



HAL
open science

Geographical mobility of PhDs: an analysis of French long-term data

Bastien Bernela, Olivier Bouba-Olga, Marie Ferru

► **To cite this version:**

Bastien Bernela, Olivier Bouba-Olga, Marie Ferru. Geographical mobility of PhDs: an analysis of French long-term data. 2013. hal-00701750v3

HAL Id: hal-00701750

<https://hal.science/hal-00701750v3>

Preprint submitted on 6 May 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Geographical mobility of PhDs: an analysis of French long-term data

BERNELA Bastien, BOUBA-OLGA Olivier, FERRU Marie

bastien.bernela@univ-poitiers.fr ; obouba@univ-poitiers.fr ; marie.ferru@univ-poitiers.fr

CRIEF EA 2249, UFR Sciences économiques

2 rue Jean Carbonnier, Bât.A1, BP 623, 86022 POITIERS Cedex, FRANCE

Temporary version

Abstract: The aim of this paper is to provide empirical analysis of the geographical mobility of PhDs during the beginning of their career, by using long term data (1970-2000) about France. First of all, we highlight the low mobility of PhDs: more than 60% of them supervise their first PhD in the region where they defended their own one. Then, thanks to a second database, we show that this mobility level is not specific to the Higher Education and Research system. We test the impact of structural determinants with a gravity model (sample selection model) and we observe in particular the influence of the scientific size of regions and spatial distance on the PhDs' mobility.

Keywords: Geographical mobility, PhDs, size effects, distances

1. Introduction

Public policy on research and innovation is acutely concerned with the scientific population. Enquiries into the system of Higher Education and Research (HER) are regularly set up at both national (French) and European levels. The European Union, committed to the Lisbon Strategy and “the era of knowledge”, paying particular attention to the establishment of a unified region of research and mobility, makes a telling contribution to such reflections (CEC, 2001, 2008). Work on the mobility of researchers is especially important given the need to move forward in various fields of research.

Among these, a group of investigations links the question of the mobility of researchers to that of the spatial diffusion of knowledge. Associated with the field of the geography of innovation, they seek to understand the role of mobility as a vector in the circulation of knowledge between countries and regions (Jaffe et al., 1993; Almeida and Kogut, 1999; Zellner, 2003; Criscuolo, 2005; Agrawal et al., 2006; Zucker and Darby, 2008; Breschi and Lissoni, 2009; Edler et al., 2011; Tripp, 2011; Franzoni et al., 2012). The mobility of researchers is also addressed by the burgeoning and controversial literature on the creative class. Florida (2002) argues that researchers have, like all creative people, intense geographic mobility, with movement to the cities being considered “attractive”. However, several authors (Shearmur, 2007; Martin-Brelot et al., 2010) cast doubt on this result. Another series of works, initiated mainly by American researchers, analyzes the impact of mobility on research productivity and career advancement (Allison and Long, 1987; Debackere and Rappa, 1995; Bozeman and Mangematin, 2004; Hoisl, 2007; De Filippo et al., 2009; Cruz-Castro and Sanz-Menendez, 2010; Latham et al., 2011; Barrufaldi and Landoni, 2012). Some therefore approach the issue of mobility by querying the recruiting practices at the University and the functioning of the academic labor market (Gaughan and Robin, 2004; Combes et al., 2008; Musselin, 2009; Bonnal and Giret, 2010; Pezzoni et al., 2012).

The study of mobility is often made difficult by the lack of available data. The scientific community rarely has access to long-term data on the trajectory and location of researchers. When one identifies the variety of topics related to mobility, it is clear that this lack of data hinders progress in understanding the process of mobility, its determinants and its impacts. The present paper aims to empirically examine the spatial mobility of French researchers and its structural determinants. An original long-term (1970-2000) database, relating to more than 12,000 PhDs in France and listing the transitions between thesis defense and first thesis supervision, is utilized.

Knowing the location of the thesis defense and the supervision of a thesis, we can focus on these particular stages of researchers’ career paths and investigate the geographical dimension. Does an academic tend to supervise his/her first thesis in the same place he/she defended his/her own thesis? In what proportion? If he/she goes to another University, where does it tend to be situated? Can one identify certain determinants of any mobility observed? In this context, we emphasize the need to distinguish two observational scales - the scale of the establishment involved and that of the region, in order to verify the existence of spatial proximity effects (sub-regional mobility).

This paper also analyzes the determinants of interregional mobility of researchers to more accurately assess the role of certain structural features of regions. While many authors make modalities of the recruitment process the key determinant of the (non-) mobility of researchers,

we hypothesize that the size of the regions and interregional distances (geographic and disciplinary) potentially influence observed mobility. For this, we test a gravity model (Heckman's model with a selection equation): the results show the particularly strong impact of the scientific size of regions and of the geographical distance that separates them.

To make a comparison, a complementary database has been chosen. The latter is the result of an investigation by CEREQ¹ in 1998 and provides information on the career paths of nearly 10,000 individuals over the ten years after leaving the education system (all levels of qualification taken together). This data source allows us to relate the mobility behavior of French PhDs to that of other categories of the population in order to verify or refute their specificity.

The paper is organized as follows. In Section 2, a review of empirical studies on the mobility of researchers is undertaken. The advantages and disadvantages of methods identifying mobility are revisited. In Section 3, the data and the method employed are described. Section 4 presents the main results. The paper concludes by reflecting on the implications of our findings for research policy.

2. Literature review

The review of empirical work on the mobility of researchers first reveals the small number of studies focusing exclusively on the spatial mobility of researchers and their determinants. It also allows the identification of two major methods of investigating mobility. The advantages and limitations are discussed.

A first set of empirical investigations studies the mobility of researchers based on patents and/or publications data, in line with Jaffe et al. (1993). These aim to identify inventors/authors' affiliations to the different dates of patents/publications, and to deduce whether or not there was mobility by comparing the locations at these events. This method has the advantage of providing a lot of peripheral information on mobility and allows the researcher to draw simultaneously from a single data source: their spatial trajectory (*via* institutional affiliation), the collaborative processes involved (*via* co-productions) and networking effects (*via* quotations/references). This work therefore generally couples the issue of mobility to that of the spatial diffusion of knowledge (Almeida and Kogut, 1999; Agrawal et al., 2006; Breschi and Lissoni, 2009; Latham et al., 2011). The large size of these databases makes it a very powerful research tool.

However, this method has the main disadvantage of being approximate and of assuming mobility. The input of the analysis involving the patent or publication rather than the researcher him/herself, the data used cannot reliably represent the spatial career paths of researchers. The patent or publication is a moment in that career and enables one to know its location for each moment identified, without giving the possibility to interpret any changes occurring between these two moments. Thus, this type of data obscures the periods of "non-production" in the career paths of individuals. In addition, Katz and Martin (1997) highlight the problem of dual institutional affiliation of many researchers: a PhD may be affiliated to two laboratories in different cities without one being able to discriminate between these two affiliations. Primary affiliation does not necessarily indicate the place of residence. Finally, these data have the disadvantage that they cover only certain areas of science: data on patents tends to exclude

¹ Centre d'Etudes et de Recherches sur l'Emploi et les Qualifications

fields where scientific production does not consist in filing patents (Griliches, 1990). This is mainly the case for the human and social sciences. We find the same type of bias in studies based on publications: "bibliometric indicators (...) are mostly obtained from bibliographic databases that cover journal articles; thus measuring only one aspect of scientific activity" (De Filippo et al., 2009, p.192). The identification of researchers by their patents or publications results in an arbitrary division of the scientific community, introducing a selection bias. The "producers" may have individual characteristics and specific mobility behavior, this calling into question the representativeness of the results and their possible systematization.

