



**HAL**  
open science

## Interregional flows 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. Interregional flows of PhDs: an analysis of French long-term data. 2014. halshs-01114268

**HAL Id: halshs-01114268**

**<https://shs.hal.science/halshs-01114268>**

Preprint submitted on 9 Feb 2015

**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.

# Interregional flows of PhDs: an analysis of French long-term data

**Bastien Bernela<sup>1</sup>, Olivier Bouba-Olga, Marie Ferru**

[bastien.bernela@univ-poitiers.fr](mailto:bastien.bernela@univ-poitiers.fr) ; [obouba@univ-poitiers.fr](mailto:obouba@univ-poitiers.fr) ; [marie.ferru@univ-poitiers.fr](mailto:marie.ferru@univ-poitiers.fr)

CRIEF EA2249, University of Poitiers

2 rue Jean Carbonnier, Bât A1, TSA 81100, 86073 Poitiers Cedex 9

**Abstract:** The aim of this article is to provide empirical analysis of the geographical mobility of PhDs between French regions by using long term data (1970-2000). 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. We then 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:** Interregional flows, Geographical mobility, PhDs, Size effect, Distance.

## 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, strongly contributes to such reflections (CEC, 2001, 2008). Many investigations make mobility of researchers a challenge for 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; Agrawal et al., 2006; Zucker and Darby, 2008; Breschi and Lissoni, 2009; Trippl, 2011).

In line with this challenge, this article focuses on interregional flows of PhDs within the French national context. It contributes to debates about knowledge circulation, by giving empirical evidence on PhDs mobility between French regions. We aim at better understanding interregional flows by describing them in time and in space, and by testing the impact of some structural determinants on the observed flows. Jayet (1996) proposes to distinguish four ways to study economically migration by crossing two binary variables: the geographical scale (national versus international) and the analysis scale (microeconomic/choices versus macroeconomic/flows). This work consists in a macro-analysis of French interregional flows<sup>2</sup>, and not in a micro-analysis of mobility individual choice.

---

<sup>1</sup> Corresponding author

<sup>2</sup> The article focuses on French case, studying flows within national territory. We do not take into account international flows, whether from France to abroad or from abroad to France. Our database does not allow such treatments. Additionally, some scholars have demonstrated the low

*\*Structural determinants of interregional knowledge flows: the underlying debates*

Many scholars have tested the impact of individual determinants on researchers' mobility, such as age, sex, academic position, publication scores, etc. However, the influence of structural factors appears to be too rarely measured, while it is extensively tested (and appears significant) in the case of research collaborations. When investigating interregional knowledge flows (as capital and labor mobility), some key determinants have to be questioned, as the role of regions' size and geographical distance between regions.

Indeed, regarding size effects, in line with the new economic geography (Krugman, 1990; Fujita et al., 2001), Grossetti and Nguyen (2001), Hoekman et al. (2010) and Bouba-Olga et al. (2012) observe empirically the tendency to knowledge concentration in space. Concerning geographical distance, abundant literature has participated in the debate about the death of distance. 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). Several measures of geographical distance are proposed as border effects and distance-time travel.

In this article, we propose to use mobility data as an indicator for knowledge diffusion across regions, using original data on French PhDs. In a continuation of previously cited scholars, the present investigation seeks to empirically 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 and arrivals of researchers to be expected.

*\*Mobility data*

The study of mobility is often made difficult by the lack of available data, namely long-term data on the trajectory and location of researchers. The review of empirical works on the mobility of researchers allows the identification of two major methods of investigating mobility.

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 latter aim to identify inventors/authors' affiliations to the different dates of patents/publications, and to deduce whether or not there is mobility by comparing the locations that correspond to these events. This method has the advantage of providing a lot of peripheral information on mobility and allows drawing simultaneously from a single data source their spatial

---

international mobility of French researchers (Guellec and Cervantes, 2002; Harfi and Mathieu, 2006).

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<sup>3</sup>.

Aware of the limitations related to this first type of data, a second group of investigations has been 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 their *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). 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 requires a relatively large investment in time (particularly in the case of semi-structured interviews), which can restrict the size of the sample.

#### *\*Empirical strategy*

Regardless of the method chosen to identify and measure mobility, we address two limits to existing investigations that we propose to go beyond in the present research. Firstly, if mobility is to be necessarily studied on a dynamic basis (locations at time  $t$  and time  $t+n$ ), the investigations are often undertaken on short-term data, risking that results are too dependent on contextual factors. Then, existing work focuses more on individual characteristics to explain mobility and tends to obscure the structural determinants (as geographical distance and size effects), as we refer earlier.

