hal-00725073, version 2 - 24 Aug 2012

Does Geographical Proximity Still Matter? An Empirical Analysis of Two Long-Term Relational Databases

Olivier Bouba-Olga & Marie Ferru

CRIEF, University of Poitiers

Abstract: The purpose of this article is to provide possible answers, at an empirical level, to the question “Does geographical proximity still matter in collaborations for innovation?", since the lack of available data has made it impossible to provide real answers up to now. Relying on two real long-term relational databases relating to science-industry collaborations in France, and on two complementary indicators of geographical proximity, we will show that proximity continues to count (number of significant intradepartmental collaborations and low average distance between the partners). However, this dynamic analysis does allow us to state that the closest partnerships and the most distant ones increase the most over the study period. We shall also show that the role of geographical proximity and its evolution over the course of time differ according to the type of science-industry contract and the sectoral specialisation of the partners.

Keywords: geographical proximity, collaborations, research, science-industry

JEL Codes: O31, R12, O39

Acknowledgements: The authors wish to thank Michel Grossetti for making the CNRS database available to us, Nadine Massard for providing the CIFRE database, and Koen Frenken for his comments on the first draft of this article.

hal-00725073, version 2 - 24 Aug 2012

1. Introduction

As Morgan (2004, p.3) stressed, “Globalization and digitilization have been presented as ineluctable forces which signal the death of geography.”" The development of CIT and the means of transport will lead to a global economic world in which distance no longer matters (Castells, 1996; Cairncross, 1997). Innovation activities would seem to be an exception to this process in the view of many authors. Their strong polarisation seen since the late 1990s (Puga, 1999; Caniëls, 1997; Lallement et al., 2007, etc.) and the persistence of the effects of geographical proximity deriving from it are said to come from the tacit nature of the knowledge exchanged, from the existence of social networks among partners developed locally, from the intervention of policies, etc.

From an empirical standpoint, “the death of distance hypothesis has not been proven” (Frenken et al., 2010) in research collaborations and the studies done suffer from important limitations, given the poor availability of data. Indeed, not only is it the case that “most of the existing contributions on this matter lack data at the individual level” (Giuri and Mariani, 2007), but also and most especially “many studies on proximity and innovation have been analysed from a static perspective” (Boschma, 2005). Hence, although the question of geographical proximity in research collaborations has already been raised by many studies (see in particular Oinas, 2000 or Lublinski, 2003), none really provides an answer to the question “Does geographical proximity *still* matter?”

In this article, we try to provide possible answers to this question through an empirical study based on an exploration of two real long-term relational databases. These databases include substantial information relating to more than 30,000 science-industry collaborations carried out in France between 1986 and 2005, such as the location of the scientific and technical partners, their specialisation and the date of the project. These data enable us to analyse the importance and evolution of geographical proximity for two types of collaborations, with part of the partnerships studied involving CIFRE contracts, and the other part CNRS contracts (see below for more information about these two types of contracts). This comparative analysis appears all the more interesting in that we can expect results that differ according to the nature of the contract: the need for geographical proximity could be greater in CIFRE collaborations, since these contracts rely on the research work of a doctoral candidate alternating time at the firm and time in the laboratory.

We propose to conduct an empirical analysis of these two databases by adopting complementary strategies for defining geographical proximity: i) we measure it first using a relative-distance indicator, the travel time by train between the two capitals of the regions where the two partners are located, ii) we then formulate a typology of the spatial scales of collaboration making it possible to define different degrees of geographical proximity. More specifically, we distinguish collaborations within the same department (DEP), those between departments in the same region (REG), collaborations between adjoining regions (LIMITROPHE), those with the capital region (ICAP) and finally those between nonadjoining regions outside the capital region (IREG).

Initially in our empirical analysis we measure the evolution over the period of the time-distances and each category of our typology and test the existence of a trend, relying on nonparametric tests (Kendall’s tau), making distinctions according to the nature of the contract (CIFRE vs. CNRS). Then we test the influence of various potential determinants (the year the contract was made, the firm’s

hal-00725073, version 2 - 24 Aug 2012

business sector or the laboratory's research domain) on the effects of geographical proximity using complementary econometric models: the first model (multinomial logit) measures the effects of proximity through the spatial scales typology, whereas the second (sample selection model, Heckman two-step) measures the influence of the same determinants on the time-distance between the partners.

The rest of the article is organised as follows. We conduct a review of the literature in the second section, stressing the limitations inherent in the empirical work done to that point. In the third section, we present the data before specifying the methodology applied to exploring our two databases. In the fourth section, we give the first descriptive statistics and the key results from the econometric processing done. We then conclude and discuss these results in the final section.

2. Review of the literature

For many years there has been an ongoing debate over the importance of proximity in research collaborations: "The relevance of proximity is one of the most controversially discussed topics in the context of innovative linkages and networks" (Sternberg, 1999, p.553). Knoben and Oerlemans (2006), in a review of the literature on the subject underscore the especially significant value attached to geographical proximity ("geographical proximity (…) is the most frequently used dimension of proximity in the literature" (p.74)) and inventory 19 works linking geographical proximity and collaborations

This debate is based essentially on the evolutionist theory developed in the 1980s, according to which the tacit knowledge exchanged during the collaboration process is by definition not codifiable and therefore cannot be distributed easily in space. This led some authors (Verspagen, 1999; Patel and Pavitt, 1994 in particular) to consider that only face-to-face relationships make it possible to benefit from it. For example, Foray (1995) states that "the aptitude to assimilate and transfer scientific and technological knowledge that is not completely codified is largely dependent on opportunities for direct personal contact between the concerned parties"(p. 19). Even though it has thus been commonly accepted that geographical proximity promotes the development of collaborations (by reducing all of the transaction costs), "popular belief holds that geography no longer matters in scientific collaborations with the arrival of cheap air travel, English as a global language and the internet" (Frenken and al., 2010). Geographical proximity would then constitute less and less a necessary or sufficient condition for the realisation of research partnerships (Boschma, 2005; Gilly and Torre, 2000; Bunnell and Coe, 2001; Breschi and Lissoni, 2003; Gertler, 2003).

