Innovation for Development

Teresa de Noronha Vaz (Ed.)

Centro de Investigação sobre o Espaço e as Organizações
Centro Regional de Inovação do Algarve
Universidade do Algarve

The geography of knowledge spillovers in Europe

Manfred M. Fischer

Introduction

In recent years, economic growth theorists have focused new attention on the role of knowledge capital in aggregate economic growth, with a prominent modelling role for knowledge spillovers (see, Romer 1986; Grossman and Helpman 2001; and Fischer 2009). For the purpose of this chapter, knowledge spillovers may be defined to denote the benefits of knowledge to firms not responsible for the original investment in the creation of this knowledge. It is convenient to distinguish two types of knowledge spillovers: spillovers embodied in traded capital or intermediate goods and services (so-called pecuniary externalities), and knowledge spillovers of the disembodied kind.

The focus in this chapter is on disembodied knowledge spillovers which arise because knowledge is a partially excludable and non-rivalrous good. Lack of excludability implies that it is difficult for firms to fully appropriate the benefits from their knowledge generation activities and prevent others from using the knowledge without compensation. While knowledge is subject to spillovers, it is only imperfectly excludable. Non-rivalry, on the other hand, implies that a novel piece of knowledge can be used many times and in many different circumstances.

Knowledge spillovers are notoriously difficult to measure, as pointed out by Krugman (1991, p. 53): "[k]nowledge flows… are invisible, they leave no paper trail by which they may be measured and tracked". But Jaffe, Trajtenberg and Henderson (1993) argue that knowledge spillovers may well leave a paper trail in the citations to previous patents recorded in patent documents.

This contribution lies in the research tradition that uses patent citations as a proxy for knowledge spillovers, and directs attention to knowledge spillovers within the high-technology sector. High-technology is defined in our context to include the ISIC-sectors (ISIC Rev. 2) pharmaceuticals (ISIC 3522), computers and office equipment (ISIC 3825), electronics-telecommunications (ISIC 3832), and aerospace (ISIC 3845). Though some firms may choose not to patent inventions, patenting in high-technology industries is commonly practiced and indeed a vital component of maintaining competitiveness. The European coverage of this paper is given by patent applications at the European Patent Office [EPO] that are assigned to high-technology firms located in the EU-27 member states (except Cyprus and Malta), and Norway and Switzerland.

The chapter summarizes previous research by the author and research associates published in recent years (Fischer, Scherngell and Jansenberger 2006, 2008). The structure is as follows. The second section explains in some detail the nature of patents and patent citations, briefly discusses how patent citations can be used as an indicator for knowledge spillovers, and elaborates on the patent citation data to be used. The third section shifts attention to the geographic dimension to the spillover mechanism and tests for spillover-localization. This is a most difficult problem due to the difficulty of separating spillovers from correlations that may be due to a pre-existing pattern of geographic concentration of technologically related activities. The fourth section suggests that both geographic and technological distance attenuate interregional knowledge spillovers from innovative activity. The results presented here indicate a tendency for knowledge spillovers to localize conditional on the technological relation between spillover generating and receiving region.

Patents, Patent Citations and Knowledge Spillovers

A patent is a property right awarded to inventions for the commercial use of a newly invented device. An invention to be patented has to satisfy three patentability criteria. It has to be *novel* and *non-trivial* in the sense that it would not appear obvious to a skilled practitioner of the relevant technology, and it has to be *useful*, in the sense that it has potential commercial value. If a patent is granted, an extensive public document is created. The document contains detailed information about the technology of the invention, the inventor, the assignee that owns the patent rights, and the technological antecedents of the invention. Because patent documents record the residence of the inventors they are an important resource for analyzing the spatial extent of knowledge spillovers, as captured by patent citations.

Patent related data, however, have two important limitations. *First*, the range of patentable inventions constitutes only a subset of all R&D outcomes. Purely scientific advances devoid of immediate applicability as well as incremental technological improvements which are too trite to pass for discrete, codifiable inventions are not patentable. The second limitation is rooted in the fact that is a strategic decision. It may be optimal for inventors *not* to apply for patents even though their inventions would satisfy the criteria for patentability (Trajtenberg 2001). Firms balance the time and expense of the patent process, and the possible loss of secrecy which results from patent publication, against the protection that a patent potentially provides to the inventor (Jaffe 2000). Thus, patentability requirements and incentives to refrain from patenting limit the scope of our analysis based on patent data.