Based on these observed limitations, we mobilize Docthesse, an original long-term (1970-2000) database relating to more than 12,000 PhDs in France. Knowing the location of the thesis defense and the first thesis supervision of a PhD, we can focus on these particular stages of researchers' career paths and investigate the geographical dimension, in line with the scholars on the tracking of career trajectories. Are PhDs mobile? In what proportion does a PhD supervise his/her first thesis in the same place he/she defended his/her own thesis? If he/she goes to another University, where does it tend to be situated? Do PhDs tend to concentrate in large-size regions? What are the determinants of interregional flows?

---

<sup>3</sup> Several limits can be identified: the tendency to focus on prolific scientists introducing a selection bias, the non-inclusion of "non-production" periods in the career paths of individuals, the issue of multi institutional affiliation of many researchers (Katz and Martin, 1997), the lack of representativeness of areas of science (data on patents tends to exclude fields where scientific production does not consist in filing patents, like the human and social sciences, Griliches, 1990).

We propose to complement existing empirical work by i) exploring mobility of French PhDs across regions, and ii) integrating in an econometric model (Heckman’s model with a selection equation), derived from international trade, various structural indicators largely untested to better understand the observed flows.

The article is organized as follows. In section 2, we describe database and introduce some descriptive statistics about interregional flows of PhDs within French territory. In section 3, we present the gravity model and the main results. The article concludes by reflecting on the implications of our findings for research policy.

## 2. Presentation of Docthesse database

### 2.1. From administrative data to interregional mobility data

The lack of career data generates difficulties for investigations into the spatial trajectory of researchers. In this article, we propose to mobilize an original database, Docthesse, 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, it is possible to identify individuals who completed a PhD thesis and are later found acting as thesis supervisors. This is the case for 12 261 individuals. For each, we could then match data relating to the defense of their thesis with data on the first thesis they supervised<sup>4</sup>. Tables 1 and 2 illustrate the method of identifying transitions and transforming the database:

**Table 1. Presentation of the original database (217,994 lines)**

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

**Table 2. 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
Indiv. A	Indiv. B	1979	Lyon	Economics	Indiv. C	1990	Poitiers	Economics

The number of identified transitions (12,261) may seem small compared to the initial number of theses (217,994), 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

---

<sup>4</sup> 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. The average time-lapse between PhD defense and first supervision is eleven years.

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.

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. 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.

- that concerning the establishment: does the individual defend 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: does the individual defend 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 spatial proximity and size effects. 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.

To investigate the spatial dimension of the University system, a possible scale is that relating to "academies". The division of French territory into academies corresponds to the NUTS2 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 article, the term "region" refers to the nomenclature of administrative NUTS2

regions, with the exception of Rhône-Alpes and Provence-Alpes-Côte-d'Azur which we break down into academies.

## 2.2. Descriptive statistics

We propose to analyze mobility occurring between French regions, and introduce first insights on the level and deployment of mobility of PhDs on French territory. To do this, we construct an interregional flow matrix in order to characterize patterns of mobility of French PhDs (Table 3). We look at, for each French region, 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? To our knowledge, such work has never been undertaken.

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)<sup>5</sup>. This result highlights the existence of a strong circulation of PhDs within the capital region. Even if the Parisian PhDs rarely supervise their first 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 significant 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.

Finally, we propose to explore the evolution of mobility during the observed period: Figure 1 shows mobility rate between 1972 and 1992<sup>6</sup> and distinguishes between human

---

<sup>5</sup> Godechot and Louvet (2008) have used Docthesse for measuring the importance of local recruitment 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 local recruitment. Changing spatial scale of analysis (from University to region) strongly impacts results and their interpretation.

<sup>6</sup> In this figure, we only keep years for which we have at least 200 PhDs, excluding thus years before 1972 and after 1992.

and social sciences and pure and applied sciences. We observe a decrease of mobility between 1972 and 1978, and then a relative stability: mobility rate varies between 30% and 40% since 1978. Differences of mobility between disciplines are significant (cf. Table 4). PhDs in the human and social sciences change region more between defense and first supervision (in 47.4% of cases) than PhDs in the pure and applied sciences (29.3%). PhDs in literature are on average 2.5 times more mobile than PhDs in medical and pharmaceutical sciences (mobility rate of 62.1% against 24.4%). This heterogeneity related to disciplines has to be taken into account for econometric estimations.