Aware of the limitations of using this first type of data, a second group of investigations was developed (Allison and Long, 1987; Debackere and Rappa, 1995; De Filippo et al, 2009; Cruz-Castro and Sanz-Menendez, 2010; Martin-Brelot et al, 2010; Canibano et al, 2011; Edler et al, 2011). These works reconstruct the researchers' trajectories from *curriculum vitae* (Dietz et al., 2000), online surveys or interviews. They have the intrinsic quality of collecting sufficiently accurate data on careers "rather than quantifying particular knowledge products at particular times" (Canibano et al., 2011, p.655). They thus allow the study of "the researchers' career trajectories, spatial mobility and mapping of collective capacity" (*ibid.*). The scope of the study, and more specifically the targeted individuals, are this time under the control of the researcher who will be able to define the population studied (according to geographical and/or disciplinary limitations). Creating observations directly related to the trajectory of researchers has the undeniable advantage of generating data more truly representing the reality studied. However, data collection of this sort requires a relatively large investment of time. This can restrict the size of the sample (particularly in the case of semi-structured interviews). These restrictions largely vanish when the data is collected by institutions (government ministries, the European Union), provided that researchers are given access to such databases.

Regardless of the method chosen for identifying and measuring mobility, a set of critics must be addressed to existing investigations. Firstly, if mobility is to be necessarily studied on a dynamic basis (locations at time t and time $t + 1$), the investigations are often undertaken on short-term data. Our database covers thirty years (1970-2000) and reflects a long-term reality. The results then are less dependent on contextual factors. A second limitation of existing studies on mobility is the lack of comparison between groups. Some research focuses on academics (public research), others on engineers (private research), but to our knowledge, no study aims to compare the mobility behavior of these two categories of people. We seek therefore to compare the mobility behavior of PhDs working in higher education and research with that of other PhDs. Finally, existing work focuses more on individual characteristics to explain mobility and tends to obscure the structural determinants. Econometric models (*probit* or *logit*) test the impact of individual determinants on researchers' mobility, such as age, sex or academic position. However, the influence of structural factors appears to be too rarely measured, even when it is extensively tested (and appears significant) in the case of research collaborations. Effects related to the size of the regions have been highlighted by Grossetti and Nguyen (2001), Hoekman et al. (2010) and Bouba-Olga et al. (2012). Similarly, Maggioni et al. (2007), Ponds et al. (2007), Frenken et al. (2009) and Bouba-Olga et al. (2012) show the significant role of distances between regions, a distinction being made between geographic distance and cognitive distance (Boschma, 2005). In a continuation of this work, the present investigation seeks to determine if the mobility of researchers depends on these structural determinants in the same way that research collaborations appear to do. It is supposed that the interregional mobility of researchers depends positively on the size of the region that individuals identify with, and

negatively on the distances between regions. In other words, an individual would be more likely to move from one region to another when these two regions are geographically or cognitively close. It is also supposed that the larger a region (bringing together a large number of researchers), the correspondingly larger number of departures of researchers to be expected.

Based on this literature review and observed limitations, we propose to complement existing empirical work by: i) exploring long-term French databases (in line with the scholars on the tracking of career trajectories), ii) comparing patterns of mobility between different types of individuals and iii) integrating in an econometric model, derived from international trade, various structural indicators previously largely untested. Each of these three points will be developed in turn in the following two sections.

3. Data and method

3.1 The DOCTHESE database: identifying the transition between PhD thesis defense and first PhD supervision

The lack of long-term data presents difficulties for investigations into the spatial trajectory of researchers. In this paper, we propose to mobilize an original database, DOCTHESE, which compiles in an almost exhaustive manner the PhD theses defended in France between 1970 and 2000 (over 200,000 theses). On each line appears the first and last name of the individual defending his/her thesis, the full name of his/her thesis supervisor(s), the year, the University and the discipline of defense. Having data collected over a long period identifies new PhDs who a few years later become supervisors. This is the case for 12 261² individuals. For each, we could then match data relating to the defense of their thesis with data on the first thesis they supervised. Although the database does not enable the spatial trajectory of PhDs to be comprehensively characterized, this matching work supports the method of trajectory tracking. It is reconstructed with the help of two events, the PhD thesis defense and first PhD thesis supervision, these being significant moments of the professional life of academics.

We thus have a database composed of 12,261 lines that correspond to a transition from defense to first supervision of a PhD thesis. On each line appears the first and last name of the individual whose transition was identified, the full names of the supervisor and of his first PhD student, the year, the University and the discipline of his/her own thesis and that of his/her first PhD student. We deduce from the University of defense and that of the first supervision, the region of defense and first supervision. This allows the observation of different levels of mobility. It is noted that the average time-lapse between PhD defense and first supervision is eleven years.

Godechot and Louvet (2008) have used this database for measuring the importance of local recruitment – or academic inbreeding - in French Universities, defining it as "the PhD defense and supervision of his/her first PhD thesis (...) in the same University and in the same discipline"(p.9, translated by us). They show its quantitative importance, with significant variations across disciplines and regions, Parisian Universities practicing significantly less inbreeding. Faced with such results, they call for a reform of the academic hiring process in France, with the main objective being the prohibition of local recruitment. The analysis

² For more information on the building of this database, see Appendix A.

proposed by these authors, however, suffers from several limitations, highlighted by Bouba-Olga et al. (2008)³. To make pronouncements on the issue of recruitment in University from this database is not appropriate, because no variable encapsulates the nature of recruitment process. Thus we focus on the issue of geographical mobility of researchers at the beginning of their professional career.

To characterize this mobility, we look at, for each French region⁴, the destination of PhDs who have defended their PhD thesis in the respective region. How many of them supervise their first thesis in the same region, in the Ile-de-France region, or in a region bordering the region of the original defense? When there is no mobility, we distinguish two scales of observation:

- that concerning the establishment: the individual defends his/her own first PhD thesis, and supervises his/her first PhD thesis in the same institution, that is to say in the same University;
- that concerning the region: the individual defends his/her PhD thesis and supervises his/her first PhD thesis in the same region.

The distinction is important since it allows the highlighting of the potential effects of spatial proximity. Indeed, in regions composed of a large number of Universities, there can simultaneously be high levels of mobility across institutions, albeit with weak mobility at the regional level. This configuration reveals the existence of sub-regional mobility: PhDs of one University supervise their first thesis at another University in the same region. We will see that this state of affairs is typical for the case of the Ile-de-France.

