



Munich Personal RePEc Archive

Similarity Of RD Activities, Physical Proximity, and The Extent Of RD Spillovers

Deltas, George and Karkalakos, Sotiris

University of Illinois at Urbana-Champaign, University of Piraeus

1 April 2007

Online at <https://mpra.ub.uni-muenchen.de/45962/>
MPRA Paper No. 45962, posted 09 Apr 2013 11:22 UTC

SIMILARITY OF R&D ACTIVITIES, PHYSICAL PROXIMITY, AND THE EXTENT OF R&D SPILLOVERS¹

by

George Deltas* and Sotiris Karkalagos**

This Version: April 19, 2007

Abstract: The diffusion of knowledge generates positive externalities if knowledge flows increase the productivity of Research and Development (R&D) by the recipients of these flows. This paper investigates the extent to which this positive spillover effect of knowledge diffusion depends on the similarity of research activities by the originator and recipient of the knowledge. The paper also investigates at what rate these spillover effects diminish as the distance between the originator and recipient increases. We find, using regional patent and R&D expenditure data from the European Union, that similarity between R&D activities is not only statistically significant, but salient: regions with completely dissimilar R&D activities exhibit essentially no spillovers at all. We also find an increase in the distance between the originating and recipient region by 550 kilometers reduces spillovers by 75% (as low as 55% in some specifications). Unlike much of the extant literature, the rate of spatial decay of spillovers is estimated jointly with the remaining parameters of the model rather than through specification searches over a set of alternative weight matrices. Our results are robust to the inclusion of unobserved country effects and border barriers.

Keywords: Technological similarities, diffusion, spatial effects.

JEL Classification: C2, O4, R0.

*Department of Economics, University of Illinois at Urbana-Champaign, 1206 South Sixth Street Champaign, IL 61820, USA; e-mail: deltas@uiuc.edu

**Department of Economics, University of Exeter, Streatham Court, Rennes Drive, Exeter EX4 4PU, UK; e-mail: S.Karkalagos@exeter.ac.uk

¹ The authors would like to thank Christos Kotsogiannis and seminar participants in CRETE 2006 for helpful comments and discussion.

1. Introduction

Alfred Marshall (1890) argues that industrial agglomerations exist in part because individuals learn from each other when they live and work in close proximity, and increasing amounts of evidence confirms this. It is now well documented that there are substantial geographical spillovers in R&D activities and that these have important theoretical implications (e.g., Aghion and Howitt, 1992, Romer, 1990), though there is still considerable debate on the measurement of these spillovers. Economists have assumed that diffusion of ideas depends on proximity in space, technological specialization, the stage of economic development, labor mobility, and a multitude of other factors (Acs and Varga, 2002). These are plausible assumptions, but each of them requires explicit testing. An all-encompassing estimation framework, though, is impossible, and thus each empirical study focuses on a particular facet of technological spillovers and employs an estimation approach that is tailored to that task.

This study focuses on the relationship between the similarity in the research activities of any pair of regions and the extent of spatial technological spillovers between them. This investigation can shed light on the nature of spatial spillovers. If the spillovers are due to the local availability of relevant know-how embodied in human capital, the spatial correlation between R&D successes would be stronger between regions with relatively similar R&D successes (or activities), and conversely the spatial correlation between R&D successes would be weaker between regions with dissimilar R&D successes (or activities). This pattern of spatial correlation would be absent if the spillovers were driven by general rather than specific knowledge, or if the spatial correlation between R&D successes were spurious.

This paper also makes a contribution on the methodological front. In particular, we employ a non-linear spatial econometric framework in which the spatial decay (spatial lag) matrix is a

function of a parameter that measures the rate at which the spatial effects dissipate. We assume an exponential rate of decay, so that an increase in physical distance by a given increment leads to (a further) decrease in spillovers by a fixed percentage. This exponential decay parameter is jointly estimated with all the other parameters of the model. In other words, we do not employ a set of possible spatial lag matrices and choose among them on the basis of some measure of fit, as usually done in the related literature; rather we estimate the spatial lag matrix jointly with the all the other parameters in the model. Our approach avoids the possibility of under-estimating the standard errors in spatial models, since the standard errors are not conditional on the correct choice of the spatial lag matrix.

As an auxiliary contribution, we investigate the robustness of the spillover estimates to the incorporation of the similarity effects on the spatial interactions. The presence of substantial differences in the estimates would suggest that ignoring the role that similarities play in spillovers could lead to incorrect inferences. Thus, we contribute to the findings on the robustness (or non-robustness) of spatial model estimates to the employment of different types of spatial weight matrices.

We utilize data from the European Union at the district level from the late 1990's. As a proxy for R&D inputs and the generation of technological ideas we use the financial outlays in R&D, while as a proxy of technological output we use the number of patents issued in a region. Our findings show that a similarity of research activities between regions is effectively a precondition for substantial technological spillovers between them. In the absence of any similarity, spillovers are small. Moreover, technological spillovers decay at a rate of 75% for every 550 kilometers of distance. Thus positive spillovers extent beyond a day's driving range, and thus are not due solely to due the Griliches' notion of exchanging ideas "over breakfast." Rather, much of

these spillovers are likely due to interaction in regional conferences and meetings, through hiring of personnel, and through supplier-client relationships.² However, spillovers at a distance of over 1,000 kilometers do become rather negligible. Finally, our findings are robust to the inclusion of border and country fixed effects, except (in some specifications) for a reduction in the estimated rate of spatial decay of technological spillovers.

