Paper to be presented at the DRUID Academy 2012

on

January 19-21
at
University of Cambridge /The Moeller Centre

Complementarities in scientific production. Evidence from scientific research.

Christian Martinez Diaz
University of Strasbourg
Faculty of Economics
camartinezdiaz@unistra.fr

Abstract

Complementarities in scientific production.
Evidence from scientific research

Christian MARTINEZ DIAZ
Laboratory: BETA - Université de Strasbourg
Currently in my 4th year of enrolment
Expected final date: November 2012
Email address: camartinezdiaz@unistra.fr

Abstrasct
Evidence from studies of scientific productivity reveals a skewed distribution of individual contributions. Using data from research groups affiliated with the University of Strasbourg, I apply the methodology of supermodularity to investigate complementarities between principal investigators who are highly productive and other members of their research groups. The results reveal interdependences between principal investigators and their collaborators.

Jelcodes:O32,O15

Complementarities in scientific production:

Evidence from collective research.

Abstract

Evidence from studies of scientific productivity reveals a skewed distribution of individual contributions. Using data from research groups affiliated with the University of Strasbourg, I apply the methodology of supermodularity to investigate complementarities between principal investigators who are highly productive and other members of their research groups. The results reveal interdependences between principal investigators and their collaborators.

1. Introduction

The scientific production process is characterized by a skewed distribution of output among individual researchers, and therefore we may say that within a scientific research group, where principal investigators (PI's) produce most of the scientific output, there must be a complementary link between PI's and other types of scientific personnel, such as fellow investigators, assistants, PhD students, and administrative personnel.

Lotka (1926) performed a study on the publication counts of scientists in the fields of physics and chemistry and observed that the distribution of scientific contributions is positively skewed with the number of scientists making *n* contributions equals $1/n^2$ of those making only one, and that 60% of the scientific population only accounts for one. This became known as Lotka's law on scientific distribution explaining the irregularities the scientific productivity; moreover, Bradford (1934) observed citations in refereed journals, and concluded there were "exponentially diminishing returns to extending a search in scientific journals" in order to find relevant scientific contributions to a specific topic in a set of journals confirming the existence irregularities in the scientific productivity distribution, where only a few scientists produce high quantities of contributions.

Today it is commonly acknowledged that the skewness of scientific production is due to differences of scientific fields, age of scientists, cohorts etc, which raises our interests on other issues such as the interactions driving it.

Therefore, the issue I want to address is the skewness of the scientific productivity distribution, which combined with the notion of collective scientific efforts among researchers, allows to believe there is a relationship between sectors of the scientific productivity distribution. In fact, those PI's located at the tail of the curve with several contributions must rely on their assistants, post doctors, and PhD students to attain such high numbers of contributions, therefore, a relationship of complementarities between these categories of scientists must account for "research delegation" or "collective scientific effort" in scientific production.

Therefore I study the notion of complementarities between PI's and other members of their research group, performing an empirical analysis of the scientific production of research groups using data from a set of public research laboratories affiliated to the University of Strasbourg (formerly University Louis Pasteur) with the objective to reveal the effects of the research group composition in terms of human resources on the scientific output, and based on the theory of supermodularity (Topkis, 1998), I propose an alternative approach to verify the existence of complementarities among different types of scientific personnel. The results provide evidence to support the existence of links among different types of scientific personnel. They highlight senior, postdoctoral researchers, and administrative personnel as determinant to the scientific production process, and verify that the estimated scientific production is supermodular, implying the existence of complementarities among different couples of researchers.

2. Literature Review

Over the past century, it had been observed that citations are an important determinant of a scientist's salary. Citations are usually regarded as a non-pecuniary mechanism for recognition, and/or a proxy for the measurement of quality research and productivity, therefore their marginal value should differ depending on the quality of research conducted by the scientist. Diamond (1986) performed an estimation of the earnings function of scientific researchers, where citations are the main determinant of salary. His results showed that scientific departments really take into account the citation variable to determine salaries, and that citations counts are correlated with scientific productivity at the moment of evaluation of the quantity and quality of research performed by a member of the department, therefore they can be associated to an output index. These notions imply a best strategy for a scientist to create a base of citable research along a network of peers that agree on citing each other, however, other variables such as the time the scientist took to complete his PhD, the quality of the institution where he received it, or his experience may also be as important determinants of salaries as his citation counts.