To investigate the spatial dimension of the University system, another possible mode of enquiry is that relating to “academies”. The division of French territory into academies corresponds to the regional breakdown, except in three cases: the Ile-de-France region is divided into three academies (Créteil, Paris and Versailles), the Provence-Alpes-Côte-d’Azur region into two academies (Marseille and Nice) and the Rhône-Alpes region into two academies (Grenoble and Lyon). For these two last-mentioned regions, the scale of the academy seems more appropriate since there are two University towns widely separated in the same region. Thus the study of mobility between academies should not be overlooked. By contrast, the Ile-de-France region is characterized by a high concentration of establishments in the same geographical continuum: to migrate from one academy to another in this case does not involve a change of agglomeration and therefore does not represent a spatially significant mobility. In other words and in the rest of the paper, the term “region” refers to the nomenclature of administrative regions, with the exception of Rhône-Alpes and Provence-Alpes-Côte-d’Azur which we break down into academies.

This is to bring to attention the interregional flow matrix of French PhDs observed using the DOCTHESE database. To our knowledge, to date such work on the mobility of researchers has not been undertaken. We can, using this matrix, obtain first insights on the level and deployment of mobility of PhDs on French territory.

³ The discussion between these authors can be read on the website *La Vie de Idées*: <http://www.laviedesidees.fr/Le-localisme-universitaire.1566.html>.

⁴ In this paper, the term “region” corresponds to the administrative breakdown of France in NUTS2 regions.

3.2 The CEREQ database: comparing the level of mobility of PhDs

The DOCTHESE database suffers from two major limitations. We propose to overcome them by drawing on a complementary database. The limitations are:

i) The trajectory of PhDs is reconstructed from information about the access to supervision of PhD theses. We therefore do not know what happens between the PhD defense and the first thesis supervision, a period covering eleven years on average. In other words, we have no information on the individual relating to the first ten years after obtaining his/her PhD. Not knowing when movements most often occur in the lives of individuals, we cannot assume that the mobility identified is representative of the career trajectory of the individual.

ii) Using DOCTHESE, we only address those academics who supervise PhD theses. Those who have not made a career in University, who did not choose to supervise theses, or who have not yet obtained the “Habilitation à Diriger des Recherches” (French certification for PhD supervision), escape the database. In addition, as we noted earlier, studies on mobility often suffer from inadequate comparative data between groups of individuals. When considering the geographical mobility of a given population, we can question the specificity of its behavior. It therefore seems important to consider in this paper elements of comparison between the mobility of PhDs and that of nearby populations. Several authors (Greenwood, 1975; Schwartz, 1976; Yankow, 2003) highlight the positive relationship between educational attainment and individual mobility: the more qualified an individual, the more mobile he/she tends to be. We want to test the hypothesis that the PhDs would be the most mobile individuals. Secondly, we establish an essential distinction between those PhDs who engage in higher education and research, and the others. For an equivalent level of qualification, does the “professional world” have a significant impact on mobility?

In relation to this question, we mobilize a database compiled by CEREQ entitled the “enquête génération”. This sets out to follow a cohort of 1998 graduates during the first ten years of their professional lives. This ten year study supplies information on the period missing from the DOCTHESE data. In other words, we are able to take into account the period corresponding to the transition between PhD defense and first supervision.

Within the CEREQ database, composed of 9,121 individuals, we define four types of people: those qualified to the level of baccalaureate or less (representing 49.5% of individuals), those with qualifications lying between baccalaureate and Master (40.7%), individuals holding a Master degree (6.8%) and PhDs (3%). Within this last category, we distinguish between PhDs inserted in University (1%) and outside University (2%). In the rest of this paper we will refer to the former as HER (Higher Education and Research) PhDs and the latter as non-HER PhDs. For each individual, we have information on location (at the regional level) and thus are able to reconstruct his/her spatial trajectory: place of graduation in 1998, and places of work in 2001 and 2008. From this information, we construct two binary variables to capture the spatial trajectory of individuals:

- Mob98-01 = 0 if the individual is located in the same region in 1998 and 2001, Mob98-01 = 1 otherwise;
- Mob98-08 = 0 if the individual had no mobility between 1998 and 2008, Mob98-08 = 1 if there is at least one mobility.

We then estimate the probability of observing mobility between 1998 and 2001 (Mob98-01) and between 1998 and 2008 (Mob98-08) from a *probit* model. The variables of interest in explaining

mobility are the level of qualification and the “professional world” of PhDs as described in the paragraph above. We also include three sociological control variables (age, gender and having/not having a child) and regional dummy variables (region where the individual studied in 1998). The control variables are binary and are constructed as follows:

- The variable "age"⁵ takes the value 1 when the individual is older than the average age of his/her group and 0 otherwise.
- The variable "man" takes the value 1 if the individual is male and 0 otherwise.
- The variable child" takes the value 1 if the individual had a child before graduation and 0 otherwise.

We show that HER PhDs are not any less mobile than Masters and non-HER PhDs and it thus do not seem relevant to attribute the low mobility of researchers to the organization of the University system. This result vindicates the interest not to focus on recruitment procedures, but to explain the spatial mobility by structural determinants, little studied in the literature.

3.3 Testing the impact of structural determinants on PhDs' mobility

After comparison thanks to the CEREQ data, we return to the DOCTHESE data to see if mobility behavior responds to structural variables. We seek to explain the probability of observing mobility between two regions and the number of movements between these pairs of regions. To do this, we test a gravity model with a selection equation (Heckman, 1979) to assess the existence of the size and distance effects on the mobility behavior of researchers. The focus now is only on interregional mobility: individuals who have defended and supervised in the same region are therefore excluded from the model. The tested model, commonly used in research on international trade, has recently been put into use to explain the geography of collaborations (Ponds et al., 2007; Bouba-Olga et al., 2012).

More precisely, we define a typology of regions such that r_1 corresponds to the regions of PhD defense and r_2 to the regions of PhD supervision. The coupling r_1/r_2 thus does not correspond to the coupling r_2/r_1 : there is a direction involved, this being from the region of departure to the region of arrival. We assume first that the existence of mobility for each pair (i) of regions r_1 and r_2 depends on a first latent variable, d_i^* , not observable and continuous, which is a linear combination of a vector of exogenous variables Z_i . The latter are the scientific size of the region where the PhD thesis was defended and the difference in size between that and the region of first PhD supervision, on the one hand. On the other hand there are various forms of distance (spatial distance and difference in disciplinary profile):

$$d_i^* = Z_i\gamma + \mu_{i1}$$

When the variable d_i^* is above a certain threshold, there is at least one movement between the two regions and if it remains below this limit, there is no evidence of mobility, hence:

$$d_i = \begin{cases} 1 & \text{if } d_i^* \geq 0 \\ 0 & \text{if } d_i^* < 0 \end{cases}$$

⁵ Regarding the effect of age, absolute age is collinear with the level of education, so we cannot integrate it as such into the model. This is why we test for a relative age effect: regardless of the effect of level of education, does being older or younger than the average age of the group significantly impact the likelihood of being mobile?

