging between 0.18 and 0.20). As their estimates of country-border effects for knowledge flows are as high as those estimated in the literature for trade flows, we wonder if the process of patent revision at the European Patent Office generates an excessive own-country bias in the citation procedures.

Our estimates, in summary, reveal a degree of localization of learning consistent with those revealed by other studies of patent citations, but significantly lower than the localization of trade.

VI. Effects of Accessible External R&D on Innovation

Finally, table 5 reports the estimates of γ and μ in equation (8). As the top 20 innovating regions perform the majority of R&D in our sample and flows of knowledge from them reach substantially farther than average knowledge flows, we first consider them as the only source of relevant flows in constructing A_{it}^{a}. This choice allows us to minimize potential endogeneity problems in estimating the coefficient of A_{it}^{a}. In fact the basic specification (column I of table 5) does not include the top 20 regions in the regression as receivers of R&D spillovers, but rather includes only the remaining 93 regions, for the period 1975–1996. External accessible R&D in specification I is thus measured as $A_{it}^{a} = \Sigma_{j \in \text{Top 20}} (\hat{\phi}_{ij}^{\text{10yr}} A_{jt})$. The weights $\hat{\phi}_{ji}^{\text{10yr}}$ measure the share of knowledge generated in j and learned by i within ten years. They are estimated using the formula in equation (4) and the coefficients from specification I of table 4.[18] The measure of innovation (dependent variable) is ln P_{it}, where P_{it} is the count of patents granted to region i in year t, weighted for the citations received during the first 4 years.[19] We fully control for country $\times$ year fixed effects. Column II of table 5 checks to see whether limiting our attention to accessible R&D from the top 20 regions is a reasonable strategy. There we include as external accessible R&D the flow-weighted stocks originating from all regions (not only the top 20). In spite of potential worsening of the endogeneity problem, estimates obtained with this approach are not very different from those in column I. The elasticity of innovation to own R&D is estimated at

[18] With the usual standardization $a = 0$, which implies $\phi_{ii} = 1$ for all i.
[19] The elasticity estimates are similar using unweighted patents as the measure of innovative output (not reported).

