Marburg Geography

Working Papers on
Innovation and Space

# The structure and dynamics of R&D collaborations in Europe and the USA

## (A longitudinal and comparative perspective)

# 04.14

Sidonia von Proff, Rafael Lata and Thomas Brenner

Philipps Universität Marburg

***Impressum:***

Working Papers on Innovation and Space
Philipps-Universität Marburg

Herausgeber:

Prof. Dr. Dr. Thomas Brenner
Deutschhausstraße 10
35032 Marburg
E-Mail: thomas.brenner@staff.uni-marburg.de

Erschienen: 2014

# The structure and dynamics of R&D collaborations in Europe and the USA

# (A longitudinal and comparative perspective)

**Sidonia von Proff**[1] **and Thomas Brenner**

both from Economic Geography and Location Research, Philipps-University, Marburg.

**Rafael Lata**

Foresight and Policy Development Department, Austrian Institute of Technology (AIT), Vienna.

**Abstract:**

Today it is generally accepted that innovation, knowledge creation, and the diffusion of new knowledge are crucial factors for economic growth at the regional, national, as well as supra-national level, and that successful innovation is increasingly based on interactions and collaborative research activities between research actors. This study focuses on diverse dimensions of distance shaping R&D collaborations in Europe and the US during the time period 1999 to 2009. We take a comparative perspective by analyzing two different R&D collaboration networks (patents and publications) and two different economic areas, namely Europe and the US, in order to examine differences in collaboration activities. In particular, we investigate how the collaboration intensity between regions has been influenced by spatial, technological, and cultural distance and whether these distances have lost importance over time in the distinct networks. The study adopts a panel spatial interaction modeling perspective. In doing so, we explicitly take account of spatial autocorrelation issues of flows by means of Eigenvector spatial filtering techniques. European coverage is achieved by using 1260 NUTS-3 regions of the 25 pre-2007 EU member-states, as well as Norway and Switzerland. The US coverage is attained by using 955 core based statistical areas (CBSAs). The results reveal how collaborative knowledge creation and the spatial range of knowledge diffusion differs between Europe and the US, and provide direct evidence on the differences in cooperation patterns between different types of collaborative R&D from a longitudinal and comparative perspective

**Keywords:** R&D Networks, Patents, Publications, Spatial Interaction Modeling, Eigenvector Spatial Filtering, distance

**JEL Classifications:** C23, O38, L14, R15

# Introduction

Today it is generally accepted that innovation, knowledge creation, and the diffusion of new knowledge are crucial factors for economic growth at the regional, national, as well as supra-national level. Successful innovation is increasingly based on interactions and collaborative research activities between research actors (Wuchty et al. 2007). Despite of globalization and the increased use of information and communication technologies (ICT), geographical distance continues to be an impediment for R&D collaboration, and thus a barrier for expansive knowledge diffusion between different actors in a system of innovation. There is a strong tendency towards spatial concentration of collaboration, even though geographical proximity can be partially substituted by other forms of proximity (Agrawal et al. 2006; Boschma 2005; Breschi and Lissoni 2009). The European Union seeks to facilitate inter-regional collaboration and aims at establishing a "European Research Area", i.e. improving conditions for innovation and knowledge diffusion to ensure that innovative ideas are efficiently turned into new products and services that create growth and employment (European Commission 2011). In order to facilitate collaborative R&D activities over spatial distance, other forms of distance have to decrease, for example social distance; or spatial proximity has to be established temporarily (Torre 2008). Several policy measures have been applied in this respect during the last decade. There are some hints that indeed the importance of spatial distance has decreased in Europe since the 1990s (Cappelli and Montobbio 2013; Chessa et al. 2013; Lata et al. 2012). However, a decreasing importance of distance could also be induced by increased ICT usage, the overall trend of globalization in science, and cultural factors. Up to now, in the US, no such policy measures for enhancing collaboration over distance have been established. Since the US is the economic area most similar to Europe in terms of culture and economics compared with other economic areas in the world, the US could serve as a suitable comparison for the development over time. In addition, research on inter-regional innovation collaboration is a specific European topic, so that up to now there are hardly any comparable studies available for the US.

The paper at hand investigates whether the importance of spatial and further forms of distance have developed differently in the US and Europe during the last decade. In detail, the aim of our research is threefold: Firstly, we analyze the development over time, i.e. investigate whether the importance of distance for innovation/scientific collaboration has decreased during the last decade. Secondly, we distinguish between different types of distance in order to see whether there is a development in the importance of certain types of distance while others remain unchanged. Thirdly, we want to compare Europe and the US in order to investigate whether possible differences in the development could be policy induced. We employ a spatial interaction modeling perspective in order to identify and compare the evolution of distinct measures of distance that influence the probability for collaborative activities in different collaboration networks in Europe and the US. In order to have a strong basis for the analysis, we use two collaboration output measures: co-patent networks and co-publication networks.

The importance of the diverse forms of distance for innovative collaboration can be analyzed on several levels (Katz and Martin 1997). The micro-level can be represented by individuals (e.g. Ter Wal 2010), but distance can also be measured on the level of the firm (e.g. Autant-Bernard et al. 2007) and on a regional level (e.g. Hoekman et al. 2009), which is the approach of the paper at hand. However, we use rather small area units: the NUTS3 level in Europe and the core based statistical areas (CBSA) in the USA, which are of comparable size. By using these fine-grained spatial units regional characteristics and their influence on collaboration behavior can be measured quite detailed. For example, the technological activities in small areas is often quite homogenous, while it becomes heterogeneous if larger areas are used. Former studies (e.g. Fischer et al. 2006; Maurseth and Verspagen 2002) have been made mostly with larger spatial units (NUTS2), a few exemptions are the studies by Chessa et al. (2013), who, however, constrain their analysis on spatial distance, Hoekman et al. (2009) and Moreschalchi et al. (2014).

The remainder of the paper is structured as follows. The next section gives an overview of the literature on collaboration over distance and lists possible differences between the US and Europe, especially regarding the development over time. The third section presents our methods, the fourth section presents the datasets and descriptive statistics. Results of the regressions and their discussion follow in the fifth section before section six concludes.

# Collaboration over distance: differences between the US and the EU

### Theoretical background

The main topic of this paper is the impact of geographic distance on the probability of R&D collaboration. Especially, the analyses focus on the differences between US and EU and the changes with time in this context. Hence, we have to discuss why geographic distance should play a role for R&D collaboration. The literature provides a number of reasons, which we will discuss one after the other in the following.

First, the traditional geographic approach connects distance with costs, mainly travelling costs. In the last decades travel costs have decreased tremendously and seem to have lost their relevance (Cairncross 1997). However, besides direct monetary travel costs, travelling implies indirect costs due to the time that is spend in travelling and cannot be used for other activities. Travel times depend on the available transportation infrastructure, which has improved in the last decades. This implies a decrease in the relevance of geographic distance. Comparing US and EU, there are no clear differences in travel times. Travelling by car is cheaper in the US. However, the spatial distribution of metropolitan areas differs with the US having the main centres at the borders of the country implying long travel distances, while in Europe there are economic centres more evenly distributed in space.

Second, many studies claim that geographic distance does not matter because of the distance itself but because of the occasions and options it provides for meeting each other, even unintentionally (Bathelt et al. 2004). One argument is based on the common belief that it is only possible to exchange certain kinds of knowledge face-to-face. Another argument is based on the assumption that interacting frequently, and especially face-to-face, enhances trust and, therefore, the willingness to exchange knowledge (see Storper and Venables 2004 for an overview of the arguments). Finally, it is argued that the search for collaboration partners has a regional bias, leading to more collaborations with nearby partners (Broekel and Binder 2007). These arguments seem not to lose relevance with time and there is also no obvious difference between US and EU.

Third, it is argued in the literature that social proximity is the reason for the findings on the impact of geographical proximity. Breschi and Lissoni (2003 and 2009) have shown in their works that collaboration occurs mainly between actors that know each other and that more geographically nearby actors are well-known so that collaboration has a strong regional bias. The driving force according to this argument is social proximity, which in most cases comes with geographic proximity, except if people move locations. As a consequence, mobility becomes an issue. There are clear differences in the mobility of people between US and EU. People in the US are more mobile, especially with respect to large distances (see e.g. Ihrke and Faber 2012: Table 1, Statistisches Bundesamt Deutschland 2012: p.23). For some years now, EU policies have been directed at the integration of innovation activities, e.g. policies aiming at a European Research Area (ERA), students' mobility programs (like Erasmus), and the European Regional Development Fund (ERDF) aiming at the cohesion of European regions. If these approaches were successful, we should see a decreasing relevance of geographic distance in the EU, especially in co-publications, while there is no obvious reason to expect changes in the US.

Fourth, in the context of globalization the last decades have been characterized by an increasing similarity of regulations and laws between countries, especially among developed countries and an increasing share of the English-speaking population in countries in which English is not the official language (European Commission 2012). These development could be responsible for decreasing cultural and institutional barriers. However, this holds only for the EU and not for the US, where these barriers are not present.

Besides geographical distance, we also study the impact of cognitive distance (in form of technological and scientific distance). The importance of an optimal distance in interaction and collaboration has been intensively discussed in the literature (e.g. Nooteboom 2000 and Nooteboom et al. 2007). Neither very small nor large cognitive distances lead to good outcomes and learning effects in interactions. Hence, most collaborations can be expected to be characterised by a rather medium cognitive distance between the involved actors. However, recent literature argues that new technological developments are more and more based on the combination of rather distant technological fields (Choi and Valikangas 2001, Porter and Rafols 2009). As a consequence, it can be expected that the optimal cognitive distance between actors becomes larger with time, especially in the technological sphere.