The variable d_i allows us to see when it is possible to observe a number of movements between the two regions, the number being represented by a second latent variable m_i^* explained by the same exogenous variables. To this is added a variable denoted "Ile-de-France" to capture the effect of the capital region:

$$m_i^* = W_i\theta + \mu_{i2}$$

The number of movements m_i for the pair i of regions considered is obtained as follows:

$$m_i = \begin{cases} m_i^* & \text{if } d_i^* \geq 0 \\ 0 & \text{if } d_i^* < 0 \end{cases}$$

We assume that the error terms μ_{i1} and μ_{i2} are normal.

With respect to the explanatory variables of the model, which are indicators of region size and distance between regions, we have constructed these partly from the DOCTHESE database.

The scientific size of regions

To approximate the scientific size of regions, we construct an indicator from the regional distribution of PhD thesis referenced in the original DOCTHESE database. It is, in other words, the contribution of each region to the national production of PhDs⁶. We therefore consider that a region is bigger scientifically when many PhDs have defended their theses in this region. In the model, we integrate the variable of scientific size of the region of departure in order to know if the fact of having defended one's thesis in a large region significantly affects the probability of being mobile. We also test the effect of size difference between regions⁷ to determine the possible tendency of PhDs to go from small regions to large regions, or vice versa.

Spatial distance between regions

To check possible effects of spatial distance, we incorporate a matrix of interregional distances, measured in hours by train between the regional capitals. For this variable, we tested several indicators: time of train journeys between regional capitals, distance in kilometers between the regional capitals and the number of regional borders to cross to reach the other region. The correlation being very high (using Kendall's tau) between all these indicators, we selected the first variable as the indicator of spatial distance between regions.

Distance between disciplinary profiles of regions

This variable is used to test the existence of an effect related to the (non-)similarity of regions in terms of disciplinary structure. It was created from the original data relating to disciplines⁸. First, we calculate for each region i , the part p that represents each of the disciplines j ($p_{i,j}$)⁹. The distance indicator relating the disciplinary profiles of regions 1 and 2 is calculated as follows:

⁶ For more information on the regional distribution of PhDs, see Appendix B.

⁷ It is more precisely the relationship between the size of the region of departure and that of arrival. The variables being expressed as Napierian logarithms, the size differential between the regions $r1$ and $r2$ corresponds to the difference between $\ln(\text{size of } r1)$ and $\ln(\text{size of } r2)$. A positive difference means that the PhD migrates to a smaller area.

⁸ For more information on the distribution of PhDs by discipline, see Appendix C.

⁹ For example, the share of PhDs in economics (discipline 10) in the Poitou-Charentes region (region 20) is named $p_{20,10}$.

$$\text{Distance between disciplinary profiles (r1; r2)} = \frac{\sum_{j=1}^{19} |p_{r1,j} - p_{r2,j}|}{2}$$

This indicator is constructed so that the closer it is to 1, the more the profiles of the two disciplinary regions are considered different. Using this indicator we approximate the impact of the disciplinary structures on observed mobility between the two regions. Do PhDs migrate more to regions that are close from the point of view of the discipline of research? One might think that this element plays a role, since if a discipline is important in the region, the number of positions offered in this discipline would be high. PhDs are therefore more likely to be recruited in a region specialized in their discipline.

To generate results in relation to disciplines, it seems appropriate to decompose the gravity model. Testing a gravity model cohort of PhDs for each discipline is not possible given the small number of observations. This is why we reorganized them into two groups: pure and applied sciences versus the human and social sciences. To check if this breakdown improves the quality of the model, we performed a Chow test, known as a coefficients stability test. From simple regressions between the dependent variable and the explanatory variables of the model, the Chow test shows that there is a significant difference between the sum of squared residuals for the entire population and the addition of the sum of squared residuals calculated from the two sub-populations. This led us to test a gravity model for the two groups of disciplines. All variables are expressed as Napierian logarithm.

4 Results

4.1 The “low” mobility of French PhDs

We propose to analyze mobility occurring between French regions. To do this, we construct an interregional flow matrix in order to characterize patterns of mobility (or immobility) of French PhDs (table 1). On average, 64.0% remain in the region of their PhD thesis defense, 8.9% go to the Ile-de-France region, 7.8% to neighboring regions and 19.3% to non-neighboring regions. In other words, the rate of interregional mobility of PhDs who become supervisors is only 36.0%. If we remove the mobility to neighboring regions, it comes down to less than a third of the individuals studied.

The difference between the first and the second column is the rate of mobility between institutions within the same region. This difference is thus zero for regions with one University. It can be noted that the variation between the two measures is particularly sensitive. The rate increased from 43.9% on average for all French Universities to 64.0% at the regional level. In other words, one fifth of the mobility in the database – 2,465 individuals – is observed between two Universities of the same region.

Ile-de-France is heavily involved in this sub-regional mobility: only 29.6% of PhDs stay in the same University (score very much lower than in other regions), even though 61.2% stay in the region (near the average for the other regions). This result highlights the existence of a strong circulation of PhDs within the capital region. Even if the Parisian PhDs rarely first supervise a thesis in the same University that they defended their own thesis, a large number of them still remain in Ile-de-France.

Ile-de-France has a strong power of attraction, this being more marked for its neighboring regions, the PhDs being "sucked in" by the city. This is particularly the case for the Centre and Champagne-Ardenne regions, where the mobility rate is higher than the average for the other regions. We can describe this phenomenon as *the shadow effect* of the capital (Brouillat and Lung, 2010), suggesting that spatial distance partly explains interregional mobility.

Table 1: From the region of PhD thesis defense to the region of first supervision of a PhD thesis

	Same region	Same University	Other University of the region	Ile-de-France	Neighboring regions (apart from IDF)	Other regions	Total
Academy of Grenoble	75.6%	53.9%	21.7%	7.7%	4.4%	12.3%	100.0%
Academy of Lyon	70.3%	50.7%	19.6%	8.5%	8.9%	12.3%	100.0%
Academy of Marseille	68.5%	53.4%	15.1%	6.8%	10.3%	14.4%	100.0%
Academy of Nice	64.8%	57.1%	7.7%	10.0%	7.8%	17.4%	100.0%
Alsace	63.9%	60.0%	3.9%	9.7%	6.5%	19.9%	100.0%
Aquitaine	60.3%	53.6%	6.7%	8.0%	6.9%	24.8%	100.0%
Auvergne	57.0%	57.0%	-	10.3%	23.0%	9.7%	100.0%
Bourgogne *	65.0%	65.0%	-	7.9%	11.5%	15.6%	100.0%
Bretagne	64.3%	58.0%	6.3%	12.3%	6.6%	16.8%	100.0%
Centre *	48.5%	45.5%	3.0%	17.9%	10.4%	23.2%	100.0%
Champagne-Ardenne *	43.5%	43.5%	-	22.2%	13.3%	21.0%	100.0%
Franche-Comté	50.9%	50.9%	-	10.2%	24.1%	14.8%	100.0%
Ile-de-France	61.2%	29.6%	31.6%	-	5.9%	32.9%	100.0%
Languedoc-Roussillon	60.3%	54.0%	6.3%	10.5%	16.0%	13.3%	100.0%
Limousin	67.9%	67.9%	-	1.8%	12.5%	17.8%	100.0%
Lorraine	62.5%	44.4%	18.1%	8.2%	6.0%	23.2%	100.0%
Midi-Pyrénées	70.6%	59.9%	10.7%	8.5%	6.8%	14.0%	100.0%
Nord-Pas-De-Calais	73.7%	65.5%	8.2%	6.7%	1.7%	17.9%	100.0%
Basse-Normandie	52.9%	52.9%	-	8.8%	21.6%	16.7%	100.0%
Haute-Normandie *	74.2%	69.7%	4.5%	7.9%	0.0%	17.9%	100.0%
Pays-De-Loire	67.8%	67.1%	0.7%	9.2%	7.2%	15.8%	100.0%
Picardie *	68.9%	66.7%	2.2%	15.6%	2.2%	13.3%	100.0%
Poitou-Charentes	65.8%	65.8%	-	7.7%	10.7%	15.8%	100.0%
France	64.0%	43.9%	20.1%	8.9%	7.8%	19.3%	100.00%