Knowledge flows have been extensively analyzed in the literature following Griliches (1992) and Jaffe (1989).³ Several papers followed their outline and enhanced our understanding of the process of knowledge diffusion (see recent survey by Breschi and Lissoni, 2001, on knowledge sources and their linkages with diffusion). A widely used approach postulates a knowledge production function and assumes that knowledge flows exist between firms with similar technology. This framework has been broadly used in the United States (Acs, Anselin, Varga, 2002 and Varga, 2000), Europe (Fisher and Varga, 2001 and Fritsch, 2002), Asia (Evenson and Singh, 1997), or for multi-regional studies (Bernstein and Mohnen, 1998, and Bernstein and Yan, 1997). Not only do these studies support the fundamental conclusions of Griliches (1990) and Jaffe (1989), but they also vary widely in their methodology (spatial vs non-spatial), structure of the data (cross section vs. panel data), the set of variables, and conceptual issues (e.g., distinction among different types of knowledge sources). Moreover, the choice of geographical unit definition varies: most of the studies use regions according to geographical criteria but some employ economic criteria.⁴

² See Levin (1988) for the perceived importance of these factors as conduits for spillovers by managers of US firms.

³ Jaffe (1989) is the first to find a positive spatial relationship between the number of patents and R&D activities, though his estimation approach only considered spillovers within a geographical unit (as opposed to spillovers from neighboring geographical units). This relationship is interpreted as the proof of the existence of “technological spillovers” at firm’s level data and provides the motivation of technological similarities in this paper.

⁴ Some recent work focuses on the microfoundations of these spillovers. Almeida and Kogut (1999) investigate the premise that they are driven to a large extent by the mobility of workers across firms, while Glaeser (1999) uses a theoretical model that builds the microfoundations of a particular type of spillover effect and investigates its implications. The model can easily be adapted to regional knowledge spillovers.

Despite being predominant in terms of frequency of use, the knowledge production function approach is not the only methodology for investigating and measuring knowledge spillovers. A second approach is proposed by Jaffe, Trajtenberg and Henderson (1993) who use the distribution of patents in different regions and measure spatial autocorrelation through the number of citations between any two regions. In related work, Peri (2005) uses patent citation data as a proxy for flows of knowledge across regions, while Hall et al. (2005) provide direct evidence that patents that are more highly cited generated more value (profit) for the firms that own them (and, surprisingly, that self citations are more valuable than citations by other firms).⁵ Both approaches posit that patent counts are a reasonable measure of research output, a premise that is supported by Trajtenberg (1990) who also argues for weighing patents by citation counts.⁶

Closest in spirit to this work are the papers by Bottazzi and Peri (2003) and Peri (2005). The former uses European regional data to investigate the effect of “technological proximity” on R&D spillovers, though this is done as an additional control variable and not part of the distance weight matrix as we do. Moreover, in that paper the authors pre-specify the distance weight matrices (as the rest of the literature), rather than estimate them directly from the data, as done in this paper. The second of these two papers investigates the effects of technological proximity on citations rather than on patents, with citations in turn being an input (among others) in the production of patents. In this paper, too, technology is a control rather than an element of a distance weight matrix in a spatial econometric framework.

⁵ See also Thomson and Fox-Kean (2005) for some criticism of Jaffe, Trajtenberg and Henderson (1993), based on possibly excessive aggregation in the partitioning of relevant patent classes.

⁶ See also Keller (2002) for a relationship between the spatial distribution of R&D and productivity.

2. Econometric Model

Many economists have drawn from the large pool of patent data and used them as a convenient measure of research output (Griliches, 1990). A standard empirical specification measuring the extent of spatial spillovers in R&D successes (patents) estimates the relationship

$$\ln(P_j) = a + b \ln(R_j) + c \sum_{i=1}^I w_{ij} \ln(R_i) + dZ_j \quad (1)$$

where P_j is the number of patents in region j , w_{ij} is an element of spatial weight matrix W that depends on the distance between regions i and j (with $w_{ij} = 0$ when $i = j$), R_{ij} is R&D expenditure in region i , and Z_j is a set of regressors. Elements of the weight matrix W are typically the inverse distance between two regions, or take the value of 1 for regions that are closer than a pre-specified cut-off and the value of 0 for regions that are further than the pre-specified cut-off. There is infinity of possible choices for the weight matrix (Bruckner 1998).⁷ The use of the log transformation allows for the interpretation of the parameter estimates as elasticities and is often also justified by variable distributions that are approximately lognormal (i.e., by variables that are positive and whose distribution has, as in our data, a long right tail).

In our work, we augment the above framework in two ways. First, we allow the spillovers from the R&D of region i to region j to depend not only on the measure of physical distance between the two regions, but also on the similarity of research activities between the regions. We do so by incorporating a measure of similarity of research activities directly into the distance weight function. We define a similarity index, S_{ij} , of research activities of regions i and j using the difference in the distribution of patent shares over K different industries or sectors. Formally, the similarity of patents between regions i and j is

⁷ The *generally correct* weight-matrix does not exist (Anselin and Bera, 1998). In the spatial econometric literature a variety of different weight matrix choices are typically made, often within a single paper.

$$S_{ji} = 1 - \left[\sum_k \left| \frac{P_{jk}}{P_j} - \frac{P_{ik}}{P_i} \right| \right] / K \quad (2)$$

where P_{jk} is the number of patents in region j and sector k . By construction, this index ranges from 0 to 1, is symmetric (i.e., $S_{ij} = S_{ji}$) and is independent of the aggregate number of patents in regions i and j . In particular, if the sectoral distribution of patents is the same for two regions, then the value of similarity index is equal to 1, whereas if all the patents of a region j are in sectors for which region i has no patents, the value of the similarity index is equal to 0. The index is not defined if one of the regions has no patents, but this does not occur in our sample.