Such findings have commonly served as support for the notion that star scientists with many publications and citations were holding the scientific production process on their shoulders, which is the idea we want to overthrow in our analysis since we assume the scientific production process does not only rely on star scientists, but rather on complex and entangled relationships among researchers of different status.

2.1 Skewness of the distribution of scientific productivity

In order to understand our argument, we must first recall some other wide spread notions relative to the distribution of the scientific production process; Lotka (1926) studied the frequency of publications by authors in different fields. The results form his studies suggest that the number of authors that publish *n* contributions is a fixed ratio of $1/n^a$ which defines the Lotka's law on the skewness of the distribution of scientific productivity that states that as the number of articles increase, the number of authors publishing many contributions decreases.

Several studies have confirmed the skewness of the distribution of scientific and innovative productivity. As an example, the results obtained by a research conducted by Arora and David, 1996, on the determinants of scientific productivity suggest that aggregate publication output of the research unit may vary with the distribution of research grants, and is skewed to the right since the elasticity of quality-adjusted publications with respect to the budget for a large number of research groups is around 0.6 whereas for some individual researchers, this elasticity approaches the unity, and for a very small number of researchers this elasticity is greater than 1, implying that past performance influences future performance, and in addition, the superior performance of the group leader increases the probability of success for their research proposals.

Breschi and Lissoni (2007) stated in their study on the trade-off between patenting and publishing that measuring productivity by means of cumulated publication records might be misleading due to the fact that distribution of professors over the number of publications is skewed to the right. Since scientific productivity follows a non normal distribution they check differences between inventors and controls through yearly publication data, and improve it by weighting the citations in the publications.

Productivity distributions among scientists are highly skewed, which can be associated to a cumulative advantage process; Allison and Stewart (1974) assume that the distribution of scientific productivity becomes increasingly unequal, as a cohort of scientists grows older. Their analysis of the distribution of the scientific productivity focuses on the assessment of whether or not the distribution becomes increasingly dispersed as time passes by.

Through simulated time series data by stratifying according to career and age, so that each stratum may represent a cohort of scientists across time and may be analyzed in terms of life-course differences, they measure productivity in terms of publication counts on 5-year periods and citations on the whole published works. Their results suggest there is indeed a general increase in citation inequality as scientists grow older, and in fact, the fit between scientist's resources and productivity improves over time.

Expanding the notion of skewness of the scientific production we may recall that tools have been developed for the purpose of its evaluation, such as the H-index (Hirsch, 2005), which ranks the publications of an author in decreasing order with respect to the number of citations it has received and positions the author in the rank *h* where for which all publications in the ranks 1 to *h* have at least *h* citations. In recent years the index has gained popularity, although Egghe (2007, 2009) points out a disadvantage in the fact that it only values the first *h* citations of the publications, and formulates an extension, the G-index, which is represents highest the rank *g* for which the articles in the 1 to g ranks have at least *g* at the power of two citations. This extension results in a more discriminative index among scientists. In addition, he provides evidence that the rank-order distribution of journal-impact factors of a set of journals is S-shaped, compared to the rank order distribution of the H-index of the same set of journals which is a decreasing power law.

2.2 Scientific research is a distributed and interactive collective process

Technological paradigms are said to be the result of interactions between scientific advancement, economic factors, institutional variables and unsolved problems for the established paradigms. Dosi (1982) performed an analysis of continuous and discontinuous changes in technological innovation that are respectively related to the progress along a technological trajectory and to the emergence of new paradigms. Among different results of his research we may notice it appears that for current technological paradigms, the interaction between researchers and inventors can be represented as an embedded set of relations and links for which any group performances can hardly be analyzed in an isolated way.

Relying on the literature on innovation economics we know that the innovation process is the result of interactions between firms and organizations that are mutually influenced by the pre-existing relations in the coordinated network. There has been growing interest in joint production processes influenced by increasing collaboration and firm alliances, affecting the nature of the productive and innovative activities, and generating a loop that affects the same distributed process that generates them. Therefore, innovation acts as both an input and an output of in the process of distributed productive and innovative activities.

Coombs and Harvey (2001) studied the fact that innovation processes are often issued form coordinated network relations between firms and organizations. They introduced the concept of instituted economic processes, composed of distributed innovation initiatives that can be instituted and/or de-instituted in space and time, and proposed an analysis to assess how productive and innovative distributed relations between firms are formed, stabilized and disrupted.