The scientific world is slightly different. Scientific researchers focus mainly on publishing, which implies that whether research can be published is a major issue. Although interdisciplinary research has been propagated strongly in the last decades, most journals are still very focused and even the university system is in most countries structured in scientific fields with little interaction. Hence, it is interesting to study whether the support for interdisciplinary research, e.g. done by the EU, has effects that can be observed in our study.

Besides the above arguments, the amount of knowledge increases continuously and exponentially and each individual is able to know a decreasing share of the overall knowledge even when educated intensively and interdisciplinary. This is the reason for the increasing share of team inventions and larger teams (Wuchty et al. 2007). It might be argued that if the share of total knowledge that is known by each actor decreases, it becomes necessary to collaborate with cognitively less distant partners. This would lead to a decrease in cognitive distances, which should similarly hold for publications and patents.

**Empirical knowledge**

There are several studies on innovation collaboration between regions, firms, or individuals which investigate the influence of more than one type of distance empirically. In these studies, all types of distance have a negative effect on collaboration, sometimes there is no significant effect at all. In case it is explicitly stated, technological distance seems to have the strongest negative effect. The bulk of the studies captures Europe and is of static nature. Table 1 summarizes the main findings.

**Table 1. Literature survey (ordered alphabetically, non-exhaustive)**

| Author & Year | dependent variable | types of distance | static/dynamic analysis | main results |
|---|---|---|---|---|
| Aldieri and Cincera 2009 | inter-firm knowledge spill-over (808 large firms worldwide) | spatial and technological distance | static (USPTO and Worldscope data from 1988-1997) | technological proximity leads to larger spill-over than spatial proximity |
| Autant-Bernard 2001 | Co-patents between French departments | spatial and technological distance | static (data from 1994-1996) | both types of distance decrease collaboration frequency |
| Broekel and Boschma 2011 | dichotomous variable indicating whether organization i or j mentions the other as a relevant source of technological knowledge (Dutch aviation industry) | spatial, cognitive, social, and organizational distance | static (data from interviews made 2008/2009) | all four types of distance have a negative impact on technology exchange |
| Cassi and Plunket 2012 | Co-patents / patent citations between individual inventors in Europe | spatial, organizational, and technological distance; network position | static (data from 1990-2006) | all three types of distance hamper collaboration and citation, but geographical distance has the least effect |
| Capelli and Montobbio (2013) | Co-patents (and patent citations) in Europe (NUTS2) | Spatial, technological, and cultural distancs | static and partly dynamic (data from 1981-2000) | over 4 sub-periods decreasing importance of technological and spatial distance and same language for collaboration |
| Chessa et al. 2013 | co-patenting / co-publication /patent citation in Europe (NUTS3) and USA | only spatial distance | static and dynamic (patent data 1986-2010 OECD REGPAT / publication data 1991-2009 ISI WoS) | Research communities in the US have more cross-border links than those in Europe, but the amount is continuously increasing in |

| | | | | |
|---|---|---|---|---|
| | | | | Europe |
| Gomes-Caceres, Hagedoorn, Jaffe 2006 | inter-firm patent citations of US patents | spatial and technological distance | static (data from 1975-1999 in the NBER patent database linked with the CATI database) | both types of distance decrease citations, but the effect of technological distance is stronger |
| Hoekman, Frenken, van Oort 2009 | co-publication / co-patenting at NUTS3-level in Europe | spatial and institutional (same country) distance | static (data from 1988-2001/2004) | spatial distance hinders collaboration, national borders as well |
| Hoekman, Frenken, Tijssen 2010 | co-publication, NUTS2 level | spatial and institutional (same country and language) distance | dynamic (2000-2007) | Negative influence of national borders is decreasing, but not the sensitivity to spatial distance |
| Johnson, Lybecker 2011 | biotechnology patent citations in the US | spatial and technological (note: within biotech) distance | dynamic (data from 1975-1994) | the negative influence of spatial distance has decreased over time |
| Lata, Scherngell, and Brenner 2012 / Scherngell and Lata 2012 | co-patents / EU-FP cooperation between NUTS2-regions in Europe | spatial and technological distance, country/language border effects | dynamic (data from 1999-2006) | all types of distance have significant negative effect, but for some variables it is decreasing over time |
| LeSage et al. 2007 | patent citations at NUTS2-regions in Europe | spatial and technological distance, country/language border effects | static (data from 1985-2002) | technological distance decreases citation probability ten times as much as spatial distance |
| Maggioni and Uberti 2009 | Four types of inter-regional links between five European countries (NUTS2-level), excluding intra-national links | spatial, technological, and sectoral distance | static (data from 1998-2003) | all types of distance have significant negative effect |
| Maurseth and Verspagen 2002 | patent citations between European regions (NUTS1 and NUTS2) | spatial, cultural, and technological distance | static (data from 1979-1996) | all types of distance have significant negative effect |
| Moreschalchi et al. 2014 | co-patenting / co-publication /patent citation/inventor-applicant relations; NUTS3 for Europe, OECD TL3-regions outside Europe | spatial, cultural, and technological distance | dynamic (EPO data from 1988-2009) | all types of distance have significant negative effect, trend is first decreasing, then increasing, now stagnating |
| Picci 2010 | co-patenting between countries worldwide | spatial, cultural, and technological distance | static (data from 1990-2005) | all types of distance have significant negative effect |

| | | | | |
|---|---|---|---|---|
| Rosenkopf and Almeida 2003 | patent citations of semiconductor firms in the US | spatial and technological distance | static (patent data from 1990-1995) | both types of distance hinder citation |
| Scherngell and Barber 2011 | joint research projects in the EU FP5 program (NUTS2) | spatial and technological distance, country/language border effects | static (fifth EU framework program data) | technological distance has a stronger negative effect than spatial distance |
| Scherngell and Hu 2011 | co-publications between 31 Chinese regions | spatial and technological distance | static (publications from 2007) | technological distance has a stronger negative effect than spatial distance |

The table shows that there are still some empirical gaps regarding the analysis of inter-regional collaboration in the US and the analysis from a longitudinal perspective. We found few dynamic studies. The one by Chessa et al. (2013) uses just spatial distance and finds a decreasing influence of spatial distance in Europe (using several measures of collaboration), although not continuously decreasing. In particular, there has not been a stronger decrease in the EU than in other OECD countries since 2003. Lata et al. (2012) also find a decreasing negative impact of spatial distance (for EU framework program projects). Hoekman et al. (2010) find a decreasing importance of national borders, while the negative effect of spatial distance is not decreasing. Capelli and Montobbio (2013) compare four sub-periods of five years each and find a decreasing importance of spatial and technological distance for patent collaboration but increasing importance of spatial distance for patent citations. By comparing old and new EU member states, they find an increased amount of collaboration between old and new EU members. This hints at an indeed positive development of integration in European research activities. The study closest to our investigation is the one by Moreschalchi et al. (2014). It is a follow-on study of Chessa et al. (2013) including, spatial, technological, and cultural distance. They find a decreasing influence of spatial distance until the mid-90s, then an increase, and lately a stagnating trend. When comparing the EU with non-EU collaboration, they find an EU integration effect until 2004, but not further on. In their static examination, Chessa et al. (2013) find that spatial distance plays a smaller role in the US compared to the EU. Similarly, Fritsch and Slavtchev (2007) find a smaller radius of university-business collaboration in Germany compared to the study of Acs et al. (2002) for the US. However, we do not know any other study comparing US and Europe regarding the impact of spatial distance on innovation collaboration.

**Hypotheses**

Overall, our theoretical considerations and earlier empirical record hint at a lower sensitivity for spatial distance in the US compared to Europe. The theoretical arguments are mainly based on the higher mobility of people in the US and on the additional barriers that exist in Europe, such as language and institutional differences. Empirically, two studies support this difference between US and EU (Chessa et al. 2013 and Fritsch and Slavtchev 2007). Against this background, we will test the following hypothesis:

*H1: The negative influence of spatial distance on innovation collaboration is stronger in Europe than in the US.*

However, part of the theoretical arguments are based on the fact that the EU is still not an as cohesive space as the US. Country borders and language difference exist that hinder collaboration between actors. Hence, we may state the following hypothesis:

*H2: A main share of the difference between US and EU in the relevance of spatial distance is driven by country borders and language differences.*

From our considerations above follows that spatial as well as other forms of distance continue to play a role - everywhere. However, there are some indications of a trend that the impeding effect of spatial distance is decreasing over time, at least in Europe, but probably also in the US. The general argument is based on the decrease in transport and travel costs and improvements in information and communication technologies. However, the empirical literature does not provide clear evidence about trends in the relevance of geographic distance although a decrease is slightly more often supported. Similarly, the empirical literature does not provide a clear statement about differences between the EU and other countries. The theoretical arguments are quite clear in this context. Given the efforts of the EU government to support mobility and the cohesion within the EU, the relevance of geographic distance, country borders and language barriers should decrease. In order to investigate the effects of EU policy, we will test the following hypothesis:

*H3: The negative influence of spatial and cultural distance on innovation collaboration has decreased in Europe during the last decade stronger than in the US.*

The empirical studies that have been conducted so far do not find a change in the negative impact of technological distance over time. The theoretical arguments (see above) lead to ambiguous expectations. With respect to patents, we have two contradicting arguments: New technological developments require more and more the combination of distant technologies, while increasing specialization makes collaborations with lower cognitive distance necessary. With respect to publications, the question is whether the increased support for interdisciplinary research counter-balances the increased specialization. We assume that the increased interdisciplinarity outweighs the increased specialization and state:

*H4: The impact of cognitive distance, measured as technological and scientific distance, decreases with time. Differences between US and EU are not expected.*

# Method

**Empirical Model**

The purpose of this study is to estimate and to compare how various distance and separation aspects influence R&D collaboration in Europe and US over time. As we are dealing with origin-destination flow data in the form of collaboration activities between different regions, it seems natural to employ a spatial interaction modeling perspective. Spatial interaction models have been widely used to explain different kinds of flows across geographical space (see, for example, Sen and Smith 1995, Fischer 2002). Details of the model are explained in Appendix 1.