Our second departure from the standard framework is that the elements of the spatial weight matrix, w_{ij} , are not taken as constants in the estimation, but are rather assumed to be an estimable function of distance. In particular, we assume that $w_{ij} \propto e^{-\theta d_{ij}}$, where d_{ij} is the physical distance between regions i and j and θ is a parameter to be estimated. We also allow (in some specifications) the spatial weights to depend on whether regions i and j are on the same side of a border. Thus, our general specification framework is given by

$$\ln(P_j) = a + b \ln(R_j) + \sum_{i=1}^I w_{ij}(\theta, c, d_{ij} S_{ij}, B_{ij}) \ln(R_i) + dZ_j \quad (3)$$

where B_{ij} takes the value of 1 if regions i and j are in the same side of border and c is a vector of parameters to be estimated. We estimate many variants of equation (3). In the more parsimonious variants, which form our base results, the weight matrix is not a function of border effects, and there are no explanatory variables in the vector Z_j . Equation (3) then takes the form

$$\begin{aligned}
\ln(P_j) &= a + b \ln(R_j) + \sum_{i \neq j} e^{-\alpha l_{ij}} (c_0 + c_1 S_{ij}) \ln(R_i) \\
&= a + b \ln(R_j) + c_0 \sum_{i \neq j} e^{-\alpha l_{ij}} \ln(R_i) + c_1 \sum_{i \neq j} e^{-\alpha l_{ij}} S_{ij} \ln(R_i)
\end{aligned} \tag{4}$$

As robustness tests, we estimate a number of variants of this above specification. Country and border effects feature prominently in these additional specifications. If research activities within each country were more similar in terms of industrial sector allocation (if only because countries specialized in different industries), then one might detect a correlation between the similarity of activities and spillovers that is driven solely by differential patent productivity across different countries. The inclusion of country fixed effects would eliminate this problem by controlling for any unobserved differences in the patent productivity across countries. The inclusion of border effects captures any discontinuities in the spatial decay function due to linguistic and labor market barriers at country borders. One such specification with both country and border effects is given by

$$\ln(P_j) = a + b \ln(R_j) + c_0 \sum_{i \neq j} e^{-\alpha l_{ij}} \ln(R_i) + c_1 \sum_{i \neq j} e^{-\alpha l_{ij}} S_{ij} \ln(R_i) + c_2 \sum_{i \neq j} B_{ij} \ln(R_i) + dZ_j \tag{5}$$

where Z_j is a vector of country dummies. A number of variants of the above equation have been estimated, some including interaction of border effects with similarity of research activities between regions, others weighing the border effects by distance using the exponential distance weights. These variants are described in more detail in section 4 below.

3 Data and Variables.

The patent information, industry classification, and R&D expenditure data is from the Cronos data series of Eurostat Statistics. Our sample comes from the 1995-1999 period. We

partition the European Union into 146 regions on the basis the territorial units as identified by Eurostat, known as Nomenclature Units Territory Statistics (NUTS).⁸ These regions are internally rather homogeneous, often have a strong local identity, are administrative units in the countries they belong to, and have some degree of policy independence. Their geographical extent is shown in the maps in Figures 1 and 2. As a measure of the innovative output of a region we use the count of granted patent applications to inventors located in each region. The Eurostat data attributes each patent to the first inventor listed in the patent application, as it is generally done in this literature (see Jaffe, Trajtenberg and Henderson, 1993). Patents have long been considered, not without controversy, as the best measure of output of the innovative activity. Although not all inventions are patented, the patented ones have to fulfil minimal standards of novelty, originality and potential use. Therefore patents can be considered as a good approximation to ideal data on “economically profitable ideas” which one would like to have for testing theories on innovation. Though an aggregate patent count forms our dependent variable, Eurostat classifies patents into one of five industrial categories: Chemistry/Chemical-related, Electricity/Electromechanical, Transportation, Biology/Medicine/Human-related, and Other. We construct the similarity matrix on the basis of the distribution of patents in these five sectors in each of the districts. The mean and standard deviation of the similarity between any random pair of regions are approximately equal to 0.68 and 0.20, respectively. There is also substantial variability in the *average* similarity of any given region’s patent output with that of *all* the other regions of the EU: the standard deviation of this measure is equal to 0.13.