However, due to the priority of understanding the relations between organizations, the authors focus only on distributed innovation processes between firms rather than within firms and other organizations, which represents as well a large share of the innovation process.

They perform their study through a conceptualization based on key issues that are observed in the process of innovation driven by the relations between organizations. The concepts hovers around the notions of mode (how different organizations create and coordinate resources and competition), dynamics (changes in the mode), and scale (to which extent the agent's inter-relationships are transformed following the stimulus to innovate.

Based upon the development of a notion of distributed innovation processes, connecting several actors in a collective effort we focus our study on an attempt to analyze such group performances not at the technological innovation process, but rather at the scientific production one through the assessment of complementarities between certain types of researchers performing an integrated and collective work. Although it may be hard to disentangle isolated performances within a scientific production process due to the increasing degree of connections among actors, we may start by understanding how science is nowadays produced collectively.

2.3 Collaboration in scientific research

Publication counts are still basic objects of analysis within the framework of understanding and evaluating the determinants of scientific productivity, which is a matter of growing interest among economists and policy makers who usually perceive and evaluate the scientific production process through the achievements of those few scientists who produce a considerable amount of publications, those star scientists who are located at the right tail of the scientific productivity curve.

Since productivity and collaboration are strongly related, the question we now ask is whether this scientific production process is well understood in the context of a complex and ever growing accumulation of knowledge. Are these connections and collaborations the base for base for a better comprehension of the process? If it is so, we need to start understanding how researchers collaborate with each other mutually complementing their work.

Several institutions have thus devoted resources to boost collaboration since the general perception is that productivity and project-success rates are higher when scientists engage in collaborative efforts. Lissoni and Mairesse (2010), investigate the relationship between scientific collaboration and productivity, they classify the scientists according to the characteristics of their collaboration in order to assess how much does productivity change if scientists collaborate. They show that a substantial change in the regime of collaboration between scientists around the 1990's translated in the rapid increase of the average number of co-authors per article.

It is widely assumed that collaboration is a good thing per se, and several programs have been set with the aim of creating networks of excellence and boosting industry university links. This is only supported by the general idea that research collaboration is understood, that it behaves in the same manner for any kind of agents (individuals, institutions etc.), and that we can properly measure it and control it to increase knowledge creation.

Katz and Martin (1997) define, characterize, and classify research collaboration according to different dimensions such as individual, institutional, and international levels. They performed their analysis by defining research collaboration, its motivations, its actors, measure alternatives, and its benefits and costs. Their analysis suggests that research collaboration is hard to define, mainly because its definition obeys to social conventions among scientists for whom the frontier between collaboration and informational links is blurred. In addition, they state that measuring collaboration through co-author indicators can be misleading since there are cases where collaboration does not end in a publication, and/or where weak interactions between scientists result in one of them. Finally, they highlight that due to conceptual problems, a differentiation between inter-individual and inter-institutional collaboration is advisable.

In summary, the review discussed here connects the notions of skewness of the scientific productivity of individual scientists with the fact that scientific production is an interactive and/or distributed process where many agents are required to collaborate. These two notions support further research on the distribution of scientific productivity inside research groups, and the complementarities between the productivity of the member of one group.

3. Collective scientific production and complementarities between types of scientific personnel.

We rely on the logical assumption that the scientific production of a research group depends on the scientific effort and quality of output performed by different classes of researchers in order to study the properties of an underlying scientific production function that can only be observed as a real value denoting the total scientific output of the research group over a certain period of time.

Since these production functions form a set of ordered values which depends on several classes of scientific personnel, we want to demonstrate that the variation or evolution of one of these classes relatively to another influences the scientific production of these research organizations.

This intuition is translated into the notion of complementarities between arguments, which in our case refers to the complementarities between types of scientific personnel in terms of their representation and evolution in the research group.

Our interest is therefore focused on the scientific production as a collective process in which a group of researchers work together in order to synthesize, assimilate and create and produce new bodies of knowledge. Hence we regroup the scientific personnel of a research group around the following five different classes, which for our purpose were defined according the hierarchical structure observed within the French higher educational system:

- Senior personnel (or highly experienced: PhD. + 8 to 12 years of active research career), this category includes research directors, university professors, and medical university professors. The individuals belonging to this category represent the most elevated rank of researchers in the system and may be associated with the highly productive segment of the Lotka's curve of scientific productivity.
- Junior personnel (or just experienced: PhD. + 5 years of active research career), this category includes confirmed researchers, associate professors (or maître de conference), and medical associate professors.