At an empirical level, despite the development of studies seeking to measure the importance of geographical proximity, "little is known about the role of geographical proximity in science collaborations" (Ponds et al., 2007) and "the death of distance hypothesis has not been proven" (Frenken et al., 2010). The results obtained vary greatly from study to study: whereas some attest to an increase over time in long-distance partnerships (Frenken et al., 2009; Narin et al., 1991; Luukkonen et al., 1993; Frenken, 2002; Magioni and Uberti, 2007; Wagner and Leydesdorff, 2005), others on the contrary show that geographic distance still plays a significant role still today in the creation of European research collaborations (Autant-Bernard et al. 2007; Maggioni and Uberti 2009; Scherngell and Barber 2009). Likewise, and above all, the existing empirical studies suffer from three major limitations.

hal-00725073, version 2 - 24 Aug 2012

1) The first limitation comes from the quality of the data, and more especially of the collaboration indicators applied. The econometric models pertaining to the "geography of innovation" (Feldman, 1994) – and particularly those that initially took into account the strong polarisation of innovation activities – rely on aggregate data such as patents, volume of expenditures, number of employees in R&D, etc. In these studies, interactions between actors are therefore not really observed but merely assumed, and the models thus account for the co-presence of actors and not true collaborations. Other, more recent studies measure collaborations through citations or participation in joint projects (Autant-Bernard et al., 2007; Scherngell and Barber, 2009; D'Este and Iammarino, 2010). Although these data make it possible to better understand interactions between actors than the previous data, the realisation of a true collaboration between partners is merely assumed: in some types of collaborations, as is the case with European partnerships in particular (EU Framework Program), the number of partners is such that it is not possible for all of the partners to have really collaborated during the project. It seems that "very few papers use data on direct collaborative agreements on R&D" (Massard and Mehier, 2009). Only studies based on data on co-patents (Maggioni et al., 2007) or co-publications (Katz 1994; Ponds et al. 2007; Frenken et al., 2010; Hoekman et al., 2010; Scherngell and Hu, 2011) or even more so those bearing on dyadic collaboration relationships (Laursen et al., 2011; Levy et al., 2009; Grossetti and Nguyen, 2001) can truly account for a collaboration relationship between the actors and in our view constitute true relational databases.

2) A second problem appears as to the indicator of geographical proximity applied. "The definition of this dimension of proximity differs slightly between different authors (…) with regard to the scale at which geographical proximity is defined" (Knoben and Oerlemans, 2006, p.74). Some authors evaluate proximity using a binary variable distinguishing national collaborations from international collaborations (Ponds et al., 2007); others consider that geographical proximity exists beyond regional borders and evaluate it at the NUTS 2 level (Scherngell and Barber, 2009; Hoekman et al. (2010)) or using an ordinal measurement (for example, Drejer and Vinding (2007) distinguish intra-regional, intra-national and international collaborations). Understood as such, geographical proximity seems too broad to be able to verify the hypothesis according to which proximity makes face-to-face interactions easier. The regional level (NUTS 3 level) as well as the use of a maximum kilometric distance indicator (under 100 miles for Laursen et al., 2011, for example) seem more relevant, even though they may prove too broad for certain territories. Yet to our knowledge there are few studies that allow a proximity analysis at a more refined level (department, town). In the same way, some studies have shown the importance of geographical proximity in testing the role of administrative boundaries (Hoekman et al., 2010), but very few have considered the specificities relating to the studied territories such as "capital effects" or "adjacent-regions effects" with the exception of Grossetti and Nguyen (2001) who underscore the hypertrophy of the Ile-de-France region in French collaborations. In short, and especially, considering the development of the means of transport, it seems to us important to test relative-distance indicators, by taking into account, for example, transport time or cost. Finally, it appears that "the geographical dimension is in complex ways related to other mechanisms that are still barely identified and that take place at different geographical levels" (cf. Autant-Bernard et al., 2007), showing the need to compare different indicators of geographical proximity.

3) In short, and especially, "many studies on proximity and innovation have been analysed from a static perspective" (Boschma, 2005). Very few studies are based on longitudinal data and therefore do not generally take into account the evolution of geographical proximity, with the exception of

Hoekman et *al*. (2010) and Grossetti and Nguyen (2001). For 33 European countries, the former underscore the persistence of local collaborations (taken at the NUTS 2 level) "and collaborations within territorial borders did decrease" between 2000 and 2007; the latter show the increase in local collaborations between 1986 and 1998 for France.

On the basis of these various studies and the limitations inherent in them, we seek to complement the existing literature through an exploration of two genuine long-term relational databases and a comparison of various indicators of geographical proximity.

3. Data and method

3.1 Data

In this article, we make use of two complementary databases: the CIFRE base and the CNRS base. The former brings together all of the science-industry contracts carried out as part of Conventions Industrielles de Formation par la Recherche [Industrial Agreements for Training through Research] – CIFRE— between 1981 and 2005. Since 1981, the CIFRE system has been subsidising any company under French law that hires a doctoral candidate and places him or her in a research collaboration with a public laboratory. These research contracts are in large part funded by the Ministry of Higher Education and Research. The second database brings together various types of information relating to research contracts carried out between 1986 and 2006 between French firms and laboratories under the Centre National de la Recherche Scientifique [National Scientific Research Centre] – CNRS – the main government research agency in France.