R&D expenditure data include expenditures that “comprise creative work undertaken on a systematic basis in order to increase the stock of knowledge, including knowledge of man,

⁸ For the time period in our sample, the membership of the European Union consisted of 15 countries. However, insufficient (or no) data was available for the three latest entrants (Austria, Sweden, and Finland) and for Belgium. The 146 NUTS units in our sample pertain to the remaining 11 countries of the EU.

culture and society and the use of this stock of knowledge to devise new applications” (European Commission, 2001). The R&D expenditure data is measured in millions of euros, and is aggregate rather than sectoral, as a breakdown by sector is not available in the EuroStat database. Given that R&D expenditure in any particular year yields results (e.g., patents) with some lag and these results are typically spread over time, we use as the dependent variable the average number of patents issued in the last 3 years of our sample (1997-99) and as R&D the average of the first 3 years of our sample (1995-97).⁹ Given that the year 1996 is characterized by missing data for half of the regions, we chose to drop it in computing the average R&D. The average number of patents per region is 294, with a standard deviation of 456, while the average R&D expenditure is 762 million euros, with a standard deviation of 1,338 million euros.¹⁰

4 Estimation Results

We estimate and report results from 12 different models that are variations of equations (4) and (5). We have also estimated a few additional variations of these models obtaining similar results, but we do not report them here as they would provide effectively zero marginal value. The sampling distributions of most parameter estimates are non-symmetric, and thus we construct confidence intervals based on bootstrapping (801 replications). We report the standard 95% confidence interval, but also indicate in the tables parameter significance at the 10% level.

In all tables, $WlnRD$ denotes the distance weighted variable $\sum_{i \neq j} e^{-\theta d_{ij}} \ln(R_i)$, $WSlnRD$ denotes the

⁹ The similarity matrix is computed using our dependent variable, and is thus based on the average patent output of each region over this 3 year period.

¹⁰ Two regions are characterized by zero expenditure, and thus we add 1 to both R&D and patents before taking the log of the dependent and independent variables. Given the scale of the data, this does not materially affect our interpretation of the parameters as elasticities.

distance and similarity weighted variable $\sum_{i \neq j} e^{-\theta l_{ij}} S_{ij} \ln(R_i)$, $BlnRD$ denotes border effects on spillovers that are not distance weighted, i.e., it denotes the variable $\sum_{i \neq j} B_{ij} \ln(R_i)$, $BSlnRD$ border effects that are similarity but not distance weighted, i.e., it denotes the variable $\sum_{i \neq j} B_{ij} S_{ij} \ln(R_i)$, and $WBSlnRD$ denotes border effects that are both similarity and distance weighted, i.e., it denotes the variable $\sum_{i \neq j} e^{-\theta l_{ij}} B_{ij} S_{ij} \ln(R_i)$.

We first estimate, as a point of departure, equation (4) omitting the last term, i.e., omitting the industry similarity variable (Table 1, Model 1). The elasticity of patents to own R&D is approximately 0.82, significantly different from constant returns to scale at the 5% level. Returns to scale to own R&D investments may indeed be diminishing; however it is also possible that R&D expenditures may yield benefits with a long enough lag that they are not reflected in patent output during our sample period. The coefficient on the spatially weighted R&D of other regions is strongly statistically significant, but small in size. An increase in the R&D of other firms by 1%, even if those are effectively co-located and of zero distance, leads to an increase in patent output of the target region by slightly less than 0.01%. Measured spillovers are, in this sense, small. Moreover, the exponential rate of spatial decay is relatively fast: The estimate of the decay parameter θ is equal to 0.2719 per unit of distance, with distances measured in degrees on the Earth's surface (one degree is equal to 111km). Simple calculations using the exponential formula show that spillovers are halved for every 283km of distance increment between two locations (i.e., the spatial half-life of R&D spillovers is 283km). Nonetheless, the aggregate effect of all spillovers is quite substantial because any European region receives positive

spillovers from all other European regions. A meaningful statistic is to compute what is the effect on a region's patent output if all other regions increase their R&D expenditures by one percent. We compute this for every region and report the average value over all regions for each model as the Mean Spatial Effect. For Model 1, this is equal to 0.163. Adding this to the coefficient of $\ln RD$, we see that if all EU regions increased R&D expenditures by 1%, research output would increase by $0.813+0.163=0.976$, which is very close to constant returns to scale, when these are evaluated at the EU level, rather than the regional level.¹¹

In Model 2 we estimate equation (4) which allows the elements of the distance weight matrix to depend on the degree of similarity of activities between any pair of regions. The similarity weighted spillovers are not only statistically significant and but also larger in magnitude than the coefficient of $W\ln RD$ in Model 1. Moreover, once the similarity weighted spillovers are accounted for, the coefficient of $W\ln RD$ becomes statistically insignificant at the 5% level and of the wrong sign. Thus, an increase in the R&D activities in a region that is characterized by a completely different set of activities than its regional neighbors confers no positive externality to those neighbors (as we see below, the sign changes with the inclusion of unobserved country heterogeneity). The estimates of the parameter θ and the Mean Spatial Effects are slightly smaller than those of Model 1. When $W\ln RD$ is dropped from the regression (Model 3), the point estimate of $W\ln RD$ remains statistically significant and the model fit (as measured by R-squared) drops only marginally. In contrast, Model 2 has a statistically significant better fit than Model 1 as measured by an F-test. Thus, on the basis of the first three models we conclude that similarity of research activities is not only important for the presence of spatial spillovers, but also salient: In its absence, spillovers are essentially zero.

¹¹ We have not at this point computed standard errors for the sum of $\ln RD$ coefficient and the Mean Spatial Effect, but, given the parameter estimates, the standard error of $\ln RD$ is effectively a lower bound.

In Table 2 (Models 4, 5, and 6) we include border effects to account for the possibility that the similarity effects on the spatial weight matrix are driven by a similarity of research activities within countries. More generally, estimated spatial spillovers may just be an artifact of the possibility that there are spillovers between regions within a country (possibly with no spatial decay) and no spillovers between regions in different countries. To control for these possibilities, three different variants of border effects are added to the equation (4): A border effect that is not affected by the distance of regions within each country (Model 4), a border effect that adjusts the spatial weight matrix of R&D spillovers (Model 5), and a border effect that adjusts the spatial weight matrix of similarity weighted R&D (Model 6). In none of these models are border effects statistically significant, not even at the 10% level.¹² All other coefficients are essentially the same as in Model 2.