* neighboring regions to Ile-de-France. NB: When there is no value in the third column, it means that the region has only one University. Rates recorded in the first and second columns are thus identical

4.2 Comparative analysis of the PhDs' trajectories

In order to establish a point of comparison, we utilize a complementary database to compare the mobility behavior of PhDs with that of other populations. Before presenting the results of the *probit* analysis, some descriptive statistics are given in table 2.

Table 2: Level of qualification, age and mobility

	Nb. Obs.	Average age (years)	Mob98-01 (%)	Mob98-08 (%)
≤ Baccalaureate	4516	20.3	16.1	23.8
Baccalaureate < level < Master	3710	23.1	24.5	32.0
Master	622	24.6	45.7	51.8
PhD	273	28.7	33.0	39.6
HER PhD	81	29.0	38.3	40.7
Non-HER PhD	192	28.5	30.7	39.1
Total	9 121	22.0	22.0	29.5

The results obtained with the CEREQ database support those of the previous section. From the DOCTHESE data we obtained a mobility rate of 36.0%. Establishing a three-year horizon, mobility now is observed to be 33.0% for PhDs and 38.3% specifically for HER PhDs. Analysis of the second database thus validates our initial results.

Overall, it appears that individuals below the level of baccalaureate are less mobile, since only one fifth changes region during the three years following graduation. Masters are the most mobile category, with a mobility rate of about 50%. The PhDs are, in turn, less mobile than the Masters. This runs against the existence of a linear and positive relationship between the level of qualification and mobility. Indeed, age has a positive effect on the degree of mobility up to the level of Master, beyond which it negatively affects the degree of mobility. This result can be explained by sociological effects related to individual trajectories (spouse's work, school children, etc.). Beyond a certain age (the average is 28.7 years at the end of a PhD thesis), individuals may be more deeply rooted in their region and mobility behavior becomes less automatic.

The differences between the mobility rates for the two periods, 1998-2001 and 1998-2008, are relatively low, indicating that most movements occur early on in careers. More specifically, of the 2,694 people who had experienced at least one move between 1998 and 2008 (29.5% of the whole database), 2,009 people - nearly 75% of this group - had experienced mobility between 1998 and 2001. This finding reinforces the validity of the method used to analyze the DOCTHESE data, since the career trajectory over ten years is relatively stable, once past first recruitment. Moreover, for HER PhDs this trend is at its strongest, with a difference of only 2.4 percentage points between the two rates of mobility. This result can be explained in terms of the stability of the first employment of HER PhDs, namely the position of assistant professor (Maître de Conférences in France). Such posts often entail decades of employment for an individual. By contrast, the largest difference between the two rates of mobility is observed for non-HER PhDs: 8.4% of them experiencing mobility between 2001 and 2008.

Table 3 presents the results of *probit* models, Part (a) giving the results for mobility at three years, and Part (b) at ten years. Descriptive statistics on the explanatory variables are given in Appendix D.

Table 3: Probability of mobility (marginal effects)

	Mob98-01				Mob98-08			
	Model 1a		Model 2a		Model 1b		Model 2b	
≤ Baccalaureate	-27.4	***	-27.4	***	-27.8	***	-27.8	***
Baccalaureate < level < Master	-17.1	***	-17.1	***	-17.7	***	-17.7	***
Master	ref.		ref.		ref.		ref.	
PhD	-6.6	***			-7.2	**		
HER PhD			-2.3				-5.7	
Non-HER PhD			-8.3	***			-7.9	**
Age	+1.4		+1.4		+0.6		+0.6	
Man	+3.6	***	+3.6	***	+5.9	***	+5.9	***
Child	-10.4	***	-10.4	***	-12.2	***	-12.2	***

*P<0.1 **P<0.05 ***P<0.01

Interpretation: Having a PhD rather than a Master decreases by 6.6 percentage points the probability of being located in the same region in 1998 and 2001, it being understood that the non-moving Masters amounts to 54.3% in 2001 and 48.2% in 2008.

If mobility is logically higher for the Master's degree holders than for those less qualified, it is also higher for those with PhDs in both the short and long term (model 1a). As regards the effects of age, being older than the average for one's group has no significant effect on mobility in relation to those who are younger. As we saw earlier, age is strongly correlated with level of qualification and has an impact on mobility, but the difference in average age has no effect. However, gender has a significant effect at the 1% level, women being less mobile than men. Having had a child before graduation also significantly reduces the likelihood of being mobile. These results confirm the idea that mobility is dependent on the temporal aspects of trajectory and is explained in part by sociological determinants (Bonney and Love 1991; Shauman and Xie, 1996).

A breakdown of the statistics on PhDs in the 2a model indicates that the HER PhDs, by contrast with the non-HER PhDs, are not significantly less mobile than the Masters. The inclusion of PhDs outside the University system thus reduces the probability of a regional mobility between graduation in 1998 and the workplace in 2001, HER PhDs being more mobile.

In comparing mobility at three years and ten years, one observes very little difference. This is a somewhat surprising result given that mobility takes place mainly at the beginning of professional life. The non-HER PhDs show mobility significantly lower than that of the Masters, over both temporal horizons.

At the end of these treatments of the CEREQ data, the most surprising result is the low mobility of the non-HER PhDs in relation to the HER PhDs. To explain this, we propose several interpretative hypotheses. Positions in Universities are rarer than elsewhere: for an individual who does not wish to pursue a career in higher education and research, it is thus easier to find a job locally. In fact, to be recruited by the University of your PhD thesis defense, a post must become vacant in the year following the thesis. The patterns of non-University recruitment (recruitment in the private sector, entrepreneurship) are less dependent on time constraints: they provide greater flexibility and allow non-HER PhDs to find work in the region with more ease. In addition, local recruitment may simply be prohibited in some Universities, entailing a

change of institution if the PhD wants to continue in the field of higher education and research. Finally, before attaining a permanent academic position (assistant professor notably), more and more PhDs are undertaking transitional post-doctoral positions. These transitions are not necessarily at the local level and thus generate mobility.