[TABLE 1]

For each base, we have information for the year the contract was signed (ANNEE), the regional location of the laboratory involved (REGLAB), the regional location of the company (REGETAB), the company's sector (SECTEUR) and the laboratory's scientific domain (DOMAINE). In the case of these last two variables, a major problem arises, since the nomenclatures used by the ANRT is not the same as that ordinarily used. We therefore had to aggregate the categories available in the CNRS base and construct a transfer matrix between the ANRT and CNRS nomenclatures. In the end, we were able to divide up the contracts into three sectors and six scientific domains. We have an aggregate base of 32,764 observations completely filled in, including 23,954 CNRS contracts and 8,810 CIFRE contracts, for the period 1985-2006. Table 1 shows the qualitative variables available in the two databases used.

3.2 Method

In order to analyse the evolution of the geographical proximity of the collaborators, we propose adopting two complementary strategies.

The first strategy consists of defining a typology (TYPO) of the spatial scales used, composed of five categories:

*TYPO=DEP when the contract concerns two partners located in the same department;

*TYPO=REG when the two partners are located in two different departments in the same region,

hal-00725073, version 2 - 24 Aug 2012

*TYPO=LIMITROPHE if the two partners are located in two adjoining regions,

*TYPO=ICAP when one of the partners is located in the capital region (Ile-de-France) and the other partner is located in another region,

*TYPO=IREG when the two partners are located in two different regions outside the Ile-de-France.

In our view, this typology provides a better understanding of geographical proximity than is generally achieved in empirical studies. Four of the five categories define decreasing degrees of geographical proximity measured in terms of physical distance: the relationships within the same department are the closest geographically (DEP), followed by those in the same region (REG), then those between adjoining regions (LIMITROPHE) and, finally, those between different regions outside the capital (IREG). The last category in the typology (ICAP) takes into account the specific nature of French geography, the capital region being at the heart of the country's network of transport infrastructure, whether the road, rail or air networks.

The second, more classic strategy consists of using a relative measurement of geographical proximity between actors, namely, travel time by train between the two capitals of the regions where the actors are located (TRAIN variable). This measurement seems to us more satisfactory than the kilometric distance between regional capitals, which is often used. An even more satisfactory alternative would certainly have been to measure the distance not between the regional capitals but between the actors' precise locations. Unfortunately, the data used are not geo-localised, which prevents us from taking that approach.

Methodologically, we propose presenting, first, various descriptive statistics that capture that temporal evolution of geographical proximity by distinguishing all CNRS contracts from the CIFRE contracts. We seek to verify the existence or absence of temporal trends by conducting non-parametric tests (Kendall's tau) between each variable and the time t. We then propose estimating two complementary econometric models. The first model is a multinomial logit, with the TYPO variable as the explained variable. The second model is a Heckman model with selection equation: the first equation (selection equation) estimates the likelihood that the collaboration between two actors is interregional (INTER=1) rather than infraregional (INTER=0); the second equation (output equation) estimates the time-distance by train between the two collaborators.

In both models, the explanatory variables are the same: i) the year of the contract (ANNEE), ii) the nature of the contract (CIFRE vs. CNRS), iii) the company's sector (INDUSTRY, SERVICES, OTHERS), iv) the laboratory's scientific domain (PHYSICS-MATHEMATICS, CHEMISTRY, ASTROPHYSICAL SCIENCES, LIFE SCIENCES, HUMAN AND SOCIAL SCIENCES, ENGINEERING SCIENCES, INFORMATION).

4. Results

4.1 Descriptive statistics

After measuring the evolution of geographical proximity over the study period, we first constructed three complementary graphs (cf. graph 1a, 1b, 1c). The first one shows the evolution for all CNRS contracts of the share of each mode in our typology. The second graph does the same for all CIFRE contracts. The third measures the evolution of the average time-distance between the two regional capitals where the partners are located.

hal-00725073, version 2 - 24 Aug 2012

[GRAPH 1a, 1b, 1c]

Several of the results are worth highlighting.

In the case of CNRS contracts (graph 1a), the most noteworthy result is the significant drop in the share of interregional contracts involving the capital region (ICAP): for the period, it goes from nearly 60% to around 40%, a drop of twenty points in about twenty years. How can this significant drop be explained? One possible hypothesis, in our view, has to do with the spatial dispersal of CNRS researchers, which led to a rebalancing in favour of the provincial regions and to the detriment of the capital region. Unfortunately, we do not have the data necessary to confirm this hypothesis, but it does correspond to the results obtained at the start of our period of observation by Grossetti and Nguyen (2001). We also see an increase in the share of CNRS contracts within the same department (DEP), from around 10% to more than 20%. The typology's other modes seem more stable.

Comparatively speaking, the evolution observed for CIFRE contracts (graph 1b) shows a very strong inertia in the geography of the collaborations, with the share of each mode seeming especially stable. A comparison of the two graphs (1a and 1b) thus shows that the evolution of the geography of collaborations differs greatly according to the nature of the science-industry contracts (CIFRE vs. CNRS). However, if we focus on the end-of-period data, we see that the classification of the various typology modes is the same for both types of contracts: these are interregional relationships, with the capital region dominating, followed by intradepartmental relationships and then by infraregional relationships between departments.

The graph (1c) illustrating the evolution of the average time-distances also provides interesting clarifications. If we focus first on all contracts (CIFRE_TOTAL and CNRS_TOTAL curves), we observe that CIFRE contracts are implemented with partners who are geographically closer than the CMRS contracts. This can be viewed as a confirmation of the hypothesis formulated in our introduction: under CIFRE contracts, the doctoral candidate must alternate periods with the firm and periods at the laboratory; greater geographical proximity between actors may be sought in order to reduce travel costs – and more generally all the transaction costs – that the doctoral candidate has to bear. In the case of the subset of interregional contracts, the average distances differ only slightly between the two types of contracts. However, we do note one important difference: the average distance appears stable for CIFRE contracts, whereas it increases for CNRS contracts.