**Variables**

Since the seminal paper on types of proximity by Boschma (2005) - and in some cases even before - many studies have included more than one type of proximity in models explaining collaboration or knowledge flows (e.g. co-patenting, patent citations, collaboration in research projects). Depending on the object of research and data availability types of distances were selected and operationalized. For the paper at hand, we developed six measures of distance.

Spatial distance was measured in 100km between the centers of each region. Three dummy variables are used for cultural distance (in a broad sense): (1) non-neighboring region equals one if two regions are not neighbored, (2) country border indicates whether there is a country border between two regions, (3) language border equals one if in two regions different languages are spoken. The first of these dummies (non-neighboring region) can be interpreted as a measure for cultural distance or as a measure for spatial distance. If the population is spatially distributed within a region, the distance between some parts of the population in neighboring regions might be very small and not adequately reflected by the distance between the regions' centers. This might facilitate collaboration between the neighboring regions. While in the US the used areas are in most cases defined such that they are built around the dominating city, the areas are defined in the EU more on a historical background. Thus, the spatial aspect within the non-neighboring dummy should be stronger in the EU. Besides this, neighboring regions often show cultural similarity, in Europe even if they belong to different countries. This seems to hold for US and EU similarly.

Technological distance is proxied by $1-r^2$ with $r^2$ being the uncentered correlation between the patent class distribution (aggregated to 121 classes) of patents applied for in the regions (see Hoekman et al. 2010, Moreno et al. 2005). In a similar way, "knowledge distance" is defined on the basis of publication data. To this end, the science classification of the Web of Science is used and aggregated to 68 classes (reflecting the classification of university activities used by the German Statistical Office). Again $1-r^2$ with $r^2$ being the uncentered correlation between the publication class distributions of regions is used as measurement for the knowledge distance. Through this, we obtain two measures that express in a similar way the similarity between the activities in two regions: The technological distance based on the patent activity and the knowledge distance based on the publication activity.

The population distributions differ between EU and US and so do inventor/researcher distributions. By including the number of inhabitants for each region, we control for this with an origin and destination variable as it is common in gravity models. Especially in our case the population distribution of the origin variable indicate the capability of the origin region to generate collaboration flows. In contrast, the population distribution of the destination variable characterize the affinity to attract collaboration flows to the destination region (Fischer and Wang 2011).

# Datasets and descriptive analysis

**Datasets**

In order to analyse different R&D collaboration networks across Europe and the US two datasets are used. We use the Web of Science database (WoS) to construct the co-publication networks and the

OECD REGPAT database (January 2013 edition) to build the co-patent networks. The Web of Science database is a bibliographical database maintained by Thomson Reuters containing information on publications in more than 12,000 journals and 148,000 conference proceedings. Furthermore, it contains information on the institutional addresses of authors. We use this information to construct our regional setting of the co-publication networks. For our co-publication networks, a network link is given when a publication contains two or more authors. The REGPAT database is a regionalised patent application database, provided by the OECD covering information on the applicants and inventors addresses for more than 30 countries. In this study, we use the inventor addresses in order to trace the origin of the invention. For the co-patent networks, a network link is given when a patent comprises two or more inventors.

If a patent or publication contains inventors or authors from more than two regions, we assign a link to each region pair that is involved. We consider only links within one of the two studied economic areas, i.e. all US-EU collaborations as well as all collaborations with actors outside US and EU are excluded.

**Regions**

Several of the studies presented in Table 1 are based on the rather large NUTS-2 level. The paper at hand takes a more detailed approach and investigates inter-regional collaboration on the NUTS-3 level (1303 regions in the European Union) in Europe. Using fine-grained spatial units has at least two advantages: (1) most large cities do not have a border with the next large city so that being neighboring is not mixed up with collaborations between large cities, (2) using the centers of the regions for calculating distances between regions results in more accurate distances. To the authors' knowledge, there is no study analyzing inter-regional collaboration in the US and there is no equivalent to the NUTS classification in the US except the OECD territorial level 3, which is comparable in size with the NUTS2 level. Most similar in size to NUTS3 are the Core Based Statistical Areas (CBSAs). They contain so-called "metropolitan and micropolitan areas", i.e. urban agglomerations of more than 10,000 inhabitants defined by the US Office of Management and Budget (OMB). The 929 CBSAs consist of 388 metropolitan and 541 micropolitan areas. By definition, this classification excludes rural areas with less than 10,000 inhabitants (referring to appr. 13% of the population). In fact, we expect that regarding patent inventors, the CBSAs cover a share which is even larger than that of the population, because innovative firms are rarely located in the countryside. The average region in the USA in 2010 had 316,000 inhabitants (sd: 1062,000), the average NUTS3 region had 382,026 inhabitants (sd: 463,503). Thus, the regions are of comparable size.

For each region we have the geographical location (latitude/longitude) and the population from government data. The US data is from the 2010 Census Gazetteer by the US Census Bureau. The EU data is available from Eurostat and refers to 2012. The assignment of inventors to the regions (NUTS-3 and CBSAs) was done according to postal codes and in ambiguous cases supported by city/county name and state.

# Results and Discussion

Comparing EU and US, we have to keep in mind that there are structural differences between these two areas under consideration: Europe contains country borders and language area borders. To deal with these borders, we defined two dummies. However, in order to better understand the meaning of these two kinds of borders, we analyze two models for Europe: (1) the same model is estimated for Europe and the US ignoring country and language borders, (2) country and language borders are added in the model for Europe. In the following we will first discuss the results for publications, then for patents, and finally compare these results.

Instead of presenting the estimated coefficients for the models, we will present and discuss the estimated incidence rate ratios (see Abramovsky et al. 2007). In our case, incidence rate ratios provide information about how many times more likely a link becomes if the independent variable is increased by one. Values above one mean that the likelihood increases, while values below one stand for a decrease in the likelihood.

Furthermore, we use two different approaches. Firstly, in order to compare EU and US (Hypothesis H1), we pool the data over the years for each area, including yearly dummies. As a consequence, we obtain for each, the EU and the US, one incidence rate ratio estimate for each independent variable. Secondly, in order to study time trends (Hypotheses H2 and H3), we analyze each year separately, so that we are able to examine the change of the estimates over the years.

**Publications**

*Comparison USA-EU.* Let us first discuss the random effects model with yearly dummies. All distance measures have significant coefficients (Table 2). If we use the same model for the US and EU, we observe a clear and significant differences in the incidence rate ratios. As expected (Hypothesis H1) geographical distance reduces the likelihood of collaboration more in the EU than in the US. However, we have to take into account also the finding for the variable reflecting neighboring regions. While in the EU this variable is insignificant, collaborations are less likely in the US for actors that are not located in neighboring regions. Hence, if we consider the distance and neighboring effects, we obtain a distance dependence that is given for some examples in Table 3. The results show that for medium distances (100-150 km) the likelihood of links is similar for the US and the EU, but collaboration between neighboring regions and regions far away from each other is much less likely in the EU.

**Table 2. Estimated incident rate ratios (standard errors in brackets) for the co-publication network using the random effects Negative Binomial specification (US and EU 1999-2009)**

| | US | EU | EU (full model) |
|---|---|---|---|
| *Origin and Destination variable* $[\alpha_1] = [\alpha_2]$ | $1.805^{***}$ (0.004) | $1.992^{***}$ (0.005) | $2.140^{***}$ (0.006) |
| *Geographical distance* | $0.696^{***}$ (0.002) | $0.449^{***}$ (0.001) | $0.728^{***}$ (0.002) |
| *Non-Neighbouring region* | $0.639^{***}$ (0.018) | 1.006 (0.009) | $0.953^{***}$ (0.009) |
| *Country border* | | | $0.487^{***}$ (0.005) |
| *Language region* | | | $0.439^{***}$ (0.004) |
| *Technological Distance* | $0.331^{***}$ (0.004) | $0.485^{***}$ (0.005) | $0.426^{***}$ (0.004) |
| *Knowledge Distance* | $0.00307^{***}$ (0.000) | $0.000461^{***}$ (0.000) | $0.000460^{***}$ (0.000) |
| *Number of significant time effects* | 10 | 10 | 10 |

| | | | |
|---|---|---|---|
| *Number of origin spatial filters* | 176 | 70 | 70 |
| *Number of destination spatial filters* | 175 | 78 | 78 |
| *Constant* [$\alpha_0$] | $0.000^{***}$(0.000) | $0.000^{***}$ (0.000) | $0.000^{***}$ (0.000) |
| *Dispersion* ($\gamma$) | $1.556^{***}$ (0.006) | $1.701^{***}$ (0.003) | $1.559^{***}$ (0.003) |
| Log Likelihood | -2112015.3 | -3992530.90 | -3966073.20 |

**Table 3. Likelihood of a co-publication link for a certain region combination in comparison to intra-regional links according to the estimated models**