Patents from different national patent offices are not comparable to each other because of different patent breadth, patenting costs, approval requirements, citation practices and enforcement rules across Europe. This makes patent data from the European Patent Office [EPO] rather than national patent offices a natural choice for our study. Our data source is the EPO database. The data on patent applications that we use in this study were drawn from the universe of European patents. By *European* patents, we mean patents assigned to corporations located in Europe, regardless of the nationality of the inventors. Our sample of patents is limited to those that are related to inventions in the high-technology industries or in other words to those patents assigned to patent classes which match the high-technology sector, at the four-digit level of the International Standard Industrial Classification, ISIC Rev. 2. We used MERIT's concordance table (see Verspagen, Moergastel and Slabbers 1994) between the four-digit ISIC-sectors and the 628 patent subclasses[1] of the International Patent Code [IPC] classification to identify the high-technology patents from the universe of European patent applications.

The database for this study contains all the high-tech patents applied at the EPO between 1985-2002, totalling 177,424 patents. Each patent application produces a highly structured public document containing detailed information on the invention itself, the technological area to which it belongs, the inventor and her/his address, and the organisation to which the inventor assigns the patent property right. By nature of the research question, we are interested in the geographical location of the inventor rather than the applicant and hence, use the postal code of the inventor address for tracing inventive activities back to the region of knowledge production.

For representing geographic space we use 188 regions that cover the EU-27 countries (excluding Cyprus and Malta) plus Norway and Switzerland. Their definition is based on the Nomenclature des Unites Territoriales Statistiques [NUTS]. The regions are essentially in line with the NUTS-2 level of the regional classification in the case of Austria, Belgium, Germany, Finland, France, Italy, The Netherlands, Portugal, Spain, Sweden and UK, and in line with the NUTS-0 level in all other cases.

Patent documents include references or citations to patents. These citations open up the possibility of tracing multiple linkages between inventions, inventors, firms and locations. In particular, patent citations enable us to analyse the geographical extent of spillovers. There are, however, also some serious limitations to the use of patent citation data. Patent citations capture only those spillovers which occur between patented inventions, and, thus, underestimate the actual extent of knowledge spillovers. Other channels of disembodied knowledge diffusion – for example, transfer of knowledge embodied in skilled labour, knowledge transfer between customers and suppliers, knowledge exchange at conferences and trade fairs – are not captured by patent citations. Patent citations do not always represent what we typically think of as knowledge spillovers. Some citations may represent only indirect knowledge spillovers since the patent examiner added them. This noise creates a bias against finding spillovers. Fortunately, bias in this direction is a problem of power which can be overcome with a sufficiently large sample size (Thompson 2003).

Patent citation is a phenomenon that derives from the relationship between two inventions or inventors as evidenced by a citing patent and a cited patent. The data on this relationship come in the form of citations *made* (that is, each patent lists references to previous patents). For identifying the citation flows we need a list of cited and associated citing patent applications. This requires access to all citation data in a way that permits efficient search and extractions of citations not by the patent number of the citing patent but by the patent number of the cited patent. In constructing the patent citation data set that forms the basis of our study we begin with the full set of issued patents that have their application year between 1985 and 2002. There are 177,424 high-technology patents. We then discard all patents that have not received any citations, leaving 101,247 patents which generate 210,667 citations.

The observation of citations is evidently subject to a truncation bias because we observe citations for only a portion of the life of an invention, with the duration of that portion varying

1. The IPC system is an internationally agreed, clear-cut non-overlapping hierarchical classification system that consists of five hierarchical levels. At the third level 628 subclasses are distinguished.

across patent cohorts. This means that patents of different ages are subject to different degrees of truncation. To overcome this problem we have used the approach of moving windows of five years and thus identified all the pairs of cited and citing patents where citations to a patent are counted for a window of five years following its issuance. This window of five years seems to be appropriate since the mean citation lag of 210,667 citations is 4.6 years. The analysis is, thus, confined to 1985-1997 in the case of cited patents while citing patents appearing in 1990-2002 are taken into account. This process reduces the number of patents to 69,814 that generate 155,462 citations. Next, we discard so-called self-citations, identified as assignee matches, because self-citations do not represent knowledge spillovers in the sense of externalities. This yields 98,191 citations or observations that link a citing patent to a cited patent.