To test the validity of these evolutions, we tried to identify the existence of increasing or decreasing trends. Considering the small number of points in our sample (20 years), we opted for nonparametric rank-correlation tests (Kendall's tau).

[TABLE 2]

The results observed confirm the reading of the graphs. We see a strong inertia for CIFRE contracts, since only the REG mode shows a rather insignificant trend. In the case of CNRS contracts, all modes present significant trends. The strongest coefficients are seen for intradepartmental relationships (increasing trend) and for relationships with the capital region (decreasing trend). An analysis of the time-distances over the study period also shows an increasing trend for interregional relationships in CNRS partnerships. The nonparametric rank-correlation tests thus confirm the existence of a special dynamic that differs according to the nature of the contract: a relative stability in the geography of CIFRE contracts vs. a significant evolution for CNRS contracts.

hal-00725073, version 2 - 24 Aug 2012

hal-00725073, version 2 - 24 Aug 2012

4.2. Estimate of econometric models

To confirm or invalidate these initial results by including any sectoral effects (business sectors on the company side and scientific domains on the laboratory side), we proceeded to estimate the two econometric models presented above. The first model is a multinomial logit, with the TYPO variable as the explained variable. The multinomial logit's reference mode is the DEP variable, which takes the value 1 if the contract is interdepartmental and the value 0 otherwise.

[TABLE 3]

For the evolution of the geography of collaborations, we find that the ANNEE [year] variable is significant for three of the spatial scales (REG, ICAP and IREG), with a negative coefficient for the first two and positive for the last. This confirms the dual evolution of the geography of collaborations: increase in the share of local collaborations and increase in the share of the most distant interregional collaborations. As for the nature of the collaborations, hen CIFREW variable is still significant, and the coefficients are systematically negative. These results demonstrate the far more local nature of the CIFRE contracts, which can be explained, as we have said, by the constraint on the doctoral candidate to alternate periods at the firm and periods in the laboratory.

We also observe sectoral effects. In the case of scientific domains, the astrophysical sciences have systematically negative and significant coefficients, which means that this is a domain with which we can associate far more local contracts than for the reference domain (mathematics and physics). In contrast, the chemical sciences have positive coefficients for three of the four modalities, which indicates a less local geography. Another finding: the human and social sciences have a very significant negative coefficient for the IREG and ICAP modes – in other words, for the more distant collaborations. This scientific domain is therefore favourable to more local collaborations, at the level of departments, region or adjacent regions. For the life sciences, the results are more contrasting: negative coefficients for REG and ICAP, positive coefficients for IREG and LIMITROPHE, signs of contradictory tendencies in terms of evolution of the geography of collaborations for this scientific domain. The last two domains (engineering sciences and information sciences) present few significant coefficients.

In terms of business sectors, the main result concerns the services sector, compared to the reference sector (industry). In fact, the services sector presents systematically significant (at the 1% threshold) and negative coefficients. In relation to industry, this sector therefore has a very strong influence on the likelihood that the contract is local (at the department level). For the more heterogeneous set of other sectors, the coefficients are also significant; three are negative (REG, IREG, LIMITROPHE) and one is positive (ICAP).

The second model is a sample selection model (Heckman two-step), with, as variable, the interregional vs. infraregional nature of the collaboration in the selection equation and time-distance by train between the capitals of the regions where the actors are located in the output equation.

[TABLE 4]

First of all, the model appears to be satisfactory overall. We should also note that the Mills ratio is significant, which confirms the choice of the econometric model adopted. Several interesting, more specific results need to be underscored. In the case of temporal effects, first, we see a negative effect

of the ANNEE [year] variable on the likelihood that the contract is interregional, proof of the tendency toward reinforcement of the role of geographical proximity over the period, which we have already reported. However, at the same time, with this model we see that this variable exerts a positive effect on the distance between the partners for interregional contracts. Hence we again find one essential result running through our article: the evolution of the geography of collaborations seems to take the form of a reinforcement of both the most local relationships and the most distant relationships.

The CIFRE variable exerts the same influence as the ANNEE [year] variable: a very significant negative influence on the likelihood that the contract is interregional and at the same time a positive but rather insignificant influence on the distance between the interregional partners.

In the case of sectoral effects, there are not very many of them, and they are not very significant in the output equation, with the exception of the "Others" business sector, which result is, however, difficult to interpret. On the other hand, several very negative effects are to be noted in the selection equation: the human and social sciences, on the scientific domain side, and the services on the business sector side, exert a very significant negative influence on the likelihood that the contract is interregional; chemistry and the engineering sciences, in contrast, exert a very significant positive influence on this same likelihood.

5. Conclusion

The goal of this article was to contribute new empirical results on the role of geographical proximity in science-industry collaborations and on the evolution of its role over time. Based on two long-term relational databases, and the construction of two complementary indicators of geographical proximity, we were able to conduct a comparative and dynamic analysis of science-industry collaborations.

We were also able to answer the question, "Does geographical proximity still matter?" and contribute a truly new result: the econometric tests conducted in fact show that proximity continues to count still today, which in particular is translated by the not-insignificant weight of intradepartmental collaborations (over 20%), stable for CIFRE contracts and increasing for CNRS contracts. The dynamic perspective enables us to clarify this result and reveals that, at the same time, the most distant interregional collaborations increase over the study period. Therefore, we arrive at a combination of reinforcement of the most local and the most distant collaborations. Such a spatial dynamic in collaborations, with evolutions in *a priori* antagonistic directions, had already been observed by Hoekman et al. (2010) at the European level.