Table 3 augments the regressions reported in Table 2 through the addition of country fixed effects. These fixed effects capture any differences in the effectiveness of R&D in leading to patented output. Such differences may be due to a number of factors, not all which can be captured through the use of controls such as education levels, GNP per capita, etc. Such factors may include differences in the experience of R&D personnel, differences in the cost of obtaining R&D inputs, differences in the propensity to patent R&D output (rather than keep it from the public domain), differences in the complexity of research undertaken within each country (some types of patents are the outcome of a very resource-intensive effort, while others may be extensions of existing work), and others. The country fixed effects are statistically significant on the basis of F-tests for all models and all conventional measures of significance. The addition of

¹² As shown in McCallum (1995) and confirmed by Helliwell (1998), migration and trade flows are much more intense between regions of the same country than of different ones. However, this may not necessarily be true for flows of ideas. Nonetheless, our findings on this regard differ from Peri (2005) who investigates citation patterns and incorporates flows from non-European regions, but more in line with Botazzi and Peri (2002) who find, using a different econometric framework, border effects that are typically not statistically significant (though positive).

the fixed effects does indeed have a measurable effect on all of the R&D coefficients: all of them are now of smaller magnitude than those of Table 2. As a consequence, *WSlnRD* narrowly misses significance at the 5% level (it is easily statistically significant at the 10% level). The sign of *WlnRD* turns positive (an easier finding to explain) but is not statistically significant at any conventional level. Thus, the inclusion of country fixed effects weakens the results without altering their basic nature: A region's research output is most responsive to its own R&D expenditure (elasticity of approximately 0.7) but exhibits diminishing returns to scale. The R&D of regions that have completely dissimilar research activities does not result in any measurable spillovers; these spillovers increase with the similarity in the research activities and this increase is statistically significant. In aggregate these spillovers are important: Increasing the R&D of all (other) EU regions by 1% increases research output in the average EU region by 0.25-0.29% depending on specification. Notice that the aggregate effect of these spillovers is higher than those reported in Table 1, despite the fact that point estimates of the coefficients of *WlnRD* and *WSlnRD* are smaller than those of Table 1.¹³ The reason is that the inclusion of country fixed effects noticeably decreases the point estimate of θ to between 0.15 and 0.16. These lower point estimates indicate that the spatial half-life of R&D spillovers increases to 500km, thus raising their quantitative importance.

In a final set of results we remove the border effects that are consistently not statistically significant and undertake a more systematic model selection among regressions with and without similarity effects. In particular, in this last set of regressions (reported in Table 4) we re-estimate the models in Table 1 with the addition of country fixed effects. Though *WSlnRD* in Model 11 barely misses statistical significance at the 5% level, the fit of Model 12 which drops *WlnRD* is

¹³ The spillover effects of other regions' R&D on patent output are about half those of own region's R&D. This is almost as high as the estimate in Peri (2005), who considers both intra-European and trans-Atlantic spillovers.

basically the same as that of Model 12, and substantially higher than Model 10 which drops $WSlnRD$ (point estimates are essentially identical to those in Table 3). Thus, the balance of evidence in the regressions with country fixed effects is that (i) similarity-weighted spillovers provide a noticeably better description of the spatial relationship between R&D and patent generation than do spillovers that ignore the similarities in the pattern of research activities, (ii) incorporating unweighted spillover effects in a model that includes similarity-weighted spillovers does not improve that model's explanatory power, while (iii) incorporating similarity weighted spillover effects on a model with unweighted spillover effects increases that model's explanatory power (at the 10% significance level).

5. Discussion and Concluding Remarks

Our analysis can be seen as a step in integrating two strands of the technology literature: The first focuses on spillovers at the firm level as a function of the type of innovation undertaken by these firms; the second looks at the regional aggregate level of innovation and research activities and focuses on spillovers as a function of physical distance between regions and other barriers. Both Bottazzi and Peri (2003) and Peri (2005) are steps in the same direction, though our approach differs in that it directly incorporates the effects of similarity of research activities into the spatial lag matrix rather than use them as an additional control in the knowledge production function or treat them as components of an intermediate input in the production of the new knowledge. Though computationally intensive, our approach eliminates the need for specification searches through the use of different alternative weight matrices in spatial econometric models. In this framework, we allow the distance-weighted R&D spillovers from the research activity in EU regions on the patenting activity of their neighbours to depend on the

similarity of their research activities. Such similarities appear to be very important: Regions with dissimilar activities appear to be characterized by essentially no spillovers.¹⁴

Our estimates of similarity effects are certainly statistically significant and robust across specifications in terms of magnitude of aggregate spillovers (though the relative contribution of the key factors varies across some specifications). But does the incorporation of similarity of research activities on the spatial decay matrix have economically meaningful effects in the geographic pattern and magnitude of spillovers in the European Union? This is examined in Figures 1 and 2 which depict the geography of spatial spillovers effects with and without technological similarities. Figure 1 shows the percentage predicted increase in a region's patent output if all other regions in the EU experience an one percent increase in their R&D when similarity effects are ignored (Model 1). Higher values of spatial effects are observed in northern and central European regions, as these regions are in areas with dense economic activity. However, taking similarities into consideration leads to a substantial "revision" of the predicted level of spillovers. Figure 2 plots the change in the predicted level of spillovers when we use Model 3 rather than Model 1 to measure spillovers.¹⁵

For some of the regions, the changes in the estimated spillovers through the incorporation of similarity effects are negative: these are regions which undertake research activities that are different than those of their neighbors and thus the standard approach over-estimates the extent of spillovers they receive. Most of these regions are in southern Europe, where estimated decreases in spillovers can be of the order of 40%. Thus, the standard approach of estimating

¹⁴ Jaffe (1989) finds that, within a U.S. state, spillovers from academic research to the industry are limited to within technical areas, with essentially no spillovers across technical areas.