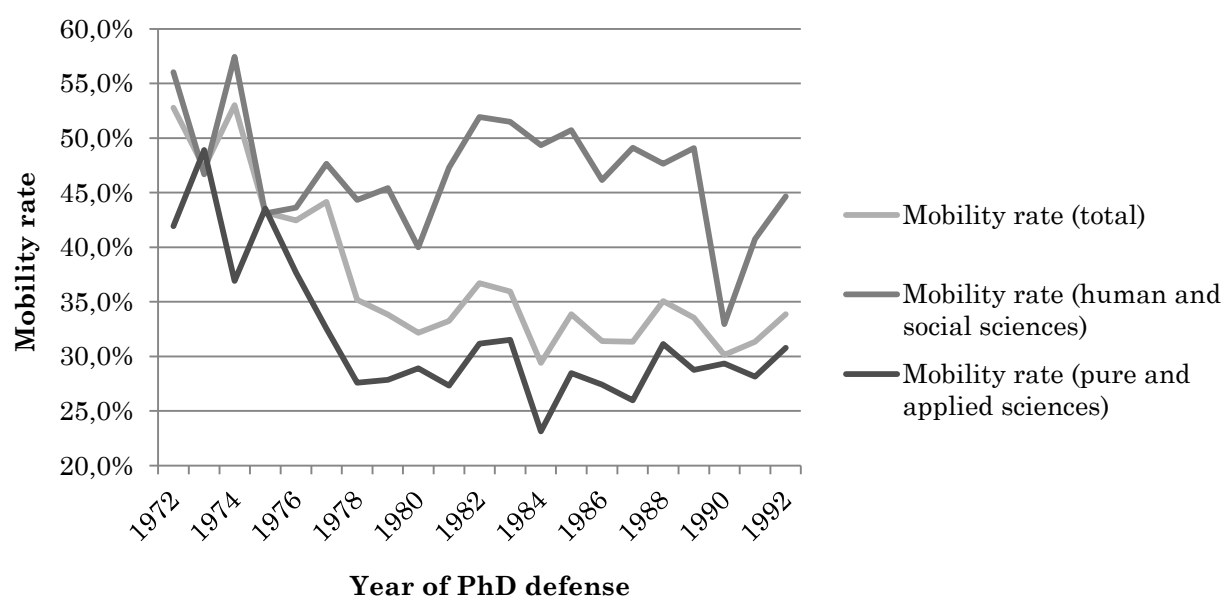
**Table 3. From the region of PhD thesis defense to the region of first supervision of a PhD thesis**

	From the region of PhD thesis defense to the region of first supervision of a PhD thesis							Regional distribution of PhDs	
	Same region	<i>Same University</i>	<i>Other University of the region</i>	Ile-de-France	Neighboring regions (apart IDF)	Other regions	Total	Region of thesis defense	Region of thesis supervision
Academy of Grenoble	75.6%	53.9%	21.7%	7.7%	4.4%	12.3%	100.0%	5.4%	6.6%
Academy of Lyon	70.3%	50.7%	19.6%	8.5%	8.9%	12.3%	100.0%	5.5%	6.4%
Academy of Marseille	68.5%	53.4%	15.1%	6.8%	10.3%	14.4%	100.0%	4.7%	5.6%
Academy of Nice	64.8%	57.1%	7.7%	10.0%	7.8%	17.4%	100.0%	1.9%	2.3%
Alsace	63.9%	60.0%	3.9%	9.7%	6.5%	19.9%	100.0%	3.3%	3.6%
Aquitaine	60.3%	53.6%	6.7%	8.0%	6.9%	24.8%	100.0%	3.8%	3.9%
Auvergne	57.0%	57.0%	-	10.3%	23.0%	9.7%	100.0%	1.0%	1.4%
Bourgogne*	65.0%	65.0%	-	7.9%	11.5%	15.6%	100.0%	1.1%	1.6%
Bretagne	64.3%	58.0%	6.3%	12.3%	6.6%	16.8%	100.0%	2.9%	3.8%
Centre*	48.5%	45.5%	3.0%	17.9%	10.4%	23.2%	100.0%	1.1%	2.1%
Champagne-Ardenne*	43.5%	43.5%	-	22.2%	13.3%	21.0%	100.0%	0.4%	0.7%
Franche-Comté	50.9%	50.9%	-	10.2%	24.1%	14.8%	100.0%	0.9%	1.2%
Ile-de-France	61.2%	29.6%	31.6%	-	5.9%	32.9%	100.0%	46.3%	33.3%
Languedoc-Roussillon	60.3%	54.0%	6.3%	10.5%	16.0%	13.3%	100.0%	3.5%	4.4%
Limousin	67.9%	67.9%	-	1.8%	12.5%	17.8%	100.0%	0.5%	0.7%
Lorraine	62.5%	44.4%	18.1%	8.2%	6.0%	23.2%	100.0%	3.0%	3.5%
Midi-Pyrénées	70.6%	59.9%	10.7%	8.5%	6.8%	14.0%	100.0%	6.3%	6.5%
Nord-Pas-De-Calais	73.7%	65.5%	8.2%	6.7%	1.7%	17.9%	100.0%	3.3%	3.9%
Basse-Normandie	52.9%	52.9%	-	8.8%	21.6%	16.7%	100.0%	0.8%	1.3%
Haute-Normandie*	74.2%	69.7%	4.5%	7.9%	0.0%	17.9%	100.0%	0.7%	1.3%
Pays-De-Loire	67.8%	67.1%	0.7%	9.2%	7.2%	15.8%	100.0%	1.2%	2.6%
Picardie*	68.9%	66.7%	2.2%	15.6%	2.2%	13.3%	100.0%	0.7%	1.3%
Poitou-Charentes	65.8%	65.8%	-	7.7%	10.7%	15.8%	100.0%	1.6%	1.9%
<b>France</b>	<b>64.0%</b>	<b>43.9%</b>	<b>20.1%</b>	<b>8.9%</b>	<b>7.8%</b>	<b>19.3%</b>	<b>100.00%</b>	<b>100.0%</b>	<b>100.0%</b>