A second important result appears: we see significantly different evolutions depending on the nature of the contracts; the role of geographical proximity seems more stable and more significant in CIFRE contracts than for CNRS contracts. This result shows the necessity of distinguishing collaborations according to their characteristics. Not only does the role of geographical proximity differ according to the nature (science-industry vs. intercompany) of the collaboration, as Ponds et al. (2007) showed, but also, within science-industry collaborations, depending on the type of contract (CIFRE vs. CNRS). Finally, our work confirms the existence of sectoral differences, already observed by Ponds et *al*. (2007) for the scientific domains.

hal-00725073, version 2 - 24 Aug 2012

In the end, the analysis of geographical proximity over a long period reveals more contrasting evolutions than is generally imagined. This has strong implications for public policies: these policies generally tend to be based on assumed transversal regularity, which is not the case. Therefore these policies need to be especially fine-tuned, since their results depend on the precise characteristics of the collaborations, as we have been able to show. However, it should be recalled that the results obtained concern France, and that, obviously, they need to be reproduced for other countries in order to assess how general they are.

6. Bibliography

AUTANT-BERNARD C., BILLAND P. and 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;

BOSCHMA R. (2005) Proximity and innovation: a critical assessment, *Regional studies* 39 (1), 61-74;

BUNNELL T. and COE N. (2001) Spaces and scales of innovation, *Progress in Human Geography* 25, 569-589;

BRESCHI S. and LISSONI F. (2003) Mobility and social networks: Localized knowledge spillovers revisited, CESPRI working paper nr. 142, March

CAIRNCROSS F. (1997) *The Death of Distance: How the Communications Revolution will Change Our Lives*, Harvard Business School Publishing, Cambridge.

CANIELS M. (1997) The geographic distribution of patents and value added accross european regions, *Working Paper, Merit,* aout, 10;

CASTELLS M. (1996) *The Rise of the Network Society*, Blackwell Publishers, Oxford.

D'ESTE P. and IAMMARINO S. (2010) The spatial profile of university-business research partnerships, *Papers in Regional Science* 86, 423-443;

DREJER I. and VINDING A. (2007) Searching Near and Far: Determinants of Innovative Firms' Propensity to Collaborate Across Geographical Distance, *Industry and Innovation* 14(3), 259-275;

FELDMAN M. (1994) *The geography of innovation*, Kluwer Academics Publishers, Dordrecht.

FORAY D. (1995) Distribution et expansion de la base de connaissances scientifiques et technologiques, *Revue STI,* OCDE, 14-71;

FRENKEN K. (2002) A new indicator of european integration and an application to collaboration in scientific research, *Economic Systems Research* 14(4), 345-361;

FRENKEN K., PONDS R., VAN OORT F.G. (2010) The citation impact of research collaboration in science-based industries: a spatial-institutional analysis, *Papers in Regional Science* 89(2), 351-371;

FRENKEN K., HARDEMAN S. and HOEKMAN J. (2009) Spatial scientometrics: towards a cumulative research program, *Journal of Informetrics* 3 (3), 222–232;

GERTLER M.S. (2003) Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there), Journal of Economic Geography, 3, 75-99;

hal-00725073, version 2 - 24 Aug 2012

GILLY JP. and TORRE A. (2000) Dynamique de proximité, Ed. L'Harmattan, Paris.

GIURI P. and MARIANI M. (2007) Inventors and the geographical breadth of knowledge spillovers, *druid working paper 08-01.*

GROSSETTI M. and NGUYEN D. (2001) La structure spatiale des relations science-industrie en france: l'exemple des contrats entre les entreprises et les laboratoires CNRS, *Revue d'Economie Régionale et Urbaine* 2, 311-328;

HOEKMAN J., FRENKEN K. and TIJSSEN R. (2010) Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe, *Research Policy* 39 (5), 662-673;

LAURSEN K., REICHSTEIN T. and SALTER A. (2011) Exploring the Effect of Geographical Proximity and University Quality on University-Industry Collaboration in the United Kingdom, *Regional Studies*, 45: 4, 507-523;

KATZ J.S. (1994) Geographic proximity and scientific collaboration, *Scientometrics*, 31 (1), 31-43;

KNOBEN J. and OERLEMANS L. (2006) Proximity and inter-organizational collaboration: a literature review, *International Journal of Management Review* 8 (2), 71–89;

LALLEMENT R., MOUHOUD EM. and PAILLARD S. (2007) Polarisation et internationalisation des activites d'innovation: incidences sur la specialisation technologique des nations», In A. Rallet and A. Torre (dir), *Quelle proximite pour innover?*, Ed. L'Harmattan.

LAURSEN, K., REICHSTEIN, T. and SALTER, A. (2010) Exploring the effect of geographical proximity and university quality on university-industry collaboration in the United Kingdom, *Regional Studies 45 (4)*, 507-523;

LEVY R., ROUX P. and WOLFF S. (2009) Study of science-industry collaborative patterns in a large european university, *Journal of Technology Transfer* 34 (1), 1-23;

LUBLINSKI E. (2003) Does geographic proximity matter? Evidence from clustered and non-clustered aeronautic firms in Germany**,** *Regional Studies*, 37(5), 453-467

LUUKKONEN T., TIJSSEN R.J.W., PERSSON O., SIVERTSEN G. (1993) The measurement of international scientific collaboration, *Scientometrics* 28, 15-36;

MAGGIONI A. and UBERTI T. (2009) Knowledge networks across europe: which distance matters?, *The Annals of Regional Science* 43(3), 691-720;

MAGGIONI M., NOSVELLI M. and UBERTI T. (2007) Space versus networks in the geography of innovation, *Papers in Regional Science* 86 (3), 471-493;

MAGGIONI M. and UBERTI T. (2007) Inter-regional knowledge flows in Europe: an econometric analysis," In Frenken K. (eds) Applied evolutionary, economics and economic geography, Edward Elgar, Cheltenhal, UK.