¹⁵ Somewhat stronger results are obtained when comparing the more fully parametrized Model 2 with Model 1. Similar results are obtained when comparing Models 10 and 12: the mean spillovers are higher than those of Model 1 and 3, but the correlation in the predicted spillovers of Models 10 and 12 is even weaker than the correlation between the predicted spillovers of Model 1 and 3 (0.974 versus 0.980).

spillovers overestimates the degree of spillovers in many regions that are already thought to be low. However, some regions in Germany also experience downward revision of the estimated spillovers (e.g., Halle experiences a 15% decrease). Apparently, many German regions are characterized by some specialization at the individual region level; for most the revision is around zero. Positive changes in the estimated spillover effects are most pronounced in France, the UK, Northern Italy, Andalusia and the Eastern fringes of Germany. These regions are characterized to a larger extent by research activities that are similar to those of the neighbors.

Finally, our findings have implications for the location of research centers and the agglomeration pattern of research activity. If it is indeed true, as we find, that the extent of spatial R&D spillovers between two regions depends on the similarity of research activities undertaken by firms in these regions, then it follows that firms will choose to locate in regions (or near regions) in which similar firms are currently located.¹⁶ Thus, not only there will be spatial agglomeration of R&D activities, but there will also be spatial *specialization* of such activities.¹⁷ Indeed, there is already some corroborating evidence. For instance, Head, Ries, and Swenson (1995) show that industry level agglomeration benefits had an important effect on the location decisions of Japanese plants in the US, with both intermediate input provision and technological/information externalities being driving forces. Further analysis of the dynamic relationship between firm location decisions and research spillovers that are related to the similarity of activities is an important avenue of future research.

¹⁶ Abstracting from any similarity effects, the importance of spatial connections, and their linkages with information diffusion, interaction, communication and innovation has received wide empirical support. For example, Audretsch and Feldman (1996) industries with high levels of innovative activity have more tendencies to cluster. The localized character of knowledge diffusion, as proxied by patent citations, is well documented in Jaffe et al. (1993).

¹⁷ This also has implications for the location of firms as a function of the differentiation of their products (see Piga and Poyago-Theotoky, 2005, for a theoretical analysis of this issue). Though firms may prefer to choose R&D approaches to minimize such spillovers (Kamien and Zang 2000), Wiethaus (2005) provides compelling arguments to the contrary.

References

- Acs Z., Anselin L. and Varga A., (2002) "Patents and innovation Counts as Measures of Regional Production of New Knowledge," *Research Policy*, 31, 1069-1085.
- Acs Z. and Varga A., (2002) "Geography, Endogenous Growth, and innovation," *International Regional Science Review*, 25, 132-148.
- Aghion P. and P. Howitt (1992) "A Model of Growth through creative Destruction", *Econometrica*, 60: 323-51.
- Almeida P. and Kogut B., (1999) "Localization of Knowledge Spillovers and the Mobility of Engineers in Regional Networks", *Management Science*, 45, 905-917.
- Anselin, L. Bera, A. K., (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, Aman, Giles, David E.A. (Eds.), *Handbook of applied economic statistics* (Marcel Dekker: New York)
- Audretsch, D., and Feldman, M., (1996) "R&D spillovers and the geography of innovation and production", *American Economic Review*, 86 (3), 630-640.
- Bernstein J. and Mohnen P., (1998) "International R&D Spillovers between U.S. and Japanese Intensive Sectors", *Journal of International Economics*, 44:2, 315-338.
- Bernstein J. and Yan X., (1997) "International R&D spillovers between Canadian and Japanese firms", *Canadian Journal of Economics*, 30:2, 276-294.
- Bottazi, L., and G. Peri. (2003) "Innovation and Spillovers in Regions: Evidence from European Patent Data," *European Economic Review*. 47: 687-710
- Breschi S. and Lissoni F., (2001) Knowledge Spillovers and Local Innovation Systems: A Critical Survey. *Liuc Papers n. 84, Serie Economia e Impresa 27*, 55-85.
- Brueckner J., (1998) "Testing for strategic interaction among local governments: The case of growth controls", *Journal of Urban Economics* 44, 438-467.
- Evenson R. and Singh L., (1997) Economic Growth, International Technological Spillovers and Public policy: Theory and Empirical Evidence from Asia, Discussion paper no 777, Yale University.
- European Commission, (2001) The Territorial Dimension of Research and Development Policy: Regions in the European Research Area, Directorate – General for research, European Commission, February 2001.
- Fischer M. and Varga A., (2001) Production of Knowledge and Geographically Mediated Spillovers from Universities. A Spatial Econometric Perspective and Evidence from