The unit of analysis is the dyad 'cited patent-citing patent'. A single originating patent, for example, that has two inventors and is cited by three subsequent patents will generate six unique observations. Figure 7 displays the 98,191 observations where each patent is assigned to one of the 188 regions based on the home address of the inventors as reported in the patent document. The nodes represent the regions, their size is relative to their spillover generating power measured in terms of citations received.

Figure 7 - Knowledge flows between European regions, as captured by interfirm patent citations in the high-technology sector, 1985-2002 [see Figure 2 in Fischer, Scherngell and Jansenberger 2008]

Geographic Localization of Knowledge Spillovers

It is widely recognized that knowledge – once generated – spills over among firms. But the geographical extent of such knowledge spillovers is greatly contested. In this section we make an effort to test for spillover-localization. This is a most difficult problem due to the difficulty of separating spillovers from correlations that may be due to a pre-existing pattern of geographic concentration of technologically related activities without resort to localization of knowledge spillovers (see Agrarwal, Cockburn and McHale 2003).

Patents linked by citations not only share a technology, but are often also developed by inventors working in a common industry. Patents linked by a citation are, thus, much more likely to share a geographic location than a pair of patents drawn at random from the entire pool of patents. To control for the tendency of inventive activities to be geographically clustered, we follow the case-control matching approach pioneered by Jaffe, Trajtenberg and Henderson (1993). The essence of this approach is to compare citing patents with control patents in terms of the frequency with which each is located in the same region as the originating patent. A finding of a disproportionate number of co-located citations relative to co-located control patents is interpreted as evidence of localized knowledge spillovers.

To derive a control frequency that is immune to contamination from localization based on the pre-existing concentration of technological activity, we went back to our patent database and identified a control patent to corresponding to each of the citing patents. For each citing patent, we identified all patents in the same patent class (measured in terms of the three-digit IPC code) with the same application year, excluding any other patents which cited the same originating patent. We then chose from this set of patents a control patent whose application date was as close as possible to that of the citing patent. This process generated, for each set of citing patents, a corresponding control sample of equal size, whose distribution across time and technological fields (defined by the 120 patent classes) is essentially identical to that of the citation data set.

Each control patent is paired with a particular citing patent. This allows us to compare the geographic location of the control patent with that of the originating patent cited by its counterpart in the citing data set. The frequency with which these control patents match geographically with the originating patent is an estimate of the frequency with which a randomly drawn patent which is not a citation, but has the same temporal and technological profile as the citation.

When we calculate the frequency with which the citations match the geographic location of the cited patents, we are estimating the probability of geographic matching for two patents, conditional on these being a citation link and conditional on the timing and technological nature of the citation. When we compute the frequency with which the control patents match geographically with the cited patents, we are estimating the probability of geographic matching for two patents, conditional on the timing and the technological nature of the citation. If the citation match frequency is significantly higher, this implies that citations are localized even after controlling for technology and timing (Jaffe, Trajtenberg and Henderson 1993).

We consider two cohorts of originating patents with corresponding sets of citing patents and control patents to test for spillover-localization. One consists of 1990 patent applications and the other of the 1995 ones. The 1990 cohort of originating patents contains 2,118 patents that have received a total of 2,362 citations including and 1,410 citations excluding self-citations by the end of 1995. The 1995 cohort of originating patents contains 1,814 patents that have received a total of 2,387 citations including and 1,366 citations excluding self-citations by the end of 2000.

Table 12 summarizes the results for both cohorts of originating patents. Localization effects are reported at two spatial levels: the regional and the country level of analysis. 'Number of citations' corresponds to the number of citations cited by the originating cohort of patents. 'overall citation matching', 'citation matching excluding self-cites' and 'control matching' are the percentage of cited patents [with and without self-citations] and controls that belong to the same geographic location as the originating patent. The *t*-statistic tests the equality of the control proportions and the citation proportions (excl. self-citations).

	1990 - Originating Cohort	**1995 - Originating Cohort**
Number of Citations		
incl. Self-Cites	2,362	2,387
excl. Self-Cites	1,410	1,366
Matching by Country		
Overall Citation Matching [%]	60.1	61.2
Citation Matching excl. Self-Cites [%]	36.6	35.9
Control Matching [%]	21.9	25.4
t-Statistic (excl. Self-Cites)	8.68 (p = 0.00)	6.01 (p = 0.00)
Matching by Region		
Overall Citation Matching [%]	36.7	37.0
Citation Matching excl. Self-Cites [%]	13.7	14.8
Control Matching [%]	5.2	5.4
t-Statistic (excl. Self-Cites)	7.91 (p = 0.00)	8.27 (p = 0.00)

Table 12 - Geographic matching fractions [see Table 3 in Fischer, Scherngell and Jansenberger 2008]

Note: The t-statistic tests equality of the citation proportion excluding self-citations and the control proportion. See text for details.