MASSARD N. and MEHIER C. (2009) Proximity and innovation through an "accessibility to knowledge" lens," *Regional Studies* 43 (1), 77-88;

MORGAN K. (2004) The exaggerated death of geography: learning, proximity and territorial innovation systems, *Journal of Economic Geography*, 4(1), 3-21;

hal-00725073, version 2 - 24 Aug 2012

NARIN, F., STEVENS, K., WHITLOW, E.S. (1991) Scientific cooperation in europe and the citation of multinationally authored papers, *Scientometrics* 21 (3), 313-323;

OINAS P. (2000) Distance and learning: does proximity matter? In s. bakkers, R.P.J.H. RUTTEN, K. MORGAN, AND F. BOEKEMA (EDS), *Learning regions, theory, policy and practice*, ALDERSHOT, 57-69;

PATEL, E and K. PAVITT (1994) Uneven (and divergent) technological accumulation among advanced countries: evidence and a framework for explanation, *Industrial and Corporate Change* 3, 739-787;

PUGA D. (1999) The rise and fall of regional inequalities, *European Economic Review* 43 (2), 303-334;

PONDS R., VAN OORT F. and FRENKEN K. (2007) The geographical and institutional proximity of research collaboration, *Paper in Regional Science* 86 (3), 423-443;

SCHERNGELL T. and HU Y. (2011) Collaborative Knowledge Production in China: Regional Evidence from a Gravity Model Approach, *Regional Studies*, 45(6), 755-772;

STERNBERG R. (1999) Innovative Linkages and Proximity: Empirical Results from Recent Surveys of Small and Medium Sized Firms in German Regions, *Regional Studies* 33 (6), 529-540;

VERSPAGEN B. (1999) Europe: one or several systems of innovation?, IN J. FAGERBERG, P. GUERRIERI and B. VERSPAGEN (eds), *The economic challenge to europe.*

WAGNER C. and LEYDESDORFF L. (2005) Network structure, self-organization, and the growth of international collaboration in science, *Research Policy* 34 (10), 1608-1618;

hal-00725073, version 2 - 24 Aug 2012

hal-00725073, version 2 - 24 Aug 2012

Figure 1a: Evolution in the share of the various spatial scales (CNRS contracts)

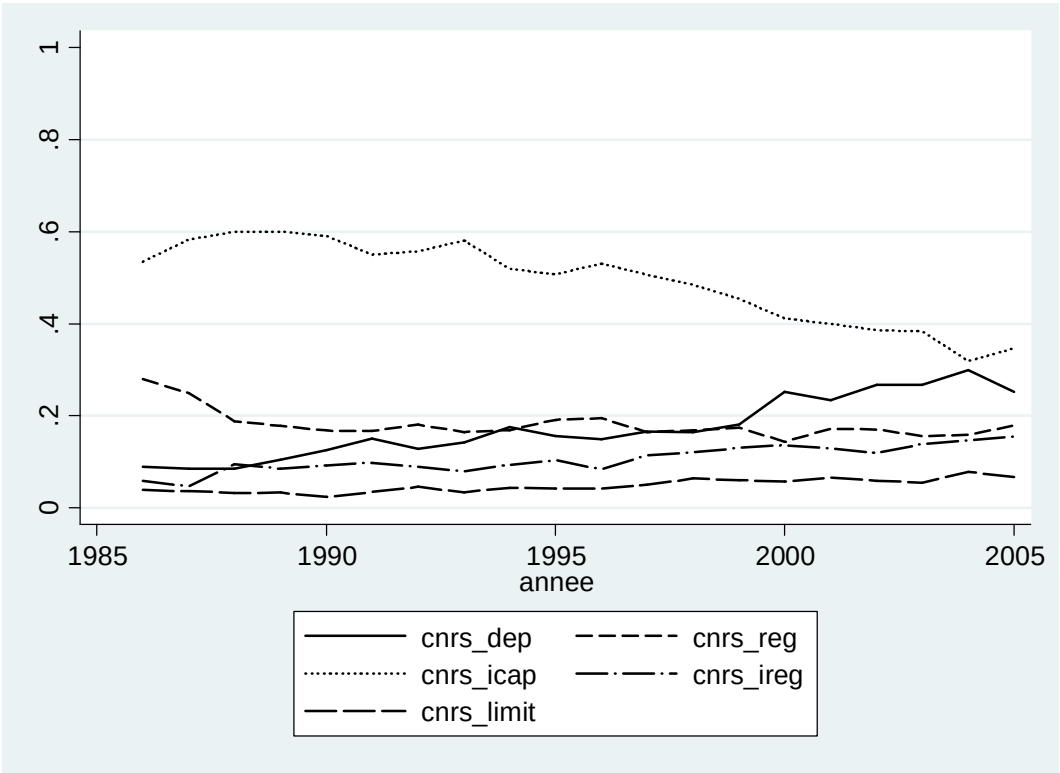
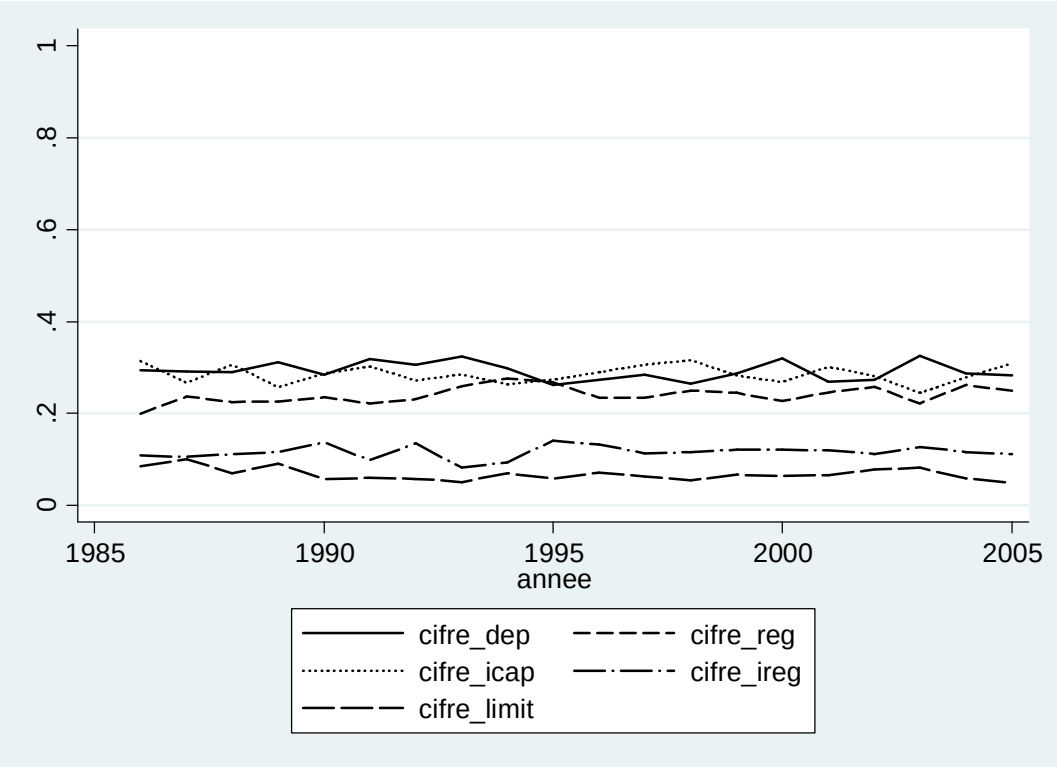