FIGURE 3.—DECAY OF KNOWLEDGE FLOWS ORIGINATING FROM TECHNOLOGICAL LEADERS

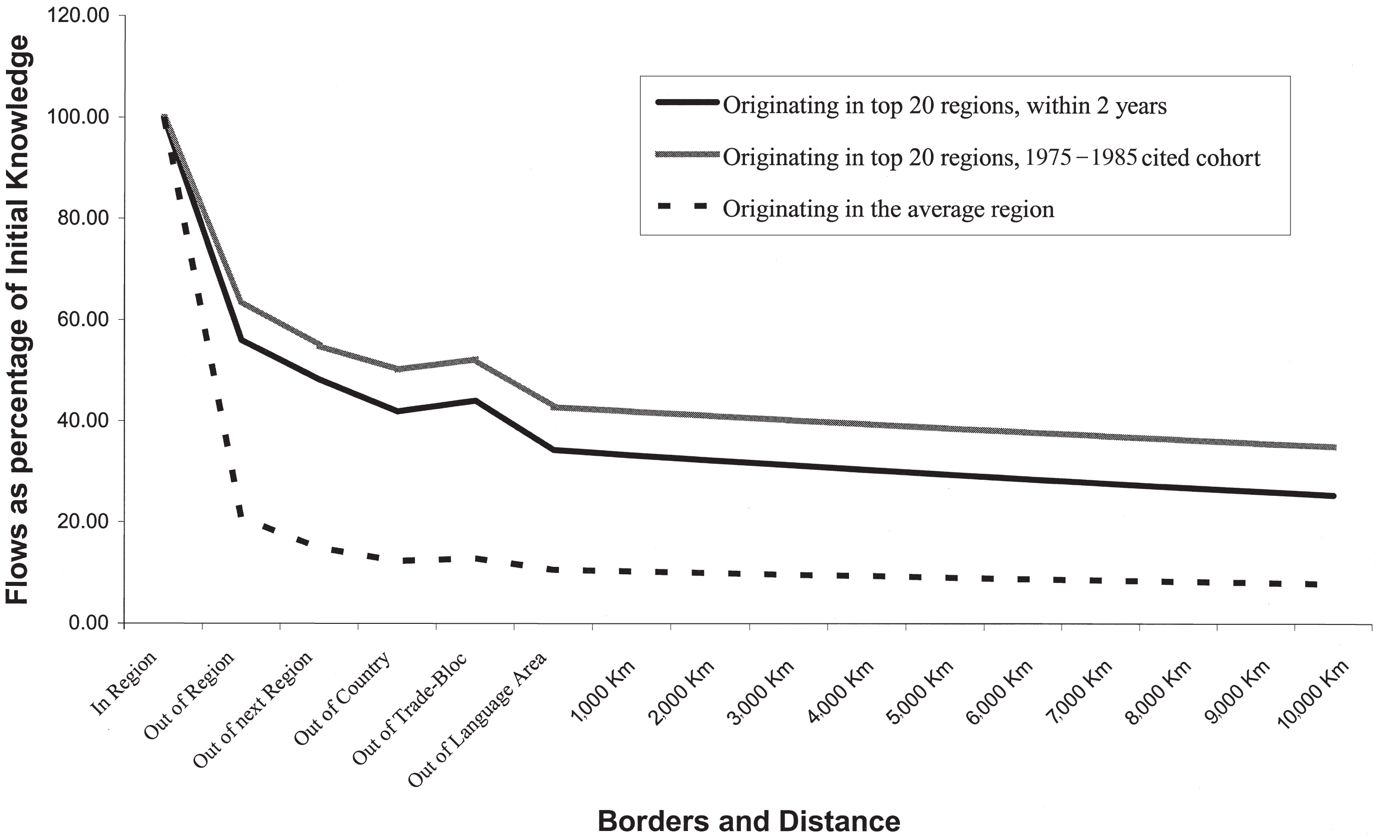


TABLE 4.—DETERMINANTS OF KNOWLEDGE FLOWS ORIGINATING FROM THE 20 MOST INNOVATIVE REGIONS

Specification:	I	II	III	IV
Flow	10 years	6 years	2 years	Citing: 75–96 Cited: 75–85
Crossing region border	−0.54*	−0.55*	−0.58*	−0.45*
	(0.07)	(0.07)	(0.08)	(0.07)
Crossing next-region border	−0.14*	−0.14*	−0.15*	−0.15*
	(0.03)	(0.03)	(0.03)	(0.03)
Crossing country border	−0.10*	−0.11*	−0.14*	−0.09*
	(0.02)	(0.02)	(0.02)	(0.03)
Crossing trade-bloc border	0.04	0.04	0.05	0.04
	(0.03)	(0.03)	(0.03)	(0.03)
Crossing linguistic border	−0.25*	−0.25*	−0.25*	−0.20*
	(0.02)	(0.02)	(0.03)	(0.02)
1000 km farther	−0.02*	−0.03*	−0.03*	−0.02*
	(0.002)	(0.002)	(0.003)	(0.002)
Difference in technological specialization[b]	−2.92*	−2.96*	−3.10*	−2.81*
	(0.09)	(0.09)	(0.10)	(0.08)
Difference in technological advancement[b]	−1.35*	−1.20*	−0.51*	−1.50*
	(0.10)	(0.15)	(0.36)	(0.22)
Citing-region fixed effects	Yes	Yes	Yes	Yes
Citied-region fixed effects	Yes	Yes	Yes	Yes
Observations	2,961	2,961	2,961	2,961
Log Likelihood	−13,754.5	−12,800.5	−9,610.6	−13,206.1

Notes: Citations are calculated omitting self-citations (citations within the same institution). Method of estimation: maximum likelihood on a negative-binomial specification. Asymptotic, heteroskedasticity-robust standard errors in parentheses. An asterisk indicates significance at 1% level.

[a] Index $(SpecDis)_{ij}$ defined in section V A within the technological class

[b] Difference in logged average real R&D spending per worker between the receiving and the originating region (1991–1996)

Specification I: Dependent variable: log of citations between patents with citing and cited patents less than 10 years apart during 1975–1996. Only top 20 regions for R&D spending included as cited regions.

Specification II: Dependent variable: log of citations between patents with citing and cited patents less than 6 years apart during 1975–1996. Only top 20 regions for R&D spending included as cited regions.

Specification III: Dependent variable: log of citations between patents with citing and cited patents less than 2 years apart during 1975–1996. Only top 20 regions for R&D spending included as cited regions.

Specification IV: Dependent variable: log of citations between patents with cited patents granted in 1975–1985. Only top 20 regions for R&D spending included as cited regions.