\*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.

**Figure 1. Evolution of mobility rate**



**Table 4. Distribution and mobility rate by discipline**

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

### 3. Testing the impact of structural determinants on PhDs' mobility

#### 3.1. Presentation of the model

We mobilize Docthesse data to see if interregional flows of PhDs respond 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  $r1$  corresponds to the regions of PhD defense and  $r2$  to the regions of PhD supervision. The coupling  $r1/r2$  thus does not correspond to the coupling  $r2/r1$ : 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  $r1$  and  $r2$  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:

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

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  $\mu_{i1}$  and  $\mu_{i2}$  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 Docthesse 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 Docthes database. It is, in other words, the contribution of each region to the national production of PhDs. 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 regions<sup>7</sup> to determine the possible tendency of PhDs to go from small regions to large regions, or vice versa.

The two last columns of table 3 indicate the regional distribution of the 12,261 PhDs from the point of view of thesis defense and thesis supervision. This distribution shows the hypertrophy of Ile-de-France, which “produces” near half 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).

#### *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 disciplines whose distribution is presented in table 4. First, we calculate for each region  $i$ , the part  $p$  that represents each of the disciplines  $j$  ( $p_{i,j}$ )<sup>8</sup>. The distance indicator relating the disciplinary profiles of regions 1 and 2 is calculated as follows:

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

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

---

<sup>7</sup> 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.

<sup>8</sup> For example, the share of PhDs in economics (discipline 10) in the Poitou-Charentes region (region 20) is named  $p_{20,10}$ .

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 (cf. table 4). 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.

To provide a dynamic analysis of interregional flows, we could have done the same method with year of PhD defense, i.e. testing separate models for different periods. However, we run a Chow test for two sub-periods (each one weights each the half of total PhDs, i.e. before 1982 versus after 1982) and there is no significant difference: it does not seem necessary to test two independent models, as suggested by the relative stability observed in figure 1. Before moving to the results of the model, some descriptive statistics on the explanatory variables are given in Table 5.

**Table 5. Explanatory variables of the gravity model**

	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

### 3.2. Determinants of interregional mobility of PhDs

In this section, estimates were obtained from a Heckman two-step correction for selection bias (Heckman, 1979). Table 6 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.

**Table 6. Determinants of researchers' interregional mobility**

	Mobility in pure and applied sciences			Mobility in human and social sciences		
<b>Selection equation</b>	<b>n=506</b>			<b>n=506</b>		
	<b>Coeff.</b>	<b>Std. Dev.</b>		<b>Coeff.</b>	<b>Std. Dev.</b>	
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	
<b>Output equation</b>	<b>n=361</b>			<b>n=307</b>		
	<b>Coeff.</b>	<b>Std. Dev.</b>		<b>Coeff.</b>	<b>Std. Dev.</b>	
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	
<b>Error term</b>	<b>Coeff.</b>	<b>Std. Dev.</b>		<b>Coeff.</b>	<b>Std. Dev.</b>	
Mills ratio	1.485	0.643	**	2.094	0.822	**

\*P&lt;0.1 \*\*P&lt;0.05 \*\*\*P&lt;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 a size effect of regions on the one hand, and 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 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.

Regarding 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.

#### **4. Conclusion**

This article 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 differing somewhat between regions. To understand these patterns of mobility, we tested the impact of structural determinants to 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 consider distance issues, being likely to move to a region close to the region where their own PhD thesis was defended.