Figure 1b: Evolution in the share of the various spatial scales (CIFRE contracts)



hal-00725073, version 2 - 24 Aug 2012

Figure 1c: evolution in time-distances by train

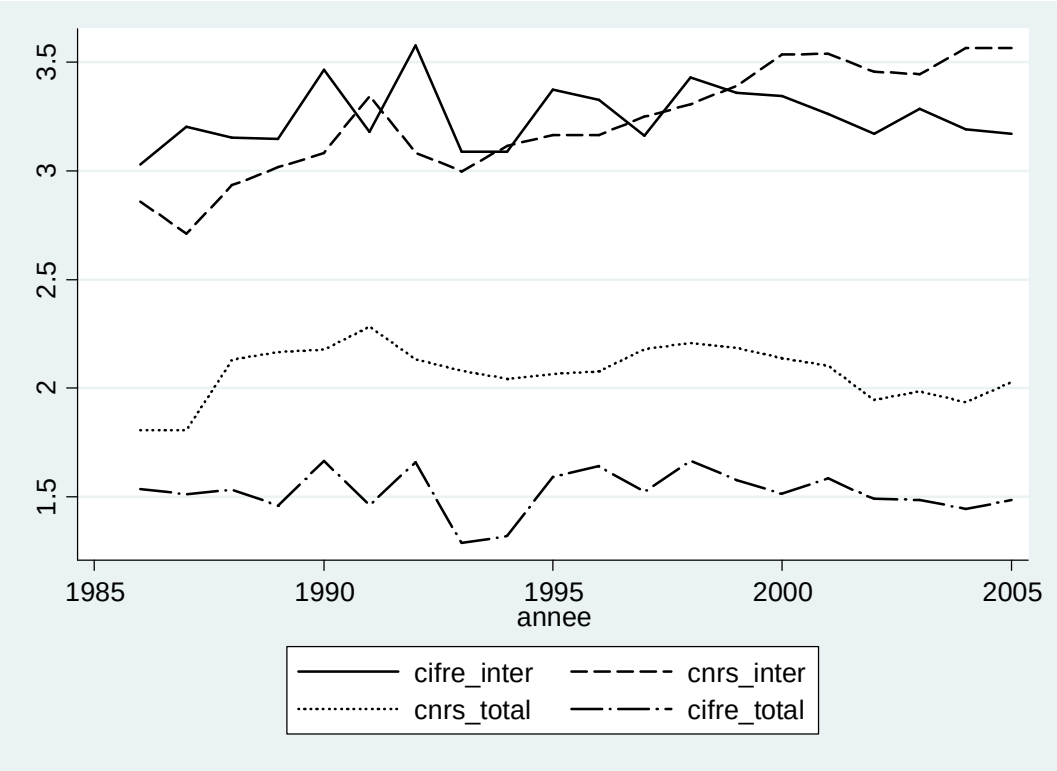


Table: available data

Variable	Modes	number of people	share
Nature of the contract	CNRS	23954	73.1%
	CIFRE	8810	26.9%
	Total	32764	100.0%
Scientific domain	Physics, maths (pnc)	3330	10.2%
	chemistry (sch)	8588	26.2%
	Astrophysical sciences ((sdu)	2122	6.5%
	Life sciences (sdv)	5512	16.8%
	Human and social sciences (shs)	2179	6.7%
	Engineering sciences (spi)	6334	19.3%
	Information sciences (sti)	4699	14.3%
	Total	32764	100.0%
Business sector	Industry	16142	49.3%
	Services	11816	36.1%
	Others	4806	14.7%
	Total	32764	100.0%
Typology of spatial scales	intradepartmental contracts (DEP)	7277	22.2%
	infraregional contracts (REG)	6302	19.2%
	Contracts between adjacent regions (LIMITROPHE)	1846	5.6%
	Contracts with the capital region (CAP)	13576	41.4%
	Other interregional contracts (IREG)	3763	11.5%
	total	32764	100.0%