- Postdoctoral researchers, or young researchers upon a well-defined research contract who are already entitled with the PhD diploma. We may notice that within this system, young researchers may prepare their qualification and go on to be associate professors without having necessarily done a post doctorate.
- PhD candidates, who are already entitled with a master's degree within the academic system. They may be referred to as young researchers although this notion is not subject to age but rather to experience.
- Assistants, including engineers, technicians, assistants, second-class professors, and other personnel who are necessary to the well development and operational performance of research groups, and who according to the definition found in the Frascati manual are considered part of the body of scientific personnel in a research institution.

These categories are thus defined not only by a notion of juridical status represented by the different diplomas a researcher may have validated, but also by a notion of experience and professional achievements the researcher may have obtained through out his scientific career and trajectory. Based on this argument, we may place these categories along the skewed curve of scientific productivity, and analyze a situation where complementary types of researchers affiliated to a same research group work collectively in order to produce a scientific output materialized in the form of scientific publications. The notion of complementarities, represented by the idea that the marginal value of an argument is increasing in the level of another argument, is well established in the economic literature. The traditional analysis relies on the assumptions of concavity and continuity of the real valued function, and is based on the property of increasing differences in the pair of variables on which the complementarities are tested.

The assumption of increasing differences implies that for all $i \neq j$, $f(x_i, x'_j) - f(x_i, x_j) \geqslant 0$, indicating that the objective function $f(x_i, x_j)$ increases in x_j, which is equivalent to:

$$\Leftrightarrow \frac{\delta[f(x_i, x'_j) - f(x_i, x_j)]}{\delta x_j} \geqslant 0$$

$$\Leftrightarrow \frac{\delta f(x_i, x'_j)}{\delta x_j} - \frac{\delta f(x_i, x_j)}{\delta x_j} \geqslant 0$$

$$\Leftrightarrow \frac{\delta f(x_i, x'_j)}{\delta x_j} \geqslant \frac{\delta f(x_i, x_j)}{\delta x_j}$$

In addition, the complementarities between arguments *i* and *j* holds if and only if $\frac{\delta f(x_i, x_j)}{\delta x_j}$ increases in x_i, which is equivalent to:

$$\Leftrightarrow \frac{\delta[\frac{\delta f(x_i, x_j)}{\delta x_j}]}{\delta x_i} \geqslant 0$$

$$\Leftrightarrow \frac{\delta^2 f(x_i, x_j)}{\delta x_j \delta x_i} \geqslant 0$$

Taken from: Topkis, 1998; Amir, 2003.

However, in our case, since the research group scientific production is only observed as a real value, which depends on several reference classes, the assumption of continuity and concavity of the objective function may or may not hold; therefore, in the impossibility to confirm these assumptions, we may avoid them by using an alternative approach, supermodularity of the objective function, to test the complementarities of pairs of reference classes.

3.1 The property of supermodularity as a working tool to assess complementarities

The notion of supermodularity formalizes the idea of complementarities between arguments. In this case, the analysis is based on the order and monotonicity of the set of the real-valued functions and of its arguments; it is also unrelated to the notions of convexity, concavity, smoothness, or returns to scale of the objective function.

Supermodularity, is a second order property of any function in R^n reflecting the condition of non-negativity on the cross differences of any pair of its arguments. This idea is analogue to the property of concavity of any continuous twice-differentiable function in R^n, which implies strong conditions on the matrix of second partial derivatives of the function.

The real valued function *f(x)* operates on a partially ordered set of arguments *X* (a sub-lattice) such that for all *x', x''* in *X*, with a partial ordering *x' < x'',* the set *X* also contains a smallest element $x' \vee x''$ under the order that is larger than both *x'* and *x''*, and a largest element $x' \wedge x''$ under the order that is smaller than both *x'* and *x''* (Topkis, 1998; Vives, 1999).