4.3 Determinants of interregional mobility of PhDs

In this section, estimates were obtained from a Heckman two-step correction for selection bias (Heckman, 1979). Table 4 shows the results of the selection equation obtained using a *probit* model (step 1) and the output equation of gravity models obtained according to discipline group (step 2). There are 145 pairs of regions for which there is no evidence of interregional mobility in the pure and applied sciences and 199 in the social and human sciences. The significance of the Mills ratio at the 5% level indicates the presence of a selection bias, controlled for in our model. Descriptive statistics on the explanatory variables are given in Appendix D.

Table 4: Determinants of researchers' interregional mobility

Selection equation	Mobility in pure and applied sciences			Mobility in human and social sciences		
	n=506			n=506		
	Coefficient	Std. Dev.		Coefficient	Std. Dev.	
Scientific size of departure region	1.235	0.143	***	1.378	0.143	***
Difference in scientific size	-0.454	0.083	***	-0.497	0.081	***
Spatial distance	-0.443	0.191	**	-0.509	0.177	***
Distance between disciplinary profiles	-0.221	0.259		0.146	0.244	
constant	-0.196	0.476		0.333	0.438	
Output equation	n=361			n=307		
	Coefficient	Std. Dev.		Coefficient	Std. Dev.	
Scientific size of departure region	1.832	0.367	***	2.072	0.583	***
Difference in scientific size	-0.730	0.162	***	-0.746	0.256	***
Spatial distance	-0.710	0.206	***	-0.947	0.304	***
Distance between disciplinary profiles	-0.115	0.286		0.138	0.435	
Ile-de-France	-0.297	0.446		0.060	0.635	
constant	-0.311	0.681		-0.726	0.975	
Error term	Coefficient	Std. Dev.		Coefficient	Std. Dev.	
Mills ratio	1.485	0.643	**	2.094	0.822	**

*P<0.1 **P<0.05 ***P<0.01

Several results are worth noting. First, with regard to the selection equation, that is to say estimating the probability of observing mobility between two regions, we note the existence of i) an size effect of regions on the one hand, and ii) a spatial distance effect on the other. The scientific size of the departure region has a positive effect on the estimated probability. The more PhDs a region produces, the more it is anticipated that some of them will migrate to other regions. The size difference between the regions of departure and arrival is also significant at the 1% level. The probability of observing mobility between two regions is much higher when the

departure region is smaller than the region of arrival. Two regions being geographically close also increases the likelihood of PhDs migrating from one area to another.

Regarding the output equation, which estimates the number of movements between two regions, we find a significant influence at the 1% level of the scientific size of the departure region. The size difference between regions of departure and arrival is still significant at the 1% level. There is even more mobility between the two regions when the departure region is smaller than the region of arrival. Distance plays a significant role in the observed number of movements, which tends to be much higher when the regions are close. This result suggests that spatial proximity is an important determinant in the choice of mobility of researchers, the individuals moving to regions close to the one where they defended their PhD thesis. The present work shows the role of spatial proximity as a determinant of mobility behavior of researchers. It is to be noted that the distance between disciplinary profiles still does not play a role. The dummy variable "Ile-de-France" has no significant effect on the number of movements observed between two regions.

There is very little difference between the results yielded by the two models. It seems that the elements for explaining mobility do not differ from one group of disciplines to another. However, we can note that if the explanatory variables have an impact on mobility in the same direction in both groups of disciplines, the coefficients will be higher for the model relating to the human and social sciences. The tested variables thus show a stronger impact in these disciplines.

5 Conclusion/Discussion

This paper highlights the low mobility of French PhDs. Identifying mobility from two events in their career paths, such as the defense of their own thesis and that of their first PhD student, we show that only 36% of people have changed region. The results differ somewhat between regions. In addition, it is clear that low mobility is not specific to the PhDs, and even less to the higher education and research system. PhDs continuing their careers at University have a degree of mobility higher than other PhDs.

To understand these patterns of mobility, we tested the impact of several factors. Sociological determinants have a strong influence on the mobility of individuals: the effect of age and the temporality of personal career paths directly affect the probability of observing mobility, likewise with family constraints. A second set of determinants, of a structural nature, explain interregional flows. In the first place, the scientific size of a region significantly increases the probability of observing mobility. Spatial proximity has a strong impact on the mobility. PhDs seem to integrate distance into their judgments, being likely to move to a region close to the region where their own PhD thesis was defended. This point joins forces with the sociological dimension in the effort to explain the system of mobility. Not only, at a certain age, does territorial anchorage seem to limit mobility behavior, but individuals, sometimes forced to migrate for a job, try to minimize the distance that separates them from their home environment.

The findings of this paper feed into two very current debates. The first relates to the work of Richard Florida on the "creative class". The latter, which includes the scientists, is presented as a strategic value for territories, these having an interest in attracting such class. For this, they need to develop "soft factors" to which creative workers are particularly sensitive. Spatial mobility of

the creative class becomes a critical parameter, since the success of local economic development strategies of attractiveness depends on the degree and determinants of their mobility. Finally, the mobility of researchers has been seen to be rather low and, when observed, it is explained in part by structural and individual determinants on which local policies have little influence. This supports the results of Martin-Brelot et al. (2010). In the light of these results, the strategy relating to the attractiveness of the creative class can without doubt be re-examined.

The second debate relates to the issue of inbreeding. This is particularly significant in French Universities (Godechot and Louvet, 2008; Bouba-Olga et al., 2008), and there are many who view the recruiting system as the main reason for the low mobility of researchers, disturbing the natural matching of the academic job market. Regardless of this issue, we show that there are individual and structural factors that explain people's choices in relation to mobility, without forgetting that this low mobility is not specific to the higher education and research system.

Finally, the issue of inbreeding raises the question of the spatial diffusion of knowledge, because those who criticize local recruitment highlight the risk of scientific sclerosis. To conclude that cognitive inertia has its basis in the low spatial mobility of researchers, however, is too extreme: the geographic mobility of a researcher may be temporary, through enrollment in research networks, participation in conferences, involvement in collaborative research programs, visit in other Universities, etc. As an extension of the present work, it would be interesting to investigate the temporary mobility of researchers (Rallet and Torre, 2005; Canibano et al., 2011; Barrufaldi and Landoni, 2012) - though such a study would not be straightforward - in order to analyze its impact on the collaborative process.