hal-00725073, version 2 - 24 Aug 2012

Table 2: identification of temporal trends

Variable	Level	Kendall's tau	
CNRS	DEP	0.842	***
CNRS	REG	-0.358	**
CNRS	LIMITROPHE	0.653	***
CNRS	ICAP	-0.821	***
CNRS	IREG	0.726	***
CIFRE	DEP	-0.147	
CIFRE	REG	0.295	*
CIFRE	LIMITROPHE	-0.174	
CIFRE	ICAP	-0.021	
CIFRE	IREG	0.111	
CNRS (total)	Train time	-0.095	
CIFRE (total)	Train time	-0.126	
CNRS (inter)	Train time	0.832	***
CIFRE (inter)	Train time	0.105	

hal-00725073, version 2 - 24 Aug 2012

hal-00725073, version 2 - 24 Aug 2012

Table 2: results of the multinomial logit (base outcome: DEP

		Coef.	Std. Err.	z	p>|z|	signif.
REG	annee	-0.034	0.004	-9.58	0.00	***
	cifre	-0.378	0.042	-9.08	0.00	***
	sch	-0.096	0.065	-1.47	0.14	
	sdu	-0.634	0.091	-6.98	0.00	***
	sdv	-0.171	0.068	-2.54	0.01	**
	shs	-0.011	0.079	-0.14	0.89	
	spi	-0.590	0.067	-8.76	0.00	***
	sti	-0.070	0.068	-1.04	0.30	
	services	-0.662	0.041	-15.99	0.00	***
	others	0.365	0.059	6.23	0.00	***
	_cons	67.941	7.048	9.64	0.00	***
LIMITROPHE	annee	-0.006	0.005	-1.15	0.25	
	cifre	-0.172	0.063	-2.72	0.01	***
	sch	0.241	0.103	2.35	0.02	**
	sdu	-0.499	0.148	-3.36	0.00	***
	sdv	0.324	0.103	3.14	0.00	***
	shs	-0.160	0.129	-1.24	0.22	
	spi	-0.070	0.105	-0.67	0.50	
	sti	-0.051	0.110	-0.47	0.64	
	services	-0.210	0.060	-3.49	0.00	***
	others	-0.170	0.102	-1.67	0.10	*
	_cons	11.104	10.715	1.04	0.30	
ICAP	annee	-0.047	0.003	-15.42	0.00	***
	cifre	-1.186	0.038	-31.2	0.00	***
	sch	0.104	0.059	1.75	0.08	*
	sdu	-0.278	0.076	-3.67	0.00	***
	sdv	-0.181	0.062	-2.9	0.00	***
	shs	-0.214	0.076	-2.82	0.01	***
	spi	0.058	0.059	0.97	0.33	
	sti	0.096	0.062	1.54	0.12	
	services	-0.725	0.035	-20.46	0.00	***
	others	0.755	0.051	14.78	0.00	***
	_cons	95.878	6.143	15.61	0.00	***
IREG	annee	0.009	0.004	2.13	0.03	**
	cifre	-0.421	0.050	-8.42	0.00	***
	sch	0.336	0.079	4.26	0.00	***
	sdu	-0.229	0.107	-2.14	0.03	**
	sdv	0.258	0.081	3.18	0.00	***
	shs	-0.638	0.116	-5.51	0.00	***
	spi	0.060	0.080	0.76	0.45	
	sti	-0.248	0.088	-2.82	0.01	***
	services	-0.665	0.047	-14.08	0.00	***
	others	-0.339	0.077	-4.38	0.00	***

_cons	-18.135	8.369	-2.17	0.03	**
Number of obs	32764		LR chi2(40)	3660.910	
Log likelihood	-44921.268		Prob > chi2	0.000	***

Table 1: result of the sample selection model (Heckman two-step)

	Coef.	Std. Err.	z	p>|z|	Signif.
Output equation (explained variable: train)					
annee	0.065	0.021	3.030	0.002	***
cifre	1.945	1.008	1.930	0.054	*
sch	-0.609	0.336	-1.810	0.070	*
sdu	0.289	0.233	1.240	0.215	
sdv	0.266	0.211	1.260	0.207	
shs	0.606	0.487	1.240	0.213	
spi	-0.720	0.474	-1.520	0.129	
sti	-0.364	0.203	-1.790	0.073	*
services	0.930	0.499	1.860	0.062	*
others	-1.107	0.417	-2.650	0.008	***
_cons	-121.817	40.673	-3.000	0.003	***
Selection equation (explained binary variable: inter)					
annee	-0.009	0.001	-6.570	0.000	***
cifre	-0.452	0.018	-25.800	0.000	***
sch	0.133	0.027	4.900	0.000	***
sdu	-0.004	0.037	-0.110	0.915	
sdv	0.044	0.029	1.520	0.128	
shs	-0.168	0.036	-4.670	0.000	***
spi	0.205	0.028	7.420	0.000	***
sti	0.035	0.029	1.210	0.226	
services	-0.224	0.017	-13.320	0.000	***
others	0.200	0.022	9.180	0.000	***
_cons	18.915	2.831	6.680	0.000	***
mills ratio	-7.073	3.735	-1.890	0.058	*
Number of obs.	32764		Wald chi2(10)	49.190	
Censored obs.	13579		Prob > chi2	0.000	***
uncensored obs.	19185				

hal-00725073, version 2 - 24 Aug 2012