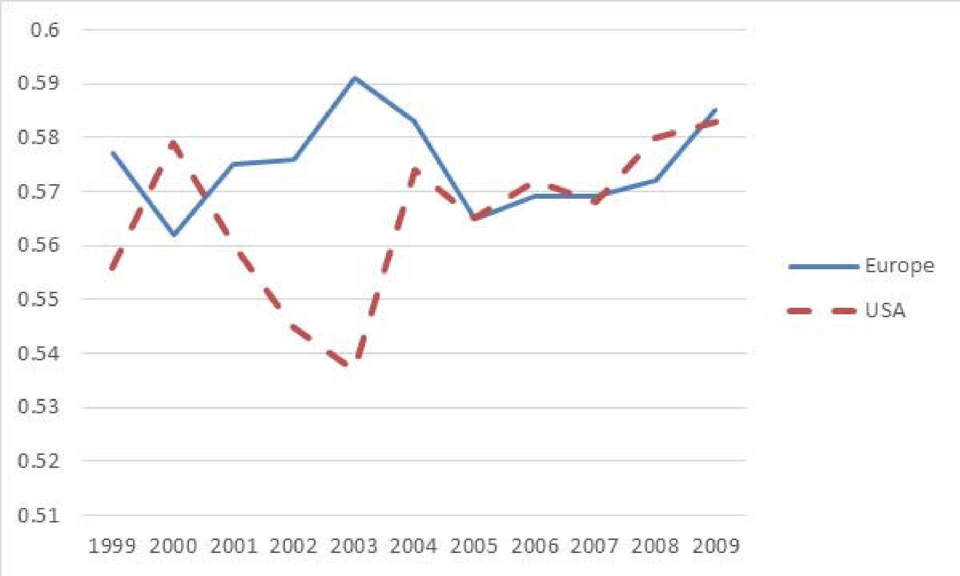
| | US | EU | EU (full model) |
|---|---|---|---|
| *Neighboring regions with a distance of 50km* | 83% | 67% | 85% |
| *Non-Neighbouring regions with a distance of 100km* | 44% | 45% | 69% |
| *Non-Neighbouring regions with a distance of 200km* | 31% | 20% | 51% |
| *Non-Neighbouring regions with a distance of 500km* | 10% | 2% | 19% |

The picture changes if we consider language and country borders in the EU. If the effects of these borders are explicitly included in the modelling (full model), the effect of distance decreases strongly. Table 3 presents also the likelihood for some exemplary inter-regional links in this case. This shows that distance matters less in Europe. All the seemingly larger effects of distance in Europe can be explained by country and language borders. Hence, Hypothesis H1 is only partly confirmed by our results. Indeed, co-publications over larger distances are less likely in the EU than in the US. However, this can be explained by country and language borders, which only exist in the EU. If these borders would not exist, our model would predict more distant co-publications in the EU than in the US. The significant findings for country and language borders in the EU clearly confirm Hypothesis H2.

Considering cognitive distances, we find that knowledge distance plays by far a larger role than technological distance for co-publication activities. This holds for the US as well as the EU, and is no surprise, since knowledge distance is calculated on publication data. Comparing US and EU, knowledge distance has the larger impeding effect in the EU. Hence, differences in the scientific specialization between regions prevents collaboration between these regions more in the EU than in the US. This contradicts Hypothesis H4. However, above we argued that the relevance of cognitive distance in scientific collaborations might be influenced by a strong structuring of university research into scientific subjects. The results might imply that these structural bindings are stronger in Europe than in the US.

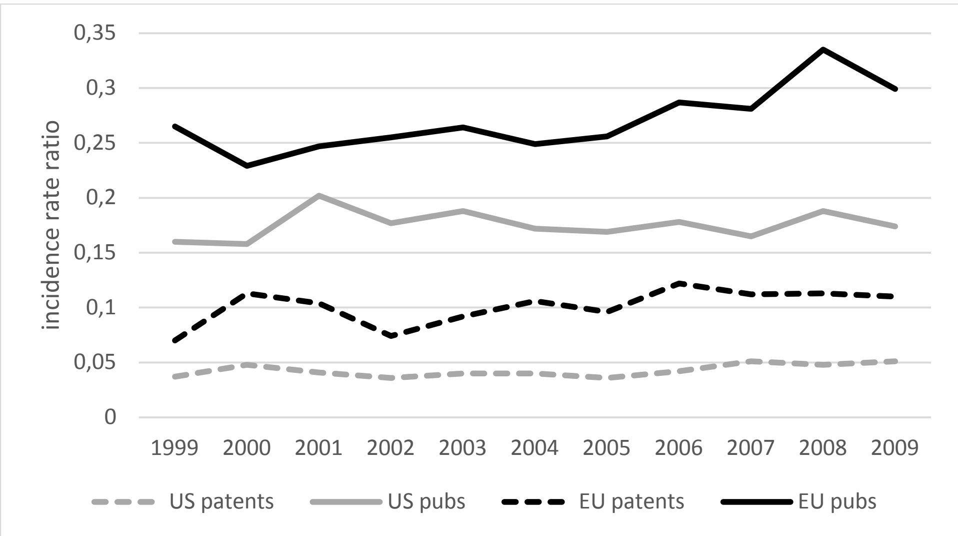
*Time trend.* In the cross section estimation, none of the distance measures exhibits a significant time trend in the USA (see Table A.1 in the appendix). In contrast, two significant trends can be found in the EU (see Table A.3 in the appendix): Technological and knowledge distance become less important. Thus, for publications Hypothesis H3 is rejected while Hypothesis H4 is partly confirmed. Let us first discuss Hypothesis H3. It was expected that the effect of distance as well as cultural difference decreases with time, especially in Europe. Besides the various aspects of globalization, the activities of the EU have been expected to have such an impact. However, as in some other studies, we do not find evidence for a decreasing effect of distance in the period from 1999 to 2009, neither in the US nor the EU. This finding is in line with Moreschalchi et al. (2014) and it shows, that the EU and all its integration endeavors are hardly or not reflected in publication collaboration. A visual inspection shows at least for the most recent few years a positive trend, likewise for the EU and the USA (Figure 1). However, space still plays a strong role in scientific collaboration.

**Figure 1: Incidence rate ratios for spatial distance (publications).**




In contrast, we observe in the EU a decreasing impeding effect of cognitive distance (see Figure 2; Figures for knowledge distance are Figures A.1-3 in the appendix). This is in line with Hypothesis H4, which was motivated by an increasing interdisciplinarity. In contrast to our expectations, the finding only holds for the EU. Hence, it could be that we observe at this point an effect of EU policy. This would mean that we are not able to prove any effect of the EU framework programs from 1999 to 2009 on the distance in co-publication collaboration, but that we are able to prove a positive effect on the interdisciplinary of such collaboration.

**Figure 2: Incidence rate ratios for technological distance**




**Patents**

*Comparison USA-EU.* As in the case of publications, we obtain significant impacts for all distance measures (see Table 5). Again, we find that spatial distance seems to be much more important in the EU when estimating the same model (excluding country and language border). At the same time, again, being neighbours matters more in the US than in the EU. Table 4 presents the combination of these two effects for a number of exemplary cases. It becomes clear that the picture is different from the one for publications. Again, the US shows higher likelihood ratios for neighboring regions and regions far away from each other, if the same model is used. In contrast to the case of publications, this does not change fundamentally if country and language borders are considered. Hence, Hypothesis H1 is partly confirmed for patents: Links between neighboring regions and between regions far away from each other are more likely in the US than in the EU. This is partly, but not completely, explained by country and language borders, which confirms Hypothesis H2. Interestingly, for medium distances between regions we find a stronger reduction of the likelihood of collaboration for the US than for the EU. This difference between medium distances and large distances between US and EU might be caused by the regional structure of the economic activity. While innovative activity shows some concentration on the East and West coast in the US, it is rather concentrated in the middle in the EU.

**Table 4. Likelihood of a co-patent link for a certain region combination in comparison to intra-regional links according to the estimated models**

| | US | EU | EU (full model) |
|---|---|---|---|
| *Neighboring regions with a distance of 50km* | 71% | 49% | 62% |
| *Non-Neighbouring regions with a distance of 100km* | 11% | 21% | 25% |
| *Non-Neighbouring regions with a distance of 200km* | 5.7% | 5.0% | 9.7% |
| *Non-Neighbouring regions with a distance of 500km* | .7% | .1% | .6% |

Considering cognitive distance, technological distance plays clearly a stronger role than knowledge distance. This is not surprising because technological distance is calculated on the basis of patent data. While we found stronger effects of cognitive distance on co-publications in the EU, for co-patents these effects are stronger in the US than in the EU. Hence, the second part of Hypothesis H4 is again not confirmed: Technological distance is not similarly important in the US and the EU, but seems to matter more in the US. We can only speculate on the reasons for that. One possible explanation is that the economic activity is more diverse within Europe offering more options for interdisciplinary collaboration. Another possible explanation is based on a potential impact of the EU policy, which fosters interdisciplinary collaboration also in the economy.

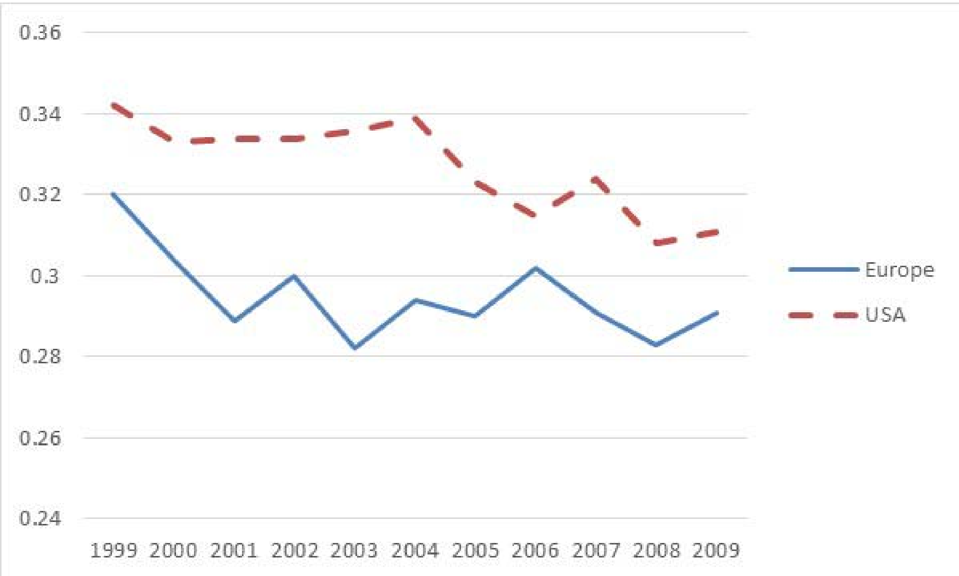
**Table 5. Estimated incidence rate ratios (standard errors in brackets) for the co-patent network using the random effects Negative Binomial specification (US and EU 1999-2009)**