The first column of Table 12 reports the 1990 results. Starting with the country match, we find that citations *including self-citations* are intranational about 38 percent points more often than the controls. Excluding self-citations cuts this difference roughly in half. The remaining difference between the citations excluding self-citations and the controls is strongly significant statistically. Looking at the 1990 results for regions (see the lower part of the table), we find that citations of patents come from the same region about 37 percent of the time. Excluding self-citations, however, makes a big difference. The proportions are cut to 13.7 percent. The matching frequency excluding self-citations is significantly greater than the matching control proportion.

The results for patent citations of 1995 patents, given in the second column of the table, are similar. For both cohorts of originating patents and for both geographical levels, the patent citations are quantitatively and statistically significantly more localized than the controls. The citation matching percentages (excl. self-cites) slightly rise at the regional level from 13.7 percent in 1990 to 14.8 percent in 1995, but slightly decrease at the country level from 36.6 percent to 35.9 percent. It is, however, impossible to tell from this comparison whether this represents a real change, or whether it is the result of differences in average citation lags. The average citation lag for the 1990 (1995) cohort of originating patents is 4.45 (4.57) compared to 4.14 (4.51) for the corresponding control patents.

The results on the extent of localization can be summarised as follows. For citations observed by 1,410 of the 1990 originating cohort of patents, there is a clear pattern of localization at the regional and country levels. Citations are about seven times more likely to come from the same region than control patents, 2.6 times more likely excluding self-cites. They are 2.7 times more likely to come from the same country as the originating patents, and 1.7 times excluding self-cites. For citations of 1995 originating patents, the same pattern emerges. All these differences are statistically significant at a level much less than one percent.

It is worth noting that localization of knowledge spillovers is not a universal phenomenon. European regions reveal different patterns in the local diffusion of knowledge externalities. Table 13 presents the results for selected regions including Île-de-France, Oberbayern, Switzerland, Noord-Brabant, Darmstadt, Lombardia and Bedfordshire which account for about one third of the inventive activities in high-technology industries in Europe, as measured in terms of EPO patent activities over 1985 to 2002. For the samples, there are significantly higher proportions of citation matches than control matches (except Noord-Brabant in 1995). Results that are significant at the 0.05 level or better are given in bold. These results indicate quite strongly that knowledge is localized at the regional level. In 1995 Île-de-France shows by far the strongest localization effect. The results for the German regions (Oberbayern, Darmstadt), Switzerland and Bedfordshire are also significant in 1990 and 1995.

	Number of Citations [excl. Self-Cites]		Citation Matching [%]		Control Matching [%]		*t*-Statistic[1]	
	1990	1995	1990	1995	1990	1995	1990	1995
Île-de-France	130	197	27.9	28.4	13.9	8.6	**3.30 (0.000)**	**6.05 (0.000)**
Oberbayern	82	88	12.1	10.2	2.4	2.4	**2.22 (0.009)**	**1.51 (0.037)**
Switzerland	73	81	17.8	28.3	9.5	6.1	**1.51 (0.046)**	**3.81 (0.000)**
Lombardia	68	43	26.4	16.2	7.3	11.6	**3.38 (0.000)**	0.70 (0.242)
Noord-Brabant	65	14	24.6	7.1	13.8	7.1	**1.72 (0.044)**	0.00 (0.500)
Darmstadt	53	76	11.3	28.9	0.2	3.9	**1.93 (0.029)**	**3.95 (0.000)**
Bedfordshire	36	13	46.1	23.0	5.5	0.0	**3.21 (0.001)**	**1.89 (0.042)**

Table 13 - Regional variations in localization: A test in selected regions [see Table 4 in Fischer, Scherngell and Jansenberger 2008]

Note:[1] Results significant at the 5 percent level of significance are in bold.