References

- Agrawal, A., Cockburn, I., Michale, J., 2006. Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships, *Journal of Economic Geography*, 6(5), 571-591.
- Allison, P.D., Long, J.S., 1987. Interuniversity mobility of academic scientists, *American Sociological Review*, 52(5), 643-652.
- Almeida, P., Kogut, B., 1999. Localization of knowledge and the mobility of engineers in regional networks, *Management Science*, 45(7), 905-917.
- Barrufaldi, S.H., Landoni, P., 2012. Return mobility and scientific productivity of researchers working abroad: the role of home country linkages, *Research Policy*, 41, 1655-1665.
- Bonnal, L., Giret, J.F., 2010. Determinants of access to academic careers in France, *Economics of Innovation and New Technology*, 19(5), 437-458.
- Bonney, N., Love, J., 1991. Gender and migration: geographical mobility and the wife's sacrifice, *The Sociological Review*, 39(2), 335-348.
- Boschma, R.A., 2005. Proximity and Innovation: a critical assessment, *Regional Studies*, 39(1), 61-74.
- Bouba-Olga, O., Grossetti, M., Lavigne, A., 2008. Le localisme dans le monde académique : une autre approche, *La Vie des Idées*, 12 mai.
- Bouba-Olga, O., Ferru, M., Pepin, D., 2012. Exploring spatial features of science-industry partnerships: a study on French data, *Papers in Regional Science*, 91(2), 355-375.
- Bozeman, B., Mangematin, V., 2004. Editor's introduction: building and deploying scientific and technical human capital, *Research Policy*, 33(4), 565-568.
- Breschi, S., Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows, *Journal of Economic Geography*, 9, 439-468.
- Brouillat, E., Lung, Y., 2010. Spatial distribution of innovative activities and economic performances: a geographical-friendly model, *Druid Summer conference*, London, 16-18 June.
- Canibano, C., Otamendi, F.J., Solis, F., 2011. International temporary mobility of researchers: a cross-discipline study, *Scientometrics*, 89, 653-675.
- CEC, 2001. A mobility strategy for the European research area, *Communication from the Commission to the Council and the European Parliament*.
- CEC, 2008. Realising a single labour market for researchers, *Report of the ERA expert group*.
- Combes, P.P., Linnemer, L., Visser, M., 2008. Publish or peer-rich? The role of skills and networks in hiring economics professors, *Labour Economics*, 15, 423-441.
- Criscuolo, P., 2005. On the road again: researcher mobility inside the R&D network, *Research Policy*, 34, 1350-1365.

- Cruz-Castro, L., Sanz-Menendez, L., 2010. Mobility versus job stability, assessing tenure and productivity outcomes, *Research Policy*, 39, 27-38.
- Debackere, K., Rappa, M.A., 1995. Scientists at major and minor universities: mobility along the prestige continuum, *Research Policy*, 24, 137-150.
- De Filippo, D., Sanz-Casado, E., Gomez, I., 2009. Quantitative and qualitative approaches to the study of mobility and scientific performance: a case study of a Spanish university, *Research Evaluation*, 18(3), 191-200.
- Dietz, J.S., Chompalov, I., Bozeman, B., Lane, E.O., Park, J., 2000. Using the curriculum vitae to study the career paths of scientists and engineers: an exploratory assessment, *Scientometrics*, 49(3), 419-442.
- Edler, J., Fier, H., Grimpe, C., 2011. International scientist mobility and the locus of knowledge and technology transfer, *Research Policy*, 40, 791-805.
- Florida, R., 2002. *The Rise of the Creative Class, and How It's Transforming Work, Leisure, Community And Everyday Life*, New York, Basic.
- Franzoni, C., Scellato, G., Stephan, P., 2012. Foreign born scientists: mobility patterns for sixteen countries, NBER Working Paper, No.18067.
- Frenken, K., Hardeman, S., Hoekman, J., 2009. Spatial scientometrics: towards a cumulative research program, *Journal of Informetrics*, 3(3), 222-232.
- Gaughan, M., Robin, S., 2004. National science training policy and early scientific careers in France and the United States, *Research Policy*, 33, 569-581.
- Génération 1998-2008, 2008. CEREQ, Centre Maurice Halbwachs (CMH).
- Godechot, O., Louvet, A., 2008. Le localisme dans le monde académique : un essai d'évaluation, *La Vie des Idées*, 22 avril.
- Greenwood, M.J., 1975. Research on internal migration in the United States: survey, *Journal of Economic Literature*, 13(2), 397-433.
- Griliches, Z., 1990. Patent statistics as economic indicators: a survey, *Journal of Economic Literature*, 28, 1661-1707.
- Grossetti, M., Milard, B., 2011. La concentration spatiale des activités scientifiques en question, *Communication pour le troisième colloque international du RESUP, « L'enseignement supérieur et la recherche en réformes »*, Paris, 27, 28 et 29 janvier 2011.
- Grossetti, M., 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 Rurale et Urbaine*, 101, 23-31.
- Heckman, J., 1979. Sample selection bias as a specification error, *Econometrica*, 47, 153-161.
- 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.

Hoisl, K., 2007. Tracing mobile inventors, the causality between inventor mobility and inventor productivity, *Research Policy*, 36, 619-636.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations, *The Quarterly Journal of Economics*, 108(3), 577-598.

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

Latham, W., Le Bas, C., Bouklia-Hassane, J.R., Volodin, D., 2011. Interregional mobility, productivity and the value of patents for prolific inventors in France, Germany and the UK, *International Journal of Foresight and Innovation Policy*, 7, 92-113.

Maggioni, M., Uberti, T., 2007. Inter-regional knowledge flows in Europe: an econometric analysis, in Frenken K. (eds) *Applied evolutionary, economics and economic geography*, Edward Elgar, Cheltenham, UK.

Martin-Brelot, H., Grossetti, M., Eckert, D., Gritsai, O., Kovacs, Z., 2010. The Spatial Mobility of the 'Creative Class': A European Perspective, *International Journal of Urban and Regional Research*, 34(4), 854-870.

Musselin, C., 2009. *The Markets for Academics*, translation of « Le marché des universitaires » published in Presses de Sciences Po in 2005, New York, Routledge.

Pezzoni, M., Sterzi, V., Lissoni, F., 2012. Career progress in centralized academics systems: social capital and institutions in France and Italy, *Research Policy*, 41, 704-719.

Ponds, R., Van Oort, F., Frenken, K., 2007. The geographical and institutional proximity of research collaboration, *Regional Science*, 86(3), 423-443.

Rallet, A., Torre, A., 2010. Proximity and localization, *Regional studies*, 39, 47-59.

Schwartz, A., 1976. Migration, Age and Education, *Journal of Political Economy*, 84(4), 701-719.

Shearmur, R., 2007. The new knowledge aristocracy: a few thoughts on the creative class, mobility and urban growth, *Work Labour and Globalization*, 1(1), 31-47.

Shauman, K.A., Xie, Y., 1996. Geographic mobility of scientists: sex differences and family constraints, *Demography*, 33(4), 455-468.

Tripl, M., 2011. Scientific Mobility and Knowledge Transfer at the Interregional and Intraregional Level, *Regional Studies*, 1-15.

Yankow, J., 2003. Migration, job change, and wage growth: a new perspective on the pecuniary return to geographic mobility, *Journal of Regional Science*, 43(3), 486-516.

Zellner, C., 2003. The economic effects of basic research: evidence for embodied knowledge transfer via scientists' migration, *Research Policy*, 32, 1881-1895.

Zucker, L.G., Darby, M.R., 2008. Star scientists, innovation and regional and national immigration, NBER Working Paper, No.13547.