| | US | EU | EU (full model) |
|---|---|---|---|
| *Origin and Destination variable* $[\alpha_1] = [\alpha_2]$ | $2.090^{***}$ (0.005) | $2.000^{***}$ (0.006) | $2.179^{***}$ (0.006) |
| *Geographical distance* | $0.499^{***}$ (0.001) | $0.242^{***}$ (0.000) | $0.384^{***}$ (0.001) |
| *Non-Neighbouring region* | $0.227^{***}$ (0.004) | $0.860^{***}$ (0.007) | $0.661^{***}$ (0.006) |
| *Country border* | | | $0.360^{***}$ (0.003) |
| *Language region* | | | $0.375^{***}$ (0.003) |
| *Technological Distance* | $0.095^{***}$ (0.001) | $0.276^{***}$ (0.003) | $0.214^{***}$ (0.002) |
| *Knowledge Distance* | $0.297^{***}$ (0.006) | $0.597^{***}$ (0.007) | $0.611^{***}$ (0.007) |
| *Number of significant time effects* | | | |

| | | | |
|---|---|---|---|
| | 9 | 9 | 9 |
| *Number of origin spatial filters* | 24 | 47 | 47 |
| *Number of destination spatial filters* | 24 | 45 | 45 |
| *Constant* [$\alpha_0$] | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| *Dispersion* ($\gamma$) | 1.134*** (0.007) | 1.458*** (0.005) | 1.656*** (0.007) |
| Log Likelihood | -631034.28 | -1647345.60 | -1604778.20 |

*Time trend.* The cross-sectional regressions for each of the eleven years under observation provide further insights: In the US, spatial distance becomes a higher barrier over time (significant at 0.1%) while in Europe the same trend is significant only on the 10% level. In fact, in the EU there is no trend visible if the first year (1999) is excluded. However, despite the negative trend in the US, spatial distance is still less a barrier in the US than in the EU (see Figure 2). Overall, it is clear that space does not lose importance, but rather increased importance of co-patent collaboration in the years 1999 to 2009. This finding corroborates earlier findings (see Morescalchi et al. 2014). This result can be seen as partly confirming Hypothesis H3, which predicts a decrease of the relevance of geographical distance only for the EU. Although we do not observe such a decrease, having a negative trend in the US and no trend in Europe could be interpreted as a hint for the success in European integration: the global trend of increasing importance of spatial distance is maybe mediated by the European integration.

**Figure 3: Incidence rate ratios for spatial distance (patents).**




Besides the trend in the relevance of geographical distance, we do not find any other significant trends for the US. In contrast, we observe a number of trends for the EU. First, the likelihood of links between non-neighboring regions decreases with time in the EU. This compensates to some extent the missing trend in the dependence on geographic distance. While large distance collaborations become less likely compared to short and medium distance collaborations in the US, the EU shows a trend of a decrease of all collaborations between non-neighboring regions, independent of the distance. Hence, in both cases regional collaboration becomes comparably more frequent.

Second, country borders lose importance in the studied period of time. This can be either seen as evidence for an impact of the European integration policies or as the impact of decreasing differences in institutions between European countries and the consequentially lower transaction costs.

Third, we also observe a decreasing relevance of the technological distance in the EU. This finding is similar to the finding for co-publications. Again, interdisciplinary collaboration seems to increase in the EU, but not in the US. Again, this partly confirms Hypothesis H4, which states that the impeding effect of cognitive distance should decrease. However, it confirms the hypothesis only for the EU and not for the US. Again, we might interpret this as a sign for effects of the EU framework program as explained above.

Table 6 sums up our findings. Interestingly, we find for most distances a stronger impeding effect in the US compared to the EU. Only knowledge distance in the case of co-publications and geographic distance in the case of co-patents have a stronger effect in the EU. In addition, more decreasing trends are found for the distance effects in the EU, which might be seen as an effect of the cohesion policy. Increasing trends are found for the effects of geographic distance in both, the US and the EU, signaling an even increasing relevance of geographic proximity, in general.

**Table 6. Comparison and trends in the impeding effects of distances (significances at 5% are presented)**

| **Distances** | **Comparison (higher effect in …)** | | **Trend (effect is +: increasing or -: decreasing)** | | | |
|---|---|---|---|---|---|---|
| | **Patent data** | **Publication data** | **patent data** | | **publication data** | |
| | | | EU | US | EU | US |
| **spatial distance** | EU | similar | 0 | + | 0 | 0 |
| **non-neighboring region** | US | US | + | 0 | 0 | 0 |
| **technological distance** | US | US | - | 0 | - | 0 |
| **knowledge distance** | US | EU | 0 | 0 | - | 0 |
| **country border** | | | - | | 0 | |
| **language border** | | | 0 | | 0 | |

# Conclusion

The paper at hand compares the impeding effects of four types of distance on innovation collaboration behavior in the USA and Europe. Three features distinguish the study from others in the field of research. Firstly, we compare the two most important economic areas. Secondly, we use fine-grained spatial units in order to have rather homogeneous units. Lastly, we analyze a period of eleven years for a dynamic view of the different types of distance. We find that space continues to be a barrier for collaboration which is rather becoming stronger than weaker. In particular, patent collaboration is more distance sensitive than publication collaboration. In this regard, Europe and the US are quite similar.

Regarding patent collaboration, spatial distance is a greater barrier in Europe than in the US, even when controlling for country and language borders. Especially, US inventors tend more to collaborate with partners in neighboring regions and in faraway regions compared to EU inventors. Regarding publications, there is no significant difference in the effect of spatial distance, with European researchers being more likely to cooperate with partners in non-neighboring regions in small and medium distance.

We find clear differences in the relevance of cognitive distance. Most kinds of cognitive distances play a stronger role in the US. Only in the case of co-publications knowledge distance has a higher impeding effect in the EU. Furthermore, there are some positive trends in Europe: technological and knowledge distance barriers become weaker for publication collaboration, technological distance barriers become weaker for patent collaboration and the impeding effect of country borders in Europe becomes weaker on patent collaboration. These trends might be the result of EU policies. There is more detailed research necessary for analyzing causal effects, which is beyond the scope of the paper at hand.

Another interesting topic would be to investigate technology-specific differences in the relevance and development of the effect of different types of distance. Running the models for a set of different industries or technologies will thus be the next step in a follow-on study.

Certainly, the investigation above has some limitations. A time period of eleven years may be too short to expose strong trends. We had to exclude data from teams composed of individuals which are not all located in one of the two economic areas under observation. Nevertheless, the study is the first one to compare Europe and USA with such large datasets and over several years. Further studies of this kind would be helpful to foster our results.

# References


**Abramovsky, L., Harrison, R., & Simpson, H. (2007). University Research and the Location of Business R&D. *The Economic Journal, 117*(519), C114-C141.**

**Acs, Z. J., Anselin, L., & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research Policy, 31*(7), 1069-1085.**

**Agrawal, A., Cockburn, I., & McHale, J. (2006). Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography, 6*(5), 571-591.**

**Aldieri, L., & Cincera, M. (2009). Geographic and technological R&D spillovers within the triad: micro evidence from US patents. *The Journal of Technology Transfer, 34*(2), 196-211.**

**Autant-Bernard, C. (2001). The geography of knowledge spillovers and technological proximity. *Economics of Innovation and New Technology, 10*(4), 237-254.**

**Autant-Bernard, C., Billand, P., Frachisse, D., & Massard, N. (2007). Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies*. *Papers in Regional Science, 86*(3), 495-519.**

**Bathelt, H., Malmberg, A., & Maskell, P. (2004). Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography, 28*(1), 31-56.**

**Boschma, R. (2005). Proximity and Innovation: A Critical Assessment. *Regional Studies: The Journal of the Regional Studies Association, 39*(1), 61-74.**

**Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography, 9*(4), 439-468.**

**Breschi, S., Lissoni, F., & Malerba, F. (2003). Knowledge-relatedness in firm technological diversification. *Research Policy, 32*(1), 69-87.**

**Broekel, T., & Binder, M. (2007). The Regional Dimension of Knowledge Transfers - A Behavioral Approach. *Industry and Innovation, 14*(2), 151-175.**

**Broekel, T., & Boschma, R. (2011). Knowledge networks in the Dutch aviation industry: the proximity paradox. *Journal of Economic Geography, online first*.**

**Cairncross, F. (1997). *The death of distance*. Boston, MA.: Harvard Business School Press.**

**Cappelli, R., & Montobbio, F. (2013). European Integration and Knowledge Flows across European Regions. *Department of Economics and Statistics "Cognetti de Martiis" Working Paper 22/13*.**

**Cassi, L., & Plunket, A. (2012). Co-patenting and inventive performance: in search of the proximity paradox. *Paper for the DRUID Conference 2012*.**

**Chessa, A., Morescalchi, A., Pammolli, F., Penner, O., Petersen, A. M., & Riccaboni, M. (2013). Is Europe Evolving Toward an Integrated Research Area? *Science, 339*(6120), 650-651.**

**Choi, D., & Valikangas, L. (2001). Patterns of strategy innovation. *European Management Journal, 19*(4), 424-429.**

**European Commission (2011). Progress Report on the Europe 2020 Strategy. Brussels.**

**European Commission (2012). Europeans and their languages. *Special Eurobarometer 386*.**

**Fischer, M. M., Scherngell, T., & Jansenberger, E. (2006). The Geography of Knowledge Spillovers Between High-Technology Firms in Europe: Evidence from a Spatial Interaction Modeling Perspective. *Geographical Analysis, 38*(3), 288-309.**