The Geographic and Technological Dimensions to the Spillover Mechanism

In the previous section we analyzed the extent to which citations by patents to previous patents are geographically localized, relative to a baseline likelihood of localization based on the predetermined pattern of technological activity. This section directs attention to the geographic and technological dimensions to the spillover mechanism and adopts a spatial interaction modelling perspective on knowledge spillovers as evidenced by patent citations (see, Fischer, Scherngell and Jansenberger 2006).

The spatial interaction modelling perspective shifts attention from individual patent citations to interregional patent citations, or in other words, from the dyad "cited patent – citing patent" to the dyad "cited region – citing region". Correspondingly, all citation data were aggregated into a region-by-region matrix, denoted by $[(c_{ij})]$, where c_{ij} denotes the number of patent citations from region j (j=1, ..., N, here N=188) to region i for i=1, ..., N. The rows of the matrix represent the origin location of the spillovers (in other words, the region of the cited patents) and the columns the destination location (the regions of the citing patents). Note that the matrix is asymmetric in nature, that is, for $c(i, j) \neq c(j, i)$ for $i \neq j$.

	Number of Matrix Elements*	Patent Citations				
		Number	Mean	Standard Deviation	Min.	Max.
All Elements	35,344	98,191	2.77	16.23	0	1,408
Intraregional Links	188	11,371	60.48	152.05	0	1,408
Interregional Links	35,156	86,820	2.46	11.14	0	351
Positive Interregional Links	11,468	86,820	7.57	18.49	1	351
National Interregional Links	3,952	25,341	6.41	20.02	0	351
International Interregional Links	31,204	61,479	1.97	9.31	0	290

* Elements of the region-by-region citation matrix

Table 14 - Descriptive statistics on the region-by-region patent citation matrix [see Table 1 in Fischer, Scherngell and Jansenberger 2006]

In the case of cross-regional inventor-teams we have used the procedure of multiple full counting that does justice to the true integer nature of patent citations, but gives – in comparison to the procedure of fractional counting – interregional cooperative inventions greater weight. Table 14 provides some basic information about the 188-by-188 citation matrix that contains 35,344 elements with a total of 98,191 citations between high-technology firms. The mean number of citations between any two regions (including intraregional flows) is 2.77, but the standard deviation is rather high. Interregional citations ($i \neq j$) show a highly skewed distribution. About two thirds of all pairs of regions never cite each other's patents. The frequency of patent citations gradually declines for more intensive citation links. There are only 90 pairs of regions for which the number of citations is about one hundred or more. The average number of citations for all interregional pairs is 2.46 and the average for those that cite each other 7.57. Table 14, moreover, indicates that national patent citations are more frequent than international ones.

The elements of the N-by-N patent citation matrix denote observations on random variables, $c(i, j)$, each of which corresponds to flows of knowledge from region i to region j. The $c(i, j)$s are assumed to be independent random variables. They are sampled from a specific probability distribution that is dependent upon some mean, say $\mu(i, j)$. Let us assume that no a priori information is given about the row and column totals of the flow matrix $[c(i, j)]$. Then the mean interaction frequencies between origin i and destination j may be modelled by

$$\mu(i, j) = const\ A(i)^{\alpha}\ B(j)^{\beta}{}_{j}\, F(i, j) \tag{1}$$

where $\mu(i, j) = E[c(i, j)]$ is the expected flow, *const* denotes a constant term, the quantities $A(i)$ and $B(j)$ are called origin and destination factors or variables, respectively, α and β indicate their relative importance, and $F(i, j)$ is a separation factor that constitutes the very core of spatial interaction models. Following Sen and Sööt (1981), we specify the separation factor in form of a multivariate exponential deterrence function

$$F(i, j) = \exp\left[\sum_{k=1}^{K} \theta_k\, {}^{k}d(i, j)\right] \tag{2}$$

where ${}^{k}d(i, j)$ is the kth measure of separation between i and j, and θ_k the associated parameter. We assume that the observed flows follow a Poisson distribution with

$$P\{c(i, j)\} = \frac{\exp\,[-\mu(i, j)]\ \mu(i, j)^{c(i, j)}}{c(i, j)!} \tag{3}$$

where $P\{.\}$ denotes probability, and the expected value, $\mu(i, j)$, is given by Eq. (1). Equation (3) models patent citations flows between origin i and destination j as inter-point movement counts. Hence, this is the specification of a discrete distribution.