- Austria. 17th Pacific Conference, Regional Science Association International, Portland, Oregon [USA], June/July.
- Fritsch M., (2002) "Measuring the Quality of Regional Innovation Systems: A Knowledge Production Approach," *International Regional Science Review* 25, 86-101.
- Glaeser E., (1999) "Learning in Cities," *Journal of Urban Economics*, 46, 254-277.
- Grilliches Z., (1992) "The search for R&D spillovers," *Scandinavian Journal of Economics*, 94, 29-47.
- Grilliches Z., (1990) "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28:4, 1961-1707.
- Hall B., A. Jaffe and M. Trajtenberg (2005) "Market Value of Patent Citations," *Rand Journal of Economics*, 36, 16-38.
- Helliwell J., (1998) *How Much Do National Borders Matter?* The Brookings Institution, Washington, DC.
- Jaffe A., (1989) "Real Effects of Academic Research," *American Economic Review* 79, 957-970.
- Jaffe A. Trajtenberg M. and R. Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108, pages 577-598.
- Kamien, M. and Zang, I., (2000) "Meet me halfway: research joint ventures and absorptive capacity", *International Journal of Industrial Organization*, 18, 995– 1012.
- Keller W., (2002) "Geographic Localization of International Technology Diffusion", *American Economic Review*, 92, 120-142.
- Levin R., (1988) "Appropriability, R&D Spending, and Technological Performance", *American Economic Review*, 78, 424-428.
- Marshall A., (1890) "Principles of Economics", Macmillan.
- McCallum J., (1995) "National borders matter: Canada-US regional trade patterns," *American Economic Review* 85 (3), 615-623.
- Peri G., (2005) "Determinants of Knowledge Flows and Their Effect on Innovation," *Review of Economics and Statistics*, 87, 308-322.
- Piga C. and J. Poyago-Theotoky (2005) "Endogenous R&D spillovers and locational choice", *Regional Science and Urban Economics*, 35, 127-139.

- Romer P. M., (1990) “Endogenous Technological Change,” *Journal of Political Economy*, 98(5), 71-102.
- Thompson P. and M. Fox-Kean, (2005), “Patent Citation and Geography Knowledge Spillovers: A Reassessment”, *American Economic Review*, 95, 450-460.
- Trajtenberg M., (1990) “A Penny for Your Quotes: Patent Citations and the value of Innovations,” *RAND Journal of Economics*, 21, 172-187.
- Varga A., (2000). Local Academic Knowledge Transfers and the Concentration of Economic Activity, *Journal of Regional Science* 40, 289-309.
- Wiethaus L., (2005) “Absorptive capacity and connectedness: Why competing firms also adopt identical R&D approaches”, *International Journal of Industrial Organization*, 23, 467–481.

Table 1
Spatial R&D Externalities with and without Industry Similarity Effects

Variables	Model 1			Model 2			Model 3		
	Coefficient	95% Conf. Interval		Coefficient	95% Conf. Interval		Coefficient	95% Conf. Interval	
Constant	-0.9544	-1.9343	-0.5541	-0.7301	-1.5896	0.0233	-0.8521	-1.6663	-0.4642
lnRD	0.8129*	0.6884	0.9807	0.7818*	0.6403	0.9443	0.7917*	0.6612	0.9605
W lnRD	0.0094*	0.0055	0.0158	-0.0074^	-0.1026	0.0019			
WS lnRD				0.0201*	0.0078	0.1500	0.0111*	0.0069	0.0204
theta-hat	0.2719	0.1448	0.3807	0.2507	0.1444	0.6306	0.2472	0.1116	0.3857
R ²	0.856			0.862			0.861		
Mean Spatial Effect	0.163			0.149			0.163		

Note: Significance at 5% is denoted by *, significance at 10% indicated by ^. Significance is not indicated for the constant and for theta (which by construction cannot be zero). Bootstrapped confidence intervals (801 replications) are presented for each regression. The number of observations is equal to 146.

Table 2
Spatial R&D Externalities with Industry Similarity and Border Effects

Variables	Model 4			Model 5			Model 6		
	Coefficient	95% Conf. Interval		Coefficient	95% Conf. Interval		Coefficient	95% Conf. Interval	
Constant	-0.6758	-1.5515	0.1514	-0.6366	-1.5663	0.1455	-0.7356	-1.5504	0.1581
lnRD	0.7903*	0.6590	0.9461	0.7897*	0.6604	0.9460	0.7910*	0.6586	0.9446
W lnRD	-0.0088^	-0.1000	0.0018	-0.0086^	-0.0991	0.0017	-0.0091^	-0.0990	0.0017
WS lnRD	0.0258*	0.0087	0.1466	0.0251*	0.0086	0.1476	0.0258*	0.0086	0.1487
B lnRD	-0.0020	-0.0065	0.0011						
BSlnRD				-0.00002	-0.00006	0.00001			
WBSlnRD							-0.00007	-0.00082	0.00009
theta-hat	0.2847	0.1642	0.5738	0.277	0.1641	0.5754	0.290	0.1663	0.5863
R ² adjusted	0.865			0.865			0.863		
Mean Spatial Effect	0.129			0.174			0.165		

Note: Significance at 5% is denoted by *, significance at 10% indicated by ^. Significance is not indicated for the constant and for theta (which by construction cannot be zero). Bootstrapped confidence intervals (801 replications) are presented for each regression. The number of observations is equal to 146.