**Fritsch, M., & Slavtchev, V. (2007). Universities and Innovation in Space. *Industry & Innovation, 14*(2), 201-218.**

**Gomes-Casseres, B., Hagedoorn, J., & Jaffe, A. B. (2006). Do alliances promote knowledge flows? *Journal of Financial Economics, 80*(1), 5-33.**

**Hoekman, J., Frenken, K., & Tijssen, R. J. W. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy, 39*(5), 662-673.**

**Hoekman, J., Frenken, K., & van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *The Annals of Regional Science, 43*(3), 721-738.**

**Ihrke, D. K., & Faber, C. S. (2012). Geographical Mobility: 2005 to 2010. Population Characteristics. United States Census Bureau.**

**Johnson, D. N., & Lybecker, K. (2011). Does Distance Matter Less Now? The Changing Role of Geography in Biotechnology Innovation. *Review of Industrial Organization, 40*(1), 21-35.**

**Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1-18.**

**Lata, R., Scherngell, T., & Brenner, T. (2012). *Integration Processes in European R&D: A comparative spatial interaction approach using project based R&D networks and co-patent networks.* Paper presented at the 52th ERSA Conference,**

**LeSage, J. P., Fischer, M. M., & Scherngell, T. (2007). Knowledge spillovers across Europe: Evidence from a Poisson spatial interaction model with spatial effects. *Papers in Regional Science, 86*(3), 393-421.**

**Maggioni, M., & Uberti, T. (2009). Knowledge networks across Europe: which distance matters? *The Annals of Regional Science, 43*(3), 691-720.**

**Maurseth, P. B., & Verspagen, B. (2002). Knowledge Spillovers in Europe: A Patent Citations Analysis. *Scandinavian Journal of Economics, 104*(4), 531-545.**

**Morescalchi, A., Pammolli, F., Penner, O., Petersen, A. M., & Riccaboni, M. (2014). The evolution of networks of innovators within and across borders: Evidence from patent data. *IMT Lucca EIC Working Paper Series #01/2014*.**

**Nooteboom, B. (2000). *Learning and Innovation in Organizations and Economies*. New York: Oxford University Press.**

**Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., & van den Oord, A. (2007). Optimal cognitive distance and absorptive capacity. *Research Policy, 36*(7), 1016-1034.**

**Picci, L. (2010). The internationalization of inventive activity: A gravity model using patent data. *Research Policy, 39*(8), 1070-1081.**

**Porter, A., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics, 81*(3), 719-745.**

**Rosenkopf, L., & Almeida, P. (2003). Overcoming Local Search Through Alliances and Mobility. *MANAGEMENT SCIENCE, 49*(6), 751-766.**

**Scherngell, T., & Barber, M. (2011). Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth EU Framework Programme. *The Annals of Regional Science, 46*(2), 247-266.**

**Scherngell, T., & Hu, Y. (2011). Collaborative Knowledge Production in China: Regional Evidence from a Gravity Model Approach. *Regional Studies, 45*(6), 755-772.**

**Scherngell, T., & Lata, R. (2012). Towards an integrated European Research Area? Findings from Eigenvector spatially filtered spatial interaction models using European Framework Programme data. *Papers in Regional Science*.**

**Statistisches Bundesamt Deutschland (2012). Bevölkerung und Erwerbstätigkeit: Wanderungen. *Fachserie 1 Reihe 1.2*. Wiesbaden.**

**Storper, M., & Venables, A. J. (2004). Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography, 4*(4), 351-370.**

**Ter Wal, A. L. J. (2010). Cluster Emergence and Network Evolution: A Longitudinal Analysis of the Inventor Network in Sophia-Antipolis. *Regional Studies*, 1-18.**

**Torre, A. (2008). On the Role Played by Temporary Geographical Proximity in Knowledge Transmission. *Regional Studies, 42*(6), 869-889.**

**Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science, 316*(5827), 1036-1039.**

# Appendix 1 – Model

**Model**

In this study, we use the spatial interaction modelling perspective in order to analyse different R&D collaboration networks across Europe and the US. Following Fischer and Wang (2011) spatial interaction models rely on three types of factors that explain the mean interaction frequencies between origins *i* and destinations *j* in time period *t*: the (*i*) origin- specific factors which characterises the origin *i* of the interaction; (*ii*) the destination specific factors which describe the destination *j* of the interaction and (*iii*) the origin-destination measures which characterize the spatial separation between a origin *i* and destination *j*. The general form of the spatial interaction model is given by

$$Y_{ijt} \mid y_{ijt} = a + x_{ijt} + \varepsilon_{ijt} \qquad i, j = 1, ..., n; \quad t = 1, ..., T \qquad (1)$$

with

$$x_{ijt} = O_{it} \; D_{jt} \; S_{ijt} \qquad i, j = 1, ..., n; \quad t = 1, ..., T \qquad (2)$$

where $Y_{ijt}$ is the dependent variable that denotes, in this case, the observations on R&D collaborations $y_{ijt}$ between region *i* and *j* in time period *t* and $\varepsilon_{ijt}$ refers to the disturbance term with the property $E\left[\varepsilon_{ijt} \mid y_{ijt}\right] = 0$ . $x_{ijt}$ is the expected mean interaction frequency of flows from regions *i* to *j* in time period *t*. $O_{it}$ and $D_{ij}$ characterizes the origin and destination factors in time period *t* , and $S_{ijt}$ denotes the function of some measure of separation between region *i* and *j* in time period *t*. It is generally accepted, that the origin and destination factors are best stated by power functions (see, for example, Fischer and Wang 2011). Thus, we define $O_{it} = o_{it}^{\alpha_1}$ and $D_{jt} = d_{jt}^{\alpha_2}$ , where $a_1$ and $a_2$ are the statistical parameters to be estimated. The principal core of the spatial interaction model is the separation function, which is specified as

$$S_{ijt} = \exp\left[\sum_{k=1}^{K} \beta_k \; s_{ijt}^{(k)}\right] \qquad i, j = 1, ..., n; \quad t = 1, ..., T \qquad (3)$$

where $s_{ijt}^{(k)}$ are *K* (*k* = 1, ..., *K*) separation variables and $ß_k$ are parameters to be estimated.

As highlighted by other studies (see, for instance, Cameron and Trivedi 1998, Long and Freese 2001, Fischer et al. 2006, Scherngell and Barber 2009) the use of least square regression requires origin-destination flows that are independent and log-normally distributed about the mean value. In our case, this assumption is violated due to the discrete nature of our dependent variable and the presence of zero flows. Consequently, the use of least square regression would produce biased estimates. The usual approach to solve this deficiency is the Poisson model specification. We therefore assume that our $Y_{ijt}$ is distributed as an independent Poisson variable. Furthermore, we follow Lata et al. (2012) and implement a panel Poisson version of the spatial interaction model, which is given by

$$x_{ijt} = \exp\left[\alpha_1 \log\left(o_{it}\right) + \alpha_2 \log\left(d_{jt}\right) + \sum_{k=1}^{K} ß_k s_{ij}^{(k)} + \sum_{t=1}^{T} H_t v_t + \gamma_{ij}\right] \qquad (4)$$

where $\gamma_{ij}$ denotes the unobserved individual specific effect (see Baltagi 2008). We use a random effect instead of a fixed effect specification as we are dealing with time-invariant coefficients. Using a fixed effect specification our variables would be wiped out by the within estimator of the coefficients on the time varying covariates (see Baltagi 2008). The random term $\gamma_{ij}$ , that is time invariant but varies across all (*i*, *j*)-region pairs, accounts in our case for region-pair specific effects that are not included in the model. In addition, we add a subset of $H_t$ time dummies in order to capture aggregate year effects (Woodridge 2008).

One shortcoming of this model specification is that this approach does not take overdispersion into account i.e. that for the origin- destination flow data the variance usually overtakes the mean. As count data is usually overdispersed, we employ a negative binomial specification

However, in regard to our research question, namely to assess how specific separation effects evolve over time, we estimate the cross-sectional negative binomial models for each year separately and compare their parameters over time.

**Eigenfunction spatial filtering**

Different studies have used spatial interaction models for modeling origin and destination flows in various regional settings (see, for instance, Sen and Smith 1995, Fischer and Reismann 2002). In this context, numerous work have pointed to the problem of spatial dependence in interaction models, also called spatial autocorrelation of flows (Lesage and Pace 2008, Fischer and Griffith 2008, Scherngell and Lata 2012). Spatial autocorrelation of flows occur when collaborations flows are related to each other e.g. when collaboration flows from region a to region b are related with collaborations flows of region a to region c (see, for example, Chun 2008). As shown by various studies (see, for example, Lesage and Pace 2008, Fischer and Griffith 2008) spatial autocorrelation of flows can lead to incorrect inferences due to inconsistence of the standard errors, and, thus, to unrealistic significances.

Thus, we follow Lata et al. (2012) and Scherngell and Lata (2013) by applying the spatial filter method to our panel and cross-section model settings. This approach consists of introducing eigenvectors in order to filter out spatial autocorrelation in our models.

However, as suggested by Griffith (2003), it is not appropriate to use the full set of extracted eigenvectors. We follow Griffith (2003) and extract a subset of eigenvectors on the basis of Moran's I values, that show a higher value than 0.25. Furthermore, we adopt this set of eigenvectors to our spatial interaction model framework by using Kronecker products (see, for details, Fischer and Griffith 2008).

At this point, we use two different procedures to construct our final spatial filter sets. The first procedure adapts the eigenvectors to our Poisson panel model, while the second procedure adjusts the eigenvectors to the cross section models. We follow Lata et.al. (2012) in order to construct the eigenvectors to our panel models. These final panel origin and destination filters are time invariant and covers the total number of space-time observations. We denote the origin and destination filters as $E_q$ and $E_r$ respectively. For our cross sectional models, we follow the method of Scherngell and Lata (2013) and define a set of eigenvectors for each time period. The resulting origin and destination filters are time variant and are denoted as $F_q$ and $F_r$ respectively.