TABLE 5.—EFFECT OF EXTERNAL ACCESSIBLE R&D ON INNOVATION

Specification: Variable	I Flows from Top 20 Innovators	II Flows from All Regions	III Flows from Top 20 Innovators	IV Flows from All Regions
Measure of weights, ϕ_{ij}	From equation (1)	From equation (1)	$(c_{ij}/P_j)/(c_{ii}/P_i)$	$(c_{ij}/P_j)/(c_{ii}/P_i)$
In A_{it}, own R&D	0.74*	0.81*	0.65*	0.60*
	(0.02)	(0.03)	(0.02)	(0.02)
In A^a_{it}, external accessible R&D	0.47*	0.40*	0.51*	0.54*
	(0.05)	(0.05)	(0.04)	(0.04)
Country × year effects	Yes	Yes	Yes	Yes
Period	1975–1993	1975–1993	1975–1993	1975–1993
R^2	0.90	0.88	0.86	0.85
Observations	2,024	2,486	2,024	2,486

Notes: Method of estimation: OLS with country × year fixed effects. Heteroskedasticity-robust standard errors in parentheses. An asterisk indicates significance at 1% level.

Specification I: Dependent variable: log of patents weighted by citation in first 4 years since granted. External accessible R&D constructed using the estimated intensity of knowledge flows from table 4, column 1. These estimates capture geographical flows of knowledge within 10 years (long run). Only the top 20 world innovators were included as senders. Only the remaining 93 regions were included as receivers. Countries covered: the United States, west Germany, the United Kingdom, Italy, Spain, France, the Netherlands, and Canada.

Specification II: Dependent variable: log of patents weighted by citation in first 4 years since granted. External accessible R&D constructed using the estimated intensity of knowledge flows from table 2, column 1. These estimates capture geographical flows of knowledge within 10 years (long run) from all regions. All 113 regions included as senders as well as receivers. Countries covered: the United States, West Germany, the United Kingdom, Italy, Spain, France, the Netherlands, and Canada.

Specification III: Dependent variable: log of patents weighted by citation in first 4 years since granted. External accessible R&D constructed using the standardized citation frequency from region i to region j, $(c_{ij}/P_j)/(c_{ii}/P_i)$, as weights. Only the top 20 world innovators were included as senders. Only the remaining 93 regions were included as receivers. Countries covered: the United States, West Germany, the United Kingdom, Italy, Spain, France, the Netherlands, and Canada.

Specification IV: Dependent variable: log of patents weighted by citation in first 4 years since granted. External accessible R&D constructed using the standardized citation frequency from region i to region j, $(c_{ij}/P_j)/(c_{ij}/P_i)$, as weights. These estimates capture geographical flows of knowledge within 10 years (long run) from all regions. All 113 regions included as senders as well as receivers. Countries covered: the United States, West Germany, the United Kingdom, Italy, Spain, France, the Netherlands, and Canada.

0.74–0.81, and the elasticity of innovation to accessible external R&D is estimated at 0.40–0.47. The contribution to innovation of external accessible R&D is large and very significant.

Column III and IV report the estimates of the innovation function using weights ϕ_{ij} calculated directly from standardized citation frequencies across regions. From the definition (6), assuming that propensities to patent (β_j) are constant across regions and that random errors (ε_{ij}) can be ignored, we obtain the relationship $\phi_{ij}/\phi_{ii} = (c_{ij}/P_j)/(c_{ii}/P_i)$. With the standardization $\phi_{ii} = 1$ we can use this formula, along with data on citations c_{ij} and patents P_i, to calculate directly the weights ϕ_{ij} for each regional couple. The estimates of own R&D elasticity are now somewhat smaller (0.60–0.65), and the estimates of the effect of external accessible R&D are somewhat larger (0.50–0.55). Overall, however, the results are similar across specifications, and the two sets of estimates give an elasticity to own R&D in the range 0.6–0.8, and an elasticity to accessible external R&D in the range 0.4–0.5. The estimates of own R&D elasticity are similar to those found in Branstetter (2001) (0.72), Pakes and Griliches (1980) (0.61), and several other studies. The estimates of the elasticity to external accessible R&D are between 50% and 80% of the elasticities to own R&D, which is within the range of the existing estimates from the micro-productivity literature (see Griliches, 1992).