Table 3
Spatial R&D Externalities with Industry Similarity, Border, and Country Fixed Effects

Variables	Model 7			Model 8			Model 9		
	Coefficient	95% Conf. Interval		Coefficient	95% Conf. Interval		Coefficient	95% Conf. Interval	
Constant	-0.3292	-10.0707	6.6325	-0.1801	-17.1280	19.6076	-0.4232	-15.2073	7.7278
lnRD	0.6970*	0.5696	0.8687	0.6963*	0.5730	0.8696	0.6977*	0.5716	0.8668
W lnRD	0.0018	-0.0146	0.0129	0.0016	-0.0149	0.0126	0.0020	-0.0259	0.0160
WS lnRD	0.0088^	-0.0005	0.0373	0.0090^	-0.0004	0.0375	0.0087^	-0.0005	0.0505
B lnRD	-0.0012	-0.1116	0.1123						
BSlnRD				-0.00002	-0.0031	0.0024			
WBSlnRD							-0.00002	-0.0053	0.0058
theta-hat	0.1549	0.0452	0.3407	0.1563	0.0485	0.3410	0.1516	0.0286	0.3901
Country dummies	Yes			Yes			Yes		
R ² adjusted	0.908			0.908			0.908		
Mean Spatial Effect	0.256			0.276			0.290		

Note: Significance at 5% is denoted by *, significance at 10% indicated by ^. Significance is not indicated for the constant and for theta (which by construction cannot be zero). Bootstrapped confidence intervals (801 replications) are presented for each regression. The number of observations is equal to 146.

Table 4
Spatial R&D Externalities with and without Industry Effects and Country Fixed Effects

Variables	Model 10			Model 11			Model 12		
	Coefficient	95% Conf. Interval		Coefficient	95% Conf. Interval		Coefficient	95% Conf. Interval	
Constant	-0.3333	-2.4829	0.4586	-0.4634	-4.2333	0.4948	-0.2587	-1.1766	0.4476
lnRD	0.7154*	0.5879	0.8698	0.6984*	0.5708	0.8671	0.6978*	0.5710	0.8553
W lnRD	0.0101*	0.0074	0.0174	0.0020	-0.0147	0.0130			
WS lnRD				0.0087^	-0.0008	0.0358	0.0113*	0.0082	0.0196
theta-hat	0.1940	0.0811	0.2886	0.1531	0.0467	0.3345	0.1725	0.0918	0.2845
Country dummies	Yes			Yes			Yes		
R ² adjusted	0.905			0.908			0.908		
Mean Spatial Effect	0.266			0.290			0.253		

Note: Significance at 5% is denoted by *, significance at 10% indicated by ^. Significance is not indicated for the constant and for theta (which by construction cannot be zero). Bootstrapped confidence intervals (801 replications) are presented for each regression. The number of observations is equal to 146.

Figure 1 Geographical mean spatial effects without technological similarities in EU (excluding Belgium)

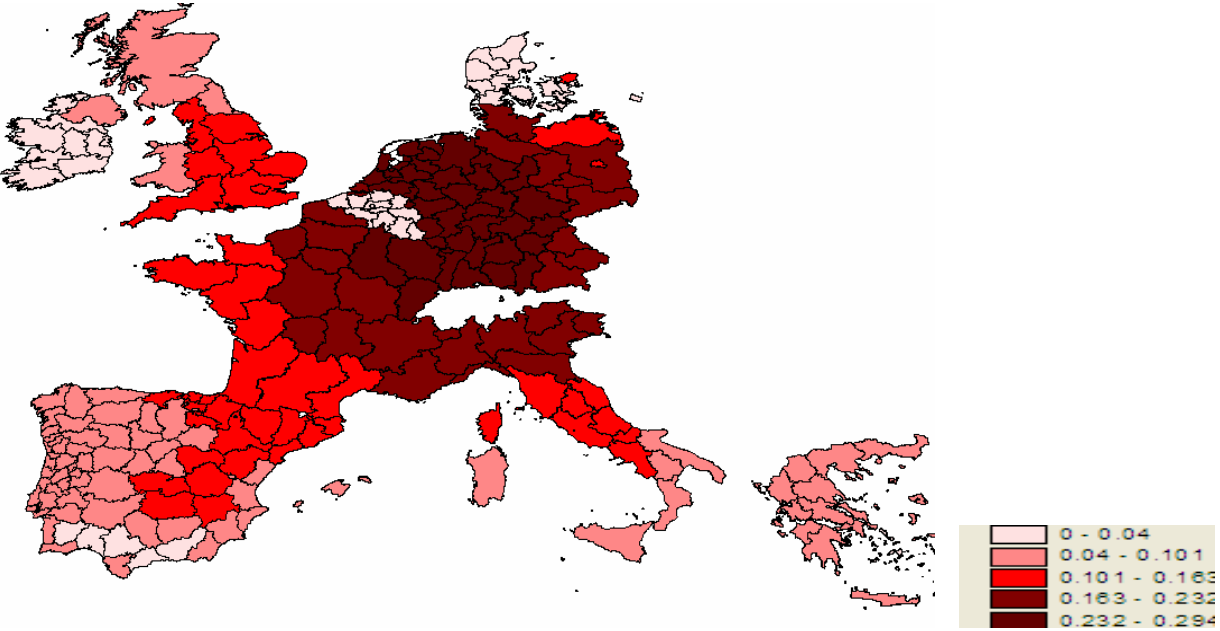


Figure 2 Percentage change of geographical mean spatial effects with technological similarities in EU (excluding Belgium).