However, in the next step we include the time invariant- and time variant spatial filters as regressors into our panel and cross sections models, respectively. Thus, the final spatially filtered negative binomial panel interaction model is given by

$$x_{ijt} = \exp\left[\alpha_0 + \sum_{q=1}^{Q} E_{iq}\psi_q + \alpha_1 \log\left(o_{it}\right) + \sum_{r=1}^{R} E_{jr}\varphi_r + \alpha_2 \log\left(d_{jt}\right) + \sum_{k=1}^{K} ß_k s_{ij}^{(k)} + \sum_{t=1}^{T} H_t v_t + \gamma_{ij}\right] \quad (5)$$

The coefficients to be estimated for the spatial filters are $\psi_q$ and $\varphi_r$ .

Concerning our cross sectional models, we add the time variant origin and destination spatial filters $F_q$ and $F_r$ to equation 3. Hence, the final spatially filtered negative binomial spatial interaction model is given by

$$x^*_{ijt} = \exp\left[\alpha_{0t} + \sum_{q=1}^{Q} F_{iqt}\varpi_{qt} + \alpha_{1t} \log\left(o_{it}\right) + \sum_{r=1}^{R} F_{jrt}\tau_{rt} + \alpha_{2t} \log\left(d_{jt}\right) + \sum_{k=1}^{K} ß_{kt} s_{ijt}^{(k)} + \xi_{ijt}\right] \quad (6)$$

and

$$\exp\left(\xi_{ijt}\right) \sim \Gamma\left(\gamma\right) \quad (7)$$

where overdispersion is modelled by an additional modell parameter $\gamma$, and $\Gamma$ (·) describes the gamma function (see Long and Freese 2001). The coefficients to be estimated for the spatial filters are $\varpi_{qt}$ and $\tau_{rt}$ .

# Appendix 2 – Additional tables and figures

**Figure A.1. Incidence rate ratios for knowledge distance (patents).**

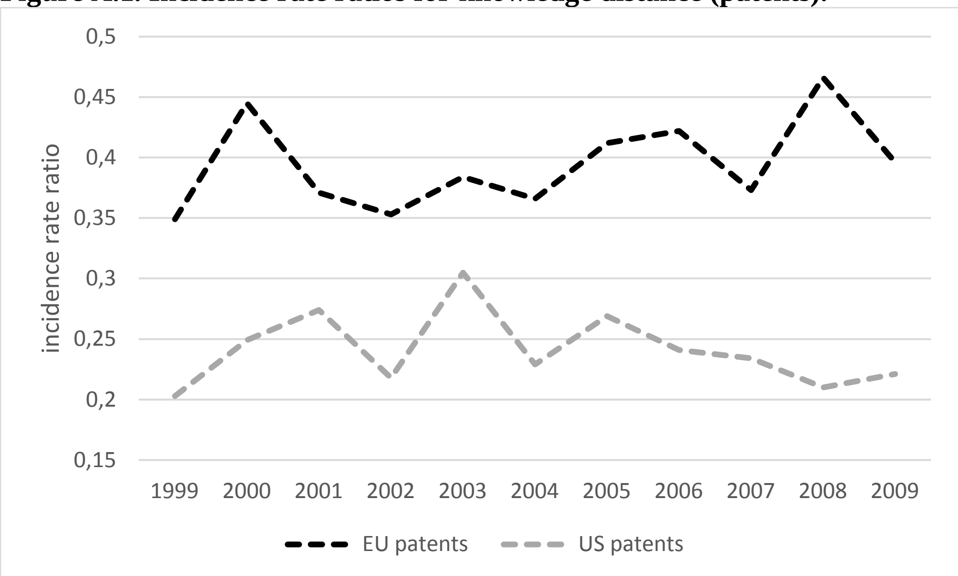


**Figure A.2. Incidence rate ratios for knowledge distance (European publications).**

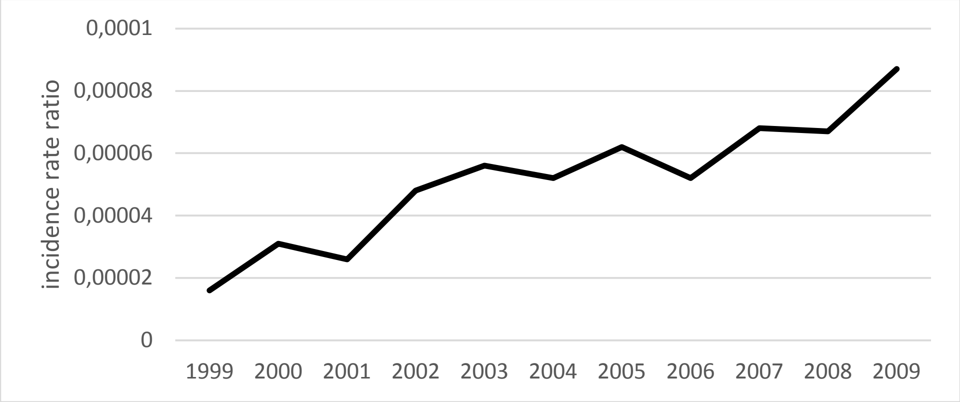



**Figure A.3. Incidence rate ratios for knowledge distance (US publications).**

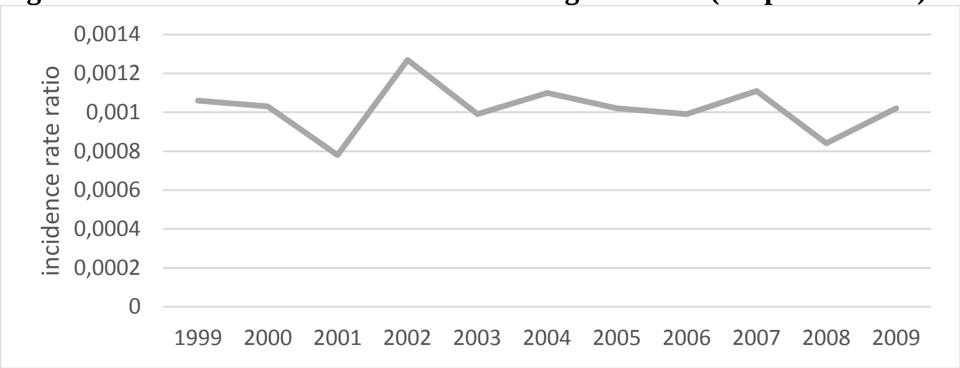


**Table A.1. Estimated incidence rate ratios (standard errors in brackets) for the co-publication network using the standard Negative Binomial specification (USA 1999-2009)**

| | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Origin and Destination variable* $[\alpha_1] = [\alpha_2]$ | 2.256*** (0.014) | 2.306*** (0.014) | 2.279*** (0.014) | 2.288*** (0.014) | 2.248*** (0.013) | 2.219*** (0.012) | 2.300*** (0.013) | 2.275*** (0.013) | 2.246*** (0.012) | 2.275*** (0.012) | 2.256*** (0.012) |
| *Geographical distance* [ $ß_1$] | 0.556*** (0.004) | 0.579*** (0.004) | 0.560*** (0.004) | 0.545*** (0.004) | 0.537*** (0.004) | 0.574*** (0.004) | 0.565*** (0.004) | 0.572*** (0.004) | 0.568*** (0.004) | 0.580*** (0.004) | 0.583*** (0.004) |
| *Non-Neighbour-ing region* [ $ß_2$] | 0.706*** (0.050) | 0.665*** (0.047) | 0.736*** (0.053) | 0.701*** (0.050) | 0.834*** (0.057) | 0.665*** (0.047) | 0.585*** (0.040) | 0.725*** (0.050) | 0.671*** (0.046) | 0.589*** (0.039) | 0.592*** (0.042) |
| *Technological Distance* [ $ß_5$] | 0.160*** (0.005) | 0.158*** (0.005) | 0.202*** (0.006) | 0.177*** (0.005) | 0.188*** (0.005) | 0.172*** (0.005) | 0.169*** (0.005) | 0.178*** (0.005) | 0.165*** (0.004) | 0.188*** (0.005) | 0.174*** (0.004) |
| *Knowledge Distance* [ $ß_5$] | 0.00106*** (0.000) | 0.00103*** (0.000) | 0.00078*** (0.000) | 0.00127*** (0.000) | 0.00099*** (0.000) | 0.00110*** (0.000) | 0.00102*** (0.000) | 0.00099*** (0.000) | 0.00111*** (0.000) | 0.00084*** (0.000) | 0.00102*** (0.000) |
| *Constant* $[\alpha_0]$ | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| *Dispersion* ($\gamma$) | 5.250*** (0.042) | 5.038*** (0.038 ) | 5.595*** (0.042 ) | 5.596*** (0.041 ) | 5.403*** (0.039 ) | 5.736*** (0.039 ) | 5.732*** (0.039 ) | 5.815*** (0.037 ) | 5.946*** (0.038 ) | 5.550*** (0.034 ) | 6.058*** (0.036 ) |
| Log Likelihood | -188,808 | -205,807 | -210,791 | -215,602 | -229,248 | -250,463 | -252,139 | -270,877 | -277,440 | -294,031 | -301,924 |
| Pseudo R2 | 0.307 | 0.307 | 0.295 | 0.293 | 0.291 | 0.281 | 0.284 | 0.278 | 0,273 | 0.277 | 0.268 |