VII. Conclusions

Trade is clearly not the only conduit of national and international knowledge flows. Indeed, there is much to be learned about knowledge diffusion at the frontier of technological innovation from the very large and detailed data set on citations across patents, developed and used extensively by the micro-productivity literature. The present work uses data on citations across patents generated in Europe and North America to construct knowledge flows across 147 regions and to estimate the effect of several resistance factors on learning. We obtain robust estimates that show that only 20% of the knowledge generated in the average region flows out of it. Moreover, another 36% drop in learning takes place when crossing the next regional border, and yet another 20% drop when passing the country border. However, two important qualifications apply. First, we find that ideas in the information and communication technologies (computers and electronics) diffuse much farther than average knowledge. Second, we find that technological leaders (the top 20 regions for total R&D) generate knowledge that also tends to diffuse farther. One advantage of our approach is that determinants of knowledge flows are estimated using a gravity-like equation and therefore could be compared quantitatively with those of trade flows. It turns out that knowledge flows are much less localized than trade flows. Finally, to confirm that these flows are relevant to regional innovative activity, we estimate the effect of accessible external knowledge on innovation, finding that the external accessible stock of R&D has an effect on the innovation of regions 50% to 80% as large as that of their own R&D stock.

REFERENCES

Adams, J., "Comparative Localization of Academic and Industrial Spillovers," *Journal of Economic Geography* 2:3 (2002), 253–278.

Anderson, J., and E. Van Wincoop, "Gravity with Gravitas: A Solution to the Border Puzzle," *American Economic Review* 93:1 (2001), 170–192.

Bernstein, J. I., and P. Mohnen, "International R&D Spillovers between U.S. and Japanese R&D Intensive Sectors," *Journal of International Economics* 44:2 (1998), 315–338.

Bernstein, J. I., and M. I. Nadiri, "Interindustry R&D Spillovers, Rate of Return and Production in High Tech Industries," *American Economic Review, Papers and Proceedings* 78:2 (1989a), 429–434.

———, "Research and Development and Intra-industry Spillovers: An Empirical Application of Dynamic Duality," *Review of Economic Studies* 56 (1989b), 249–269.

Bernstein, J. I., and X. Yan, "International R&D Spillovers between Canadian and Japanese Industries," *Canadian Journal of Economics* 30:2 (1997), 276–294.

Blomstrom, M., and A. Kokko, "Multinational Corporations and Spillovers," *Journal of Economic Surveys* 12:3 (1998), 247–277.

Branstetter, L., "Looking for International Knowledge Spillovers: A Review of the Literature with Suggestion for New Approaches," *Annales d'Economie et de Statistique* no. 49–50 (1998), 517–540.

———, "Is Foreign Investment a Channel of Knowledge Spillovers? Evidence from Japan's FDI in the United States," NBER working paper no. 8015 (2000).

———, "Are Knowledge Spillovers International or Intranational in Scope? Microeconometric Evidence from the U.S. and Japan," *Journal of International Economics* 53:1 (2001), 53–79.

Bresnahan, T. F., "Measuring Spillovers from Technical Advance," *American Economic Review* 76 (1936), 741–755.

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

Coe, D., and E. Helpman, "International R&D Spillovers," *European Economic Review* 39 (1995), 859–887.

Coe, D., E. Helpman, and A. Hoffmeister, "North-South R&D Spillovers," *Economic Journal* 107:4 (1997), 139–149.

Conley, T., and E. Ligon, "Economic Distance and Cross-Country Spillovers," *Journal of Economic Growth* 7:2 (2002), 157–187.

Eaton, J., and S. Kortum, "Trade in Ideas: Productivity and Patenting in the OECD," *Journal of International Economics* 40 (1996), 1–28.

Feenstra, R., *Advanced International Trade: Theory and Evidence* (Princeton, NJ: Princeton University Press, 2003).

Griliches, Z. "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature* 28:4 (1990), 1661–1707.

———, Z. "The Search for R&D Spillovers," *Scandinavian Journal of Economics* (Supplement 1992) 94 (1992), 29–47.

Grossman, G., and E. Helpman, *Innovation and Growth in the Global Economy* (Cambridge, MA: MIT Press, 1991).

Hall, B. H., A. B. Jaffe, and M. Trajtenberg, "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," National Bureau of Economic Research working paper no. 8498 (2001).

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

Jaffe, A., and M. Trajtenberg, *Patents, Citations and Innovations* (Cambridge, MA MIT Press, 2002).

Jaffe, A., M. Trajtenberg, and R. Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* 108:3 (1992), 577–598.