Appendix A - Building the database DOCTHESE

The DOCTHESE database relates to approximately 218,000 PhD theses completed in France between 1970 and 2000. For each of the rows that match a thesis, we have the full name of the author of the thesis and that of the thesis supervisor. From this information, it is possible to identify individuals who complete a PhD thesis and are later found acting as thesis supervisors. The tables below illustrate the method of identifying transitions and transforming the database:

Table A1: Presentation of the original database (217,994 lines)

PhD	Thesis supervisor	Year of thesis defense	University of thesis defense	Discipline of the thesis defended
Individual A	Individual B	1979	Lyon	Economics
Individual C	Individual A	1990	Poitiers	Economics

Table A2: Presentation of the database after identifying transitions (12,261 lines)

Transition identified	Supervisor of thesis	Year of thesis defense	University of thesis defense	Discipline of the thesis defended	First PhD supervised	Year of thesis supervision	University of thesis supervision	Discipline of the thesis supervised
Individual A	Individual B	1979	Lyon	Economie	Individual C	1990	Poitiers	Economie

In the case of similar surnames, we performed the matching manually: the main problem with the data was the incomplete filling in of the the surname of the author (sometimes only an initial was given). When an individual appeared several times as an author, that is to say he/she has completed several theses, we used only the first thesis. Similarly, when an individual appeared several times as a supervisor, we refer only to the first thesis he/she directed. One way of controlling the matching process was to verify that there is at least three years between defense and first supervision, so as to avoid erroneous cases. Taking the necessary precautions, we finally obtained a database of 12,261 individuals, for which we have information on each person's own thesis and on his/her first PhD thesis supervision. This number may seem small compared to the initial number of lines of data, but several factors can explain this difference:

- Loss of matches at the beginning and end of the database: logically, it is unlikely that we would find in the database the thesis of a supervisor if he/she completed it in the early 1970s. Symmetrically, a PhD who completed his own thesis at the end of the 1990s will probably not have had the opportunity to supervise a PhD thesis before the year 2000.
- Misspellings, incorrect surnames and first names: this inevitably generated missed pairings.
- The more theses an individual supervises, the more it "consumes" lines of data. For example, when a person appears ten times as a supervisor, one goes from eleven lines in the original database to a single line after the matching process. The line on which the individual appears as author and the ten lines where he/she appears as supervisor all become reduced to a single line.
- Finally, there are PhDs who never become supervisors.

Appendix B – The scientific size of regions

The table below shows the regional distribution of PhDs from 217,994 recorded theses (1st column); likewise of the 12,261 individuals from the point of view of thesis defense (2nd column) and the same 12,261 individuals from the point of view of thesis supervision (3rd column). This distribution shows the hypertrophy of Ile-de-France, which “produces” 43.3% of the PhDs, a result closely matching empirical work on the geography of science. When comparing the distribution of PhDs by region from the point of view of thesis defense and thesis supervision, the importance of the Paris region goes from almost half (46.3%) to a third (33.3%). The importance of the other regions correspondingly increases. This trend seems to illustrate a process of spatial diffusion of research from center to periphery, already identified by other authors (Grossetti and Milard, 2011).

Table B1: Regional distribution of data

Region	Original database	Region of thesis defense	Region of thesis supervision
Academy of Grenoble	5.3%	5.4%	6.6%
Academy of Lyon	5.4%	5.5%	6.4%
Academy of Marseille	4.6%	4.7%	5.6%
Academy of Nice	1.9%	1.9%	2.3%
Alsace	4.0%	3.3%	3.6%
Aquitaine	3.8%	3.8%	3.9%
Auvergne	1.2%	1.0%	1.4%
Bourgogne	1.3%	1.1%	1.6%
Bretagne	3.0%	2.9%	3.8%
Centre	1.3%	1.1%	2.1%
Champagne-Ardenne	0.4%	0.4%	0.7%
Franche-Comté	0.9%	0.9%	1.2%
Ile-de-France	43.3%	46.3%	33.3%
Languedoc-Roussillon	4.5%	3.5%	4.4%
Limousin	0.5%	0.5%	0.7%
Lorraine	3.3%	3.0%	3.5%
Midi-Pyrénées	6.3%	6.3%	6.5%
Nord-Pas-De-Calais	3.0%	3.3%	3.9%
Basse-Normandie	1.0%	0.8%	1.3%
Haute-Normandie	0.9%	0.7%	1.3%
Pays-de-la-Loire	1.6%	1.2%	2.6%
Picardie	0.9%	0.7%	1.3%
Poitou-Charentes	1.6%	1.6%	1.9%
France	100.0%	100.0%	100.0%
Nb. Obs. :	217 994	12 261	12 261

Appendix C – Disciplines

The table below shows the distribution of PhDs by discipline, from which we observe strong differences in mobility. Overall, PhDs in the human and social sciences change region more between defense and first supervision than PhDs in the pure and applied sciences. The number of observations may differ from one column to another, discipline not having been entered for some individuals.

Table C1: Distribution of disciplines of data

Discipline	Original database	Region of thesis defense	Region of thesis supervision	Regional rate of mobility
<i>Human and social sciences</i>	41.7%	40.9%	35.4%	47.4%
Law, Political sciences	6.7%	4.5%	4.1%	42.9%
Literature	3.5%	3.8%	3.2%	62.1%
Languages	7.0%	8.9%	7.7%	52.3%
Management, Communication	2.1%	1.6%	2.2%	40.5%
History	5.1%	5.2%	4.5%	47.4%
Geography	2.6%	2.8%	2.3%	51.8%
Philosophy	2.0%	2.3%	1.4%	51.5%
Psychology	2.5%	3.2%	2.7%	37.0%
Sociology, Educational sciences	3.7%	3.0%	2.7%	44.7%
Economics	4.5%	4.0%	2.9%	34.4%
Arts	2.0%	1.6%	1.7%	47.5%
<i>Pure and applied sciences</i>	58.2%	59.1%	64.6%	29.3%
Biology	14.4%	14.9%	17.4%	29.2%
Chemistry	5.7%	4.4%	4.3%	32.0%
Mathematics	2.8%	3.9%	3.7%	42.8%
Medical and pharmaceutical sciences	3.4%	3.1%	3.7%	24.4%
Physics	10.7%	12.8%	12.8%	25.7%
Engineering	12.7%	10.4%	13.1%	26.4%
Information technology	4.2%	5.2%	5.3%	28.1%
Earth sciences	4.3%	4.5%	4.3%	36.5%
Total	100.0%	100.0%	100.0%	36.0%
Nb. Obs. :	208 623	11 248	11 885	11 248

Appendix D – Descriptive statistics

Table D1: Explanatory variables of the *probit* model (CEREQ data)

	Obs.	Mean	Std. Dev.	Min.	Max.
≤ Baccaureate	9121	0.49	0.50	0	1
Baccaureate < level < Master	9121	0.41	0.49	0	1
Master	9121	0.07	0.25	0	1
PhD	9121	0.03	0.17	0	1
HER PhD	9121	0.01	0.09	0	1
Non-HER PhD	9121	0.02	0.14	0	1
Age	9121	0.38	0.49	0	1
Man	9121	0.52	0.50	0	1
Child	9121	0.04	0.18	0	1

Table D2: Explanatory variables of the gravity model (DOCTHESI data)

	Obs.	Mean	Std. Dev.	Min.	Max.
Scientific size of departure region	506	0.80	1.00	-0.93	3.77
Difference in scientific size	506	0.00	1.45	-4.70	4.70
Spatial distance	506	1.55	0.43	0.00	2.25
Distance between disciplinary profiles	506	-1.40	0.29	-2.25	-0.79
Ile-de-France	506	0.09	0.28	0	1