This notion implies that in an *n*-dimensional space, the elements $x' \vee x''$ and $x' \wedge x''$ are given by:

$x' \vee x'' = (max\{x'_1, x''_1\}, \ldots, max\{x'_n, x''_n\})$ and $x' \wedge x'' = (min\{x'_1, x''_1\}, \ldots, min\{x'_n, x''_n\})$

Carrying on with the definition, any real valued function *F(.)* on a partially ordered set *X* (a sub-lattice *(X, ≥))* such that for all x_i and x_j with $i \neq j$, the objective function $F(x_i , x_j)$ is supermodular in the arguments x_i and x_j, for all fixed x_i, if an only if it is characterized by the following inequality:

$$F(x' \vee x'') + F(x' \wedge x'') \geqslant F(x'') + F(x') \Leftrightarrow F(x' \vee x'') - F(x') \geqslant F(x'') - F(x' \wedge x'')$$

Which is equivalent to the following inequality when performing a pair wise comparison between arguments *i* and *j*:

$$F(x''_i, x''_j) + F(x'_i, x'_j) \geqslant F(x''_i, x'_j) + F(x'_i, x''_j) \Leftrightarrow F(x''_i, x''_j) - F(x'_i, x''_j) \geqslant F(x''_i, x'_j) - F(x'_i, x'_j)$$

This inequality confirms the property of increasing differences in any pair of arguments of $F(x_i , x_j)$ and implies the existence of complementarities between the arguments. Therefore the supermodular property of a function allows us to state that the marginal value of the objective function obtained through an increase of units of an argument in a pair is higher when the units of the other argument are also increased. From the inequality, we can deduct that the value of the sum of the objective function for any combination between the highest and lowest elements in a pair of arguments is higher when compared to any other combination of intermediate elements.

3.2 Application to the economics of science

We want to verify there are complementarities between these reference classes with respect to the expansion or contraction of scientific personnel and assess that the movement of each one of these classes with respect to the others influences the scientific output of the institution. Since the scientific production is only observed as a set of ordered real values, we may not directly assume the functional form of the scientific production of research groups; neither are we able to state whether they are endowed with continuity and differentiability properties, therefore, the innovation of the present analysis relies on the methods rather than the intuition of complementarities itself, which has been present in the economic literature for some period of time, and now in the domain of the economics of science.

The problem of the research group consists of choosing the amount of scientific personnel in each of the categories based upon their experience and status through the expansion or contraction of the number of positions filled in the institution, such that its scientific production is maximized.

Using the concept of a supermodular scientific production requires the observation and evaluation of the output from a sample of *m* research groups taking into account the influence of the arguments belonging to the set of researcher categories[1].

To verify that our data fulfills the condition of increasing differences under the framework of supermodular functions, we perform an analysis on several samples of combinations of 4 different research groups drawn from the dataset throughout a process of random sampling without repetition. Each one of the research groups in a given combination is ranked according to both variables of interest for a given pair wise comparison, and the difference between the sum of the estimated output of complementary research units and that of the intermediary ones is established to test whether its sign is positive through a one sided t-test in which the null hypothesis to be rejected is that the sample of differences equals zero.

If the scientific production is supermodular and shows complementarities in its categories of researchers, then the sum of variations of the estimated output when each category of researchers in pair increases separately is less important than the variation of the estimated output when both categories are jointly increased; which can be represented by the following inequality:

---

[1] Here we assume the scientific production of research groups is influenced by their composition: *X={Senior Researchers, Junior Researchers, Doctoral and Postdoctoral Researchers, Assistant Personnel}*.

$$\hat{f}(x_i'', x_j'') - \hat{f}(x_i', x_j') \geqslant [\hat{f}(x_i'', x_j') - \hat{f}(x_i', x_j')] + [\hat{f}(x_i', x_j'') - \hat{f}(x_i', x_j')]$$

Following Beresteanu (2005) who studies cost complementarities in telecommunications using the property of submodularity (the inverse of supermodularity), we may rewrite the expression and compute an indicator of the amount of complementarities. In presence of complementarities, the indicator is positive and equivalent to the increasing differences inequality.

$$1 \geqslant \frac{[\hat{f}(x_i'', x_j') - \hat{f}(x_i', x_j')] + [\hat{f}(x_i', x_j'') - \hat{f}(x_i', x_j')]}{\hat{f}(x_i'', x_j'') - \hat{f}(x_i', x_j')}$$

$$\frac{\hat{f}(x_i'', x_j'') - \hat{f}(x_i', x_j') - \hat{f}(x_i'', x_j') - \hat{f}(x_i', x_j'') + 2\hat{f}(x_i', x_j')}{\hat{f}(x_i'', x_j'') - \hat{f}(x_i', x_j')} \geqslant 0$$

4. Empirical evidence: the case of the former University Louis Pasteur

Within the framework of complementarities between different types of scientific personnel in research groups we used the University Louis Pasteur (current scientific branch of the unified University of Strasbourg – since January 2009) database, providing information on the research activities of the ULP, and built around three important axes: its personnel, research groups, and scientific output.