Subject to caveats relative to the relationship between patent citations and knowledge spillovers (see Section 2), this Poisson spatial interaction model allows us to identify and measure separation effects for interregional knowledge spillovers in Europe. Our interest is focused on K=4 separation measures. ${}^{1}d(i, j)$ represents geographic distance between regions i and j in terms of the great circle distance (in km) between their economic centres; ${}^{2}d(i, j)$ technological distance measured in terms of dissimilarity in a multidimensional technological space spanning 55 individual patent classes[2]; ${}^{3}d(i, j)$ and ${}^{4}d(i, j)$ are dummy variables representing border effects and language barriers between region i and j.

2. Each region is assigned a 55-by-1 technology vector that measures the share of patenting in each of the technological classes for a region. The technological proximity index, denoted by s, between region i and j is given by the uncentred correlation of their technological vectors. Two regions that patent exactly in the same proportion each patent class have an index equal to one, while two regions patenting only in different classes have an index equal to zero. This index is appealing because it allows for a continuous measure of technological distance by the transformation ${}^{2}d(i, j)=1-s(i, j)$.

The product $A(i)$ $B(j)$ in Eq. (1) may be interpreted simply as the number of distinct (i, j)-interactions that are possible. Thus, a reasonable way to measure the origin factor is in terms of the number of patents in the knowledge producing region i in the time period 1985-1997, and the destination factor in terms of the number of patents in the knowledge absorbing regions j in the time period 1990-2002.

Table 15 reports the results from the estimation of the Poisson spatial interaction model by maximum likelihood (ML), using Newton-Raphson. The ML-estimates of the Poisson model specification given by Eq. (3) are reported in the first column, those of a generalized Poisson model specification in the second[3]. Standard errors are presented in parentheses rather than *t*-statistics to allow comparison with the precision of the generalized model specification. The reported standard errors all assume correct specification of the variance function. They are characterized by low significance levels.

	Poisson Spatial Interaction Model	
	without Heterogeneity	with Heterogeneity
Intercept	-10.278*** (0.051)	-10.881*** (0.124)
Origin Variable [α]	0.833*** (0.002)	0.915*** (0.006)
Destination Variable [β]	0.858*** (0.002)	0.885*** (0.006)
Geographical Distance [θ_1]	-0.270*** (0.005)	-0.321*** (0.014)
Technological Proximity [θ_2]	-0.928*** (0.032)	-1.219*** (0.130)
Country Border [θ_3]	-0.050*** (0.007)	-0.533*** (0.046)
Language Barrier[θ_4]	-0.238*** (0.014)	-0.031*** (0.043)
Dispersion Parameter [δ]	–	0.725 (0.014)
Log-likelihood	-51,801.10	-37,235.05
$\{\mathrm{Corr}\,[c(i, j), \text{predicted } c(i, j)]\}^2$	0.686	0.783

*** denotes significance at the one percent level.

Table 15 - Estimation results of the Poisson spatial interaction model with asymptotic standard errors in parentheses [see Table 2 in Fischer, Scherngell and Jansenberger 2006]

Note: All independent metric variables are expressed log form in order to lessen the impact of outliers.

3. The Poisson model specification given by Eq. (3) does not allow for individual (i, j) effects, given the exogenous variables $A(i)$, $B(j)$ and $F(i, j)$. It is clear, however, that the existence of fixed effects at the individual level of (i, j) pairs is likely to exist in interregional patent citation relationships. This individual effect problem can be partly solved by introducing a heterogeneity term in the mean $\mu(i, j)$ of the Poisson distribution such that the multiplicative heterogeneity term follows a gamma distribution with mean one and variance δ. This modification yields the so-called heterogeneous Poisson model of interregional patent citations that allows for overdispersion (*i.e.* $\delta > 0$) and subsumes the Poisson model specification given by Eq. (3) if $\delta = 0$.

The estimated value of the dispersion parameter δ indicates that the basic Poisson model specification has to be rejected (H_0: δ =0, G^2 =29,256.6, $p<0.01$). The rejection of this model version is due to the situation of overdispersion, which is associated with unobserved heterogeneity among (i, j)-pairs of regions. Therefore, the Poisson model specification with heterogeneity is preferred. The variance-mean equality assumption of the basic Poisson model is too restrictive to adequately describe the patent citation flows.

This model specification with heterogeneity yields highly significant effects. Both α and β estimates are – in accordance to expectations – close to one. Geographical distance between inventors has a strong and negative effect on the likelihood of high-technology patent citations. The parameter estimate, $\hat{\theta}_1$ = -0.321 , indicates that for any additional 100 km between regions i and j the (i, j)-mean patent citation frequency decreases by 27.5 percent. This suggests spillovers between high-technology firms are impeded by geographical distance.