Jozefowicz, J. "Estimating University Spillovers with Count Panel Data," *Pennsylvania Economic Review* 11:1 (2002), 58–75.

Keller, W. "Trade and the Transmission of Technology," *Journal of Economic Growth 7* (2002a), 5–24.

———, "Geographic Localization of International Technology Diffusion," *American Economic Review* 92 (2002b), 120–142.

Krugman, P. "A Model of Innovation, Technology Transfer and the World Distribution of Income," *Journal of Political Economy* 65 (1979), 69–75.

Madden, G., S. Savage, and J. P. Bloxham, "Asian and OECD International R&D Spillovers," *Applied Economic Letters* 8:7 (2001), 431–435.

Maruseth, P. B., and B. Verspagen, "Knowledge Spillovers in Europe: A Patent Citation Analysis," *Scandinavian Journal of Economics* 104:4 (2002), 531–543.

Mansfield, E., J. Rappoport, A. Romeo, S. Wagner, and J. Beardsley, "Social and Private Rates of Returns for Industrial Innovation," *Quarterly Journal of Economics 77* (1977), 221–240.

Mohnen P. "R&D Externalities and Productivity Growth," *STI Review,* no. 18, 39–66 (1996).

Pakes, A., and Z. Griliches, "Patents and R&D at the Firm Level: A First Report," *Economic Letters* 5 (1980), 377–381.

Peri, G., "Knowledge Flows, R&D Externalities and Innovation," Zentrum für Eurepäische Wirtschaftsforschung discussion paper no. 40/03 (2003).

Rivera-Batiz, L., and P. M. Romer, "Economic Integration and Endogenous Growth," *Quarterly Journal of Economics* 106:2 (1991), 227–244.

Romer, P. M., "Endogenous Technological Change," *Journal of Political Economy* 98:5 (1990), 71–102.

Scherer F. M., "Using Linked Patents and R&D Data to Measure Interindustry Technology Flows," in Z. Griliches (Ed.), *R&D, Patents and Productivity* (University of Chicago Press, 1984).

Schmookler, J., "Invention and Economic Development," University of Pennsylvania PhD dissertation (1951).

Terlecky, N., "Direct and Indirect Effects of Industrial Research and Development on the Productivity Growth of Industries," in J. B. Kendrik and B. N. Vaccara (Eds.), *New Developments in Productivity and Measurement and Analysis* (University of Chicago Press, 1980).

Wolf, E. N., and M. I. Nadiri, "Spillover Effects, Linkage Structure, Technical Progress and R&D," New York University C.V. Starr Center research report 87/43 (1987).

APPENDIX A

Regions

Austria: Ostösterreich, Sudösterreich, Westösterreich.

Belgium: Bruxelles, Vlaams Gewest, Regione Wallonne.

Canada (provinces): Newfoundland, Prince Edward Island, Nova Scotia, New Brunswick, Quebec, 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, Mediterranée.

Germany (Länder): Baden-Wurttemberg, Bayern, Berlin, Brandenburg, Bremen, Hamburg, Hessen, Mecklenburg-Vorpommern, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, Sachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen.

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.

Luxembourg: Luxembourg.

Norway: Norway.

Portugal: Portugal.

Spain: Noroeste, Noreste, Comunidad de Madrid, Centro, Este, Sur, Canarias.

Sweden: Sweden.

Switzerland: Regione Lemanique, Espace Mittelland, Nordwestschweiz, Zürich, Ostchweiz, Zentralschweiz, Ticino.

United Kingdom: North, Yorkshire and the Humber, East Midlands, East Anglia, Southeast, Southwest, West Midlands, Northwest, Wales, Scotland, Northern Ireland.

United States of America (states): Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, D.C., Florida, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming.

APPENDIX B

Patent Classes

Chemical: agriculture, food, textile, coating, gas, organic compounds, resins, miscellaneous chemicals.

Computers: communications, computer hardware and software, computer peripherals, information storage.

Drugs: drugs, surgical and medical instruments, biotechnology, miscellaneous medical.

Electronics: electrical devices, electrical lighting, measuring and testing, nuclear and x rays, power systems, semiconductors, miscellaneous electronics.

Mechanical: material processing and handling, metal working, motors and engines, optics, transportations, miscellaneous mechanical.

Other: agriculture husbandry and food, amusement devices, apparel, earth working and wells, furniture, heating, pipes and joints, receptacles, miscellaneous other.