**Table A.2. Estimated incidence rate ratios (standard errors in brackets) for the co-patent network using the standard Negative Binomial specification (USA 1999-2009)**

| | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Origin and Destination variable* $[\alpha_1] = [\alpha_2]$ | 2.520*** | 2.642*** | 2.642*** | 2.441*** | 2.547*** | 2.634*** | 2.584*** | 2.701*** | 2.668*** | 2.694*** | 2.546*** |
| | (0.029) | (0.030) | (0.030) | (0.026) | (0.026) | (0.027) | (0.026) | (0.027) | (0.026) | (0.030) | (0.028) |
| *Geographical distance* $[ß_1]$ | 0.342*** | 0.333*** | 0.334*** | 0.334*** | 0.336*** | 0.339*** | 0.323*** | 0.315*** | 0.324*** | 0.308*** | 0.311*** |
| | (0.005) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) | (0.004) |
| *Non-Neighbouring region* $[ß_2]$ | 0.099*** | 0.145*** | 0.129*** | 0.115*** | 0.146*** | 0.116*** | 0.138*** | 0.153*** | 0.133*** | 0.134*** | 0.154*** |
| | (0.008) | (0.011) | (0.010) | (0.010) | (0.010) | (0.008) | (0.009) | (0.011) | (0.009) | (0.010) | (0.011) |
| *Technological Distance* $[ß_5]$ | 0.037*** | 0.048*** | 0.041*** | 0.036*** | 0.040*** | 0.040*** | 0.036*** | 0.042*** | 0.051*** | 0.048*** | 0.051*** |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.003) |
| *Knowledge Distance* $[ß_5]$ | 0.203*** | 0.249*** | 0.274*** | 0.218*** | 0.305*** | 0.229*** | 0.269*** | 0.241*** | 0.234*** | 0.210*** | 0.221*** |
| | (0.016) | (0.018) | (0.020) | (0.015) | (0.070) | (0.016) | (0.018) | (0.016) | (0.015) | (0.015) | (0.072) |
| *Constant* $[\alpha_0]$ | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000*** | 0.000*** |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| *Dispersion* $(\gamma)$ | 11.138*** | 12.570*** | 12.199*** | 11.572*** | 11.198*** | 10.791*** | 10.745*** | 11.305*** | 10.688*** | 12.974*** | 11.560*** |
| | (0.186) | (0.194 ) | (0.186 ) | (0.174 ) | (0.167 ) | (0.157 ) | (0.154 ) | (0.156 ) | (0.150 ) | (0.193 ) | (0.176 ) |
| Log Likelihood | -53,756 | -60,596 | -61,230 | -63,186 | -65,944 | -67,817 | -69,855 | -73,938 | -72,524 | -63,272 | -62,531 |
| Pseudo R2 | 0.327 | 0.315 | 0.319 | 0.319 | 0.313 | 0.316 | 0.316 | 0.310 | 0,312 | 0.311 | 0.317 |

**Table A.3. Estimated incidence rate ratios (standard errors in brackets) for the co-publication network using the standard Negative Binomial specification (Europe 1999-2009) – Full model**

| | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Origin and Destination variable* $[\alpha_1] = [\alpha_2]$ | 3.920*** (0.039) | 3.870*** (0.036) | 3.713*** (0.033) | 3.419*** (0.030) | 3.485*** (0.029) | 3.650*** (0.031) | 3.586*** (0.029) | 3.810*** (0.031) | 3.672*** (0.029) | 3.703*** (0.007) | 3.870*** (0.030) |
| *Geographical distance* $[\beta_1]$ | 0.577*** (0.005) | 0.562*** (0.005) | 0.575*** (0.005) | 0.576*** (0.005) | 0.591*** (0.005) | 0.583*** (0.005) | 0.565*** (0.004) | 0.569*** (0.004) | 0.569*** (0.004) | 0.572*** (0.004) | 0.585*** (0.004) |
| *Non-Neighbouring region* $[\beta_2]$ | 0.484*** (0.027) | 0.465*** (0.025) | 0.506*** (0.026) | 0.522*** (0.026) | 0.615*** (0.030) | 0.588*** (0.029) | 0.520*** (0.025) | 0.460*** (0.023) | 0.532*** (0.025) | 0.508*** (0.024) | 0.427*** (0.021) |
| *Country border* $[\beta_3]$ | 0.289*** (0.008) | 0.298*** (0.008) | 0.330*** (0.008) | 0.343*** (0.008) | 0.309*** (0.007) | 0.296*** (0.007) | 0.311*** (0.007) | 0.312*** (0.007) | 0.321*** (0.007) | 0.325*** (0.007) | 0.303*** (0.006) |
| *Language region* $[\beta_4]$ | 0.425*** (0.011) | 0.469*** (0.012) | 0.426*** (0.010) | 0.412*** (0.009) | 0.382*** (0.008) | 0.423*** (0.009) | 0.415*** (0.009) | 0.407*** (0.009) | 0.429*** (0.009) | -0.410*** (0.008) | 0.401*** (0.008) |
| *Technological Distance* $[\beta_5]$ | 0.265*** (0.008) | 0.229*** (0.007) | 0.247*** (0.007) | 0.255*** (0.007) | 0.264*** (0.007) | 0.249*** (0.006) | 0.256*** (0.006) | 0.287*** (0.007) | 0.281*** (0.007) | 0.335*** (0.008) | 0.299*** (0.007) |
| *Knowledge Distance* $[\beta_6]$ | 0.000016*** (0.000) | 0.000031*** (0.000) | 0.000026*** (0.000) | 0.000048*** (0.000) | 0.000056*** (0.000) | 0.000052*** (0.000) | 0.000062*** (0.000) | 0.000052*** (0.000) | 0.000068*** (0.000) | 0.000067*** (0.000) | 0.000087*** (0.000) |
| *Constant* $[\alpha_0]$ | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| *Dispersion* $(\gamma)$ | 7.030*** (0.040) | 6.832*** (0.038 ) | 6.483*** (0.036 ) | 6.454*** (0.034 ) | 6.185*** (0.032 ) | 6.649*** (0.032 ) | 6.382*** (0.030 ) | 6.633*** (0.030 ) | 6.527*** (0.029 ) | 6.207*** (0.027 ) | 6.284*** (0.026 ) |
| Log Likelihood | -342,770 | -363,575 | -375,131 | -396,616 | -424,095 | -455,612 | -481,599 | -510,789 | -534,268 | -570,597 | -610,027 |
| Pseudo R2 | 0.286 | 0.283 | 0.282 | 0.276 | 0.276 | 0.268 | 0.267 | 0.262 | 0,259 | 0.259 | 0.255 |

**Table A.4. Estimated incidence rate ratios (standard errors in brackets) for the co-patent network using the standard Negative Binomial specification (Europe 1999-2009) – Full model**

| | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Origin and Destination variable* $[\alpha_1] = [\alpha_2]$ | 2.742*** (0.033) | 2.891*** (0.033) | 2.798*** (0.031) | 2.808*** (0.031) | 2.880*** (0.031) | 2.706*** (0.029) | 2.843*** (0.030) | 2.903*** (0.031) | 2.820*** (0.030) | 2.900*** (0.028) | 2.698*** (0.028) |
| *Geographical distance* $[\beta_1]$ | 0.320*** (0.003) | 0.304*** (0.003) | 0.289*** (0.003) | 0.300*** (0.003) | 0.282*** (0.003) | 0.294*** (0.003) | 0.290*** (0.003) | 0.302*** (0.003) | 0.291*** (0.002) | 0.283*** (0.002) | 0.291*** (0.002) |
| *Non-Neighbouring region* $[\beta_2]$ | 0.359*** (0.018) | 0.390*** (0.018) | 0.385*** (0.017) | 0.352*** (0.017) | 0.378*** (0.017) | 0.330*** (0.015) | 0.314*** (0.014) | 0.313*** (0.013) | 0.353*** (0.015) | 0.361*** (0.015) | 0.317*** (0.014) |
| *Country border* $[\beta_3]$ | 0.240*** (0.008) | 0.243*** (0.008) | 0.250*** (0.008) | 0.247*** (0.008) | 0.342*** (0.010) | 0.328*** (0.010) | 0.326*** (0.009) | 0.288*** (0.008) | 0.344*** (0.009) | 0.287*** (0.008) | 0.293*** (0.008) |
| *Language region* $[\beta_4]$ | 0.449*** (0.015) | 0.549*** (0.018) | 0.526*** (0.017) | 0.535*** (0.017) | 0.421*** (0.012) | 0.410*** (0.012) | 0.421*** (0.012) | 0.399*** (0.011) | 0.370*** (0.010) | 0.456*** (0.012) | 0.479*** (0.013) |
| *Technological Distance* $[\beta_5]$ | 0.070*** (0.003) | 0.113*** (0.004) | 0.104*** (0.004) | 0.074*** (0.003) | 0.092*** (0.003) | 0.106*** (0.004) | 0.096*** (0.003) | 0.122*** (0.004) | 0.112*** (0.004) | 0.113*** (0.004) | 0.110*** (0.004) |
| *Knowledge Distance* $[\beta_6]$ | 0.349*** (0.015) | 0.445*** (0.018) | 0.371*** (0.015) | 0.353*** (0.014) | 0.384*** (0.015) | 0.366*** (0.014) | 0.412*** (0.015) | 0.422*** (0.015) | 0.373*** (0.013) | 0.466*** (0.016) | 0.396*** (0.014) |
| *Constant* $[\alpha_0]$ | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) | 0.000*** (0.000) |
| *Dispersion* $(\gamma)$ | 8.512*** (0.096) | 8.124*** (0.086 ) | 7.549*** (0.080 ) | 8.127*** (0.083 ) | 7.177*** (0.073 ) | 8.330*** (0.081 ) | 7.783*** (0.075 ) | 7.798*** (0.073 ) | 7.533*** (0.069 ) | 7.467*** (0.069 ) | 8.555*** (0.078 ) |
| Log Likelihood | -126,746 | -143,178 | -146,603 | -151,539 | -153,787 | -164,846 | -169,785 | -179,308 | -184,279 | -182,795 | -187,927 |
| Pseudo R2 | 0.305 | 0.300 | 0.304 | 0.298 | 0.307 | 0.291 | 0.295 | 0.290 | 0,291 | 0.292 | 0.277 |