Not only distance, but also border effects matter. The point estimate of the coefficient $\hat{\theta}_3$ is nearly twice times as large as that of $\hat{\theta}_1$, showing that border effects are more important than distance effects. Citing patents are much more likely to come from the same country as the cited patents. High-technology related knowledge flows much more easily within than between countries. Note that language barriers, though significant, have only a rather small effect ($\hat{\theta}_4$=-0.031) on interregional knowledge spillovers.

The variable technological proximity controls for spillovers that are stronger between technologically similar regions. The point estimate for the variable shows an effect that is about four times larger than the distance effect even though the estimate is not very precise. Interregional patent citation flows tend to follow particular technological trajectories as defined at the three-digit level of the IPC classification system. This indicates that patent citation flows are industry-specific and occur most often between regions not too far located from each other in technological space. Technological proximity matters more than geographical proximity.

Conclusion

It is widely recognized that disembodied knowledge – once generated – spills over among individuals, firms and regions. But the geographical range of such knowledge spillovers is greatly contested. This chapter lies in the research tradition that uses patent citations as a proxy for knowledge spillovers, and directs attention to knowledge spillovers within the high-technology sector in Europe. The European coverage is given by patent applications at the EPO which are assigned to high-technology firms located in Europe.

Using the case-control matching approach – pioneered by Jaffe, Trajtenberg and Henderson (1993) – we find strong evidence of geographic localization at two different spatial levels (region, country) even after controlling for the tendency of innovative activities in the high-technology sector to be geographically clustered. The findings not only indicate that knowledge localization exists in the aggregate, but that there are also variations of localization by region.

The results obtained in a spatial interaction modelling framework indicate the tendency for knowledge spillovers to localize is conditional on the technological relation between spillover generating and receiving firms and regions. The analysis results presented in Section 4 are consistent with intuition and existing empirical evidence which suggests that both geographical and technological distance attenuate knowledge spillovers. But it is important to note that disembodied knowledge flows more easily within European countries than across and that technological proximity tends to overcome geographical proximity. Interregional knowledge flows seem to follow particular technological trajectories and occur most often between regions with similar technological profiles.

References

Agrawal, A., I. M. Cockburn and J. McHale (2003): Gone but not forgotten: Labor flows, knowledge spillovers and enduring social capital. NBER Working Paper 9950, Cambridge [MA].

Borgatti, S. P., M. G. Everett and L.C. Freeman (1999): Ucinet for Windows: Software for social network analysis, User's guide. Analytical Technologies, Harvard [MA].

Fischer, M.M. (2009): A spatially augmented Mankiw-Romer-Weil model: Theory and evidence. Paper submitted for publication in the *Annals of Regional Science.*

Fischer, M.M., Scherngell, T. and Jansenberger, E. (2008): Geographic localisation of knowledge spillovers. Evidence from high-tech patent citations in Europe. Accepted for publication in the *Annals of Regional Science.*

Fischer, M.M., Scherngell, T. and Jansenberger, E. (2006): The geography of knowledge spillovers between high-technology firms in Europe: Evidence from a spatial interaction modelling perspective. *Geographical Analysis* 38 (3): 288-309.

Grossman GM and Helpman E (1991): Innovation and growth in the global economy. MIT Press, Cambridge [MA].

Jaffe, A. B. (2000): The U.S patent system in transition: Policy innovation and the innovation process. *Research Policy* 29(5): 531-57.

Jaffe A B, Trajtenberg M and Henderson R (1993): Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 108(3): 577-598.

Krugman P (1991): Geography and trade. MIT Press, Cambridge [MA].

Romer PM (1986) Increasing returns and long-run growth. *Journal of Political Economy* 94(5): 1002-1037.

Sen, A. and Sööt, S. (1981): Selected procedures for calibrating the generalized gravity model. *Papers of the Regional Science Association* 48: 165-176.

Trajtenberg, M. (2001): Innovation in Israel 1968-1997: A comparative analysis using patent data. *Research Policy* 30(3): 363-89.

Verspagen B, Moergastel T van and Slabbers M (1994): MERIT concordance table: IPC - ISIC (rev. 2). Maastricht Economic Research Institute on Innovation and Technology, University of Limburg, Maastricht.