The findings of this article feed into a very current debate, related to scholars about 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 according to Florida (2002). 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 determinants – that can be associated with "hard factors" (Florida, 2002) - on which local policies have little influence. In the light of these results, the strategy relating to the attractiveness of the creative class can probably be re-examined, as Martin-Brelot et al. (2010) have already suggested.

To conclude, we consider that this type of work sheds light on the mobility of PhDs but cannot allow commenting the potential relation between the low spatial mobility of researchers and the risk of cognitive inertia. 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), pp. 571-591.
- Allison, P.D. & Long, J.S. (1987) Interuniversity mobility of academic scientists, *American Sociological Review*, 52(5), pp. 643-652.
- Almeida, P. & Kogut, B. (1999) Localization of knowledge and the mobility of engineers in regional networks, *Management Science*, 45(7), pp. 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, pp. 1655-1665.
- Boschma, R.A. (2005) Proximity and Innovation: a critical assessment, *Regional Studies*, 39(1), pp. 61-74.
- 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), pp. 355-375.
- Breschi, S. & Lissoni, F. (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows, *Journal of Economic Geography*, 9, pp. 439-468.
- Brouillat, E. & Lung, Y. (2010) Spatial distribution of innovative activities and economic performances: a geographical-friendly model, Paper presented at the 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, pp. 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.
- Cruz-Castro, L. & Sanz-Menendez, L. (2010) Mobility versus job stability, assessing tenure and productivity outcomes, *Research Policy*, 39, pp. 27-38.
- Debackere, K. & Rappa, M.A. (1995) Scientists at major and minor universities: mobility along the prestige continuum, *Research Policy*, 24, pp. 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), pp. 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), pp. 419-442.
- Edler, J., Fier, H. & Grimpe, C. (2011) International scientist mobility and the locus of knowledge and technology transfer, *Research Policy*, 40, pp. 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*, 18067.



- Frenken, K., Hardeman, S. & Hoekman, J. (2009) Spatial scientometrics: towards a cumulative research program, *Journal of Informetrics*, 3(3), pp. 222-232.
- Fujita, M., Krugman, P.R. & Venables, A.J. (2001) The spatial economy: Cities, regions, and international trade, *MIT press*.
- Godechot, O. & Louvet, A. (2008) Le localisme dans le monde académique : un essai d'évaluation, *La Vie des Idées*, 22 avril.
- Griliches, Z. (1990) Patent statistics as economic indicators: a survey, *Journal of Economic Literature*, 28, pp. 1661-1707.
- Grossetti, M. & Milard, B. (2011) La concentration spatiale des activités scientifiques en question, Paper presented at the third international RESUP conference, Paris.
- 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*, pp. 101, 23-31.
- Guellec, D. & Cervantes, M. (2002). International mobility of highly skilled workers: From statistical analysis to policy formulation, *OECD - International mobility of the highly skilled*, pp. 71-98.
- Harfi, M. & Mathieu, C. (2006). Mobilité internationale et attractivité des étudiants et des chercheurs, *Horizons stratégiques*, 1(1), pp. 28-42.
- Heckman, J. (1979) Sample selection bias as a specification error, *Econometrica*, 47, pp. 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), pp. 662-673.
- 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), pp. 577-598.
- Jayet H. (1996) L'analyse économique des migrants, une synthèse critique, *Revue Economique*, 47(2), pp. 193-226.
- Katz, J.S. & Martin, B.R. (1997) What is research collaboration? *Research Policy*, 26, pp. 1-18.
- Krugman, P. (1990) Increasing returns and economic geography. *NBER Working Paper*, 3275.
- 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, pp. 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).
- 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), pp. 854-870.
- Ponds, R., Van Oort, F. & Frenken, K. (2007) The geographical and institutional proximity of research collaboration, *Papers in Regional Science*, 86(3), pp. 423-443.

- Rallet, A. & Torre, A. (2010) Proximity and localization, *Regional Studies*, 39, pp. 47-59.
- Schwartz, A. (1976) Migration, Age and Education, *Journal of Political Economy*, 84(4), pp. 701-719.
- Trippl, M. (2011) Scientific Mobility and Knowledge Transfer at the Interregional and Intraregional Level, *Regional Studies*, pp. 1-15.
- Zellner, C. (2003) The economic effects of basic research: evidence for embodied knowledge transfer via scientists' migration, *Research Policy*, 32, pp. 1881-1895.
- Zucker, L.G. & Darby, M.R. (2008) Star scientists, innovation and regional and national immigration, *NBER Working Paper*, 13547.