The first source of information upon which the database holds are detailed records on the research activity within the former ULP that were tracked in order to match the financial inputs and scientific output (publications, patents and industrial contracts) of the institution. These records are part of the 4-year ministry surveys that academic research groups in France fill out with the objective of obtaining the right from the ministry of higher education to pursue research activities. These surveys are compulsory for each one of the operational research groups within the academy, and ensure their survival subject to the ministry's decision making; therefore we may regard the information provided by this dataset as a reliable source of institutional data on the scientific research groups of the university.

A second source of information are two databases from the statistics office of the university concerning researchers habilitated to conduct research on the one hand, and the doctoral students on the other. Finally, complementary information was drawn from other sources such as the Internet and research reports.

From the three different data lists described above we proceeded to match the information on the basis of a two level matching: publications-personnel, and personnel-laboratory, this operation allowed us to establish an aggregated dataset providing information on the composition of the university research groups on the one hand, and their scientific output (publications) on the other.

Table 1: Details on the type and source of the initial information found in the database.

Variables	Type	Source
Scientific Production:		
Total Publications	Count variable. Represents the total number of publications in the year X belonging to authors affiliated to a research group of the university.	Interrogation on the database "Publications".
Knowledge:		
Number of contributions	Count variable. Represents the total number of scientific contributions done by internal researchers in the year of analysis.	Dataset "Publications".
Number of internal authors	Count variable. Represents the number of authors from the ULP university during the year.	Database "Publications".
Number of external authors	Count variable. Represents the number of authors non affiliated to the ULP university during the year.	Dataset "Publications".
Research group Composition:		
Status (rank) of personnel	Set of count variables for the number of individuals in the research group corresponding to each status. Notice that information on PhD's and postdocs was not available in the "Personnel" dataset.	Datasets "Laboratories" and "Personnels".
Number of HDR	Count variable. Represents the number of researchers in the research group holding the state authorization to direct research projects.	
Year of enrolment	Count variable. Represents the year in which the scientific personnel was affected to the research group. Notice this information was only available for the database "Personnel" and therefore information on the year of enrolment of PhD's and Postdocs was not available.	Dataset "Personnels"
Administrative personnel	Continuous variable. Represents the amount of scientific administrative personnel (support staff) affiliated to the research group. It is measured in terms of equivalent of full time activity.	Matching datasets "Personnels" and "Laboratoires"
Research group discipline code (DISC-3 and DISC-10)	Set of categorical variables for the research group discipline.	Dataset "Laboratoires"

In addition to the variables concerning the researchers and the administrative personnel in each research group, five types of scientific personnel were gathered from the overall database according to their status within the French academic research system. These types were created according to the classification mentioned in former sections: senior and junior personnel, postdoctoral researchers, PhD candidates, and assistant researchers. We may clarify here that the information on PhD's and postdoctoral researchers was available in the "Laboratory" dataset, while the information on the types of researchers was available in the "Personnel" data set.

We also derived from the information contained within dataset several complementary variables such as the average age and average time of affiliation of researchers across their type, the ratio of defended theses per number of PhD students, and the ratio of juniors, assistants, PhDs, and post doctors per senior researchers. All of these variables can provide us with a better view of the composition and output of the research groups.

4.1 Descriptive Statistics

4.1.1 Dataset "Laboratories"

The first part of the database structure is the Research group data list, which covers the period 1996-2008 and provides information on the research groups of the university (83 research groups in 1996, 82 in 2000, 74 in 2004, and 46 in 2008). The main information contained in this dataset concerns the composition of the research group, its expected financial needs and to some extent its available equipment. For a better understanding of this dataset we may cite some variables it contains:

- The research group code, which identifies each research group within the university and creates a link with the individual personnel.
- The number of research – teaching personnel per research group and per host institution (in full time activity ratios).
- The number of researchers affiliated to the CNRS, INSERM, and other institutions.
- The number of PhD candidates.
- The number of defended PhD theses.
- The total number of post-doctoral researchers (locals and foreigners).
- The total number of personnel (representing other administrative staff, different from researchers) per research group and per host institution.
- The total number of researchers holding the habilitation to conduct research per research group and per host institution.

These research groups are distributed across different disciplinary fields following notation and correspondence of the Observatoire des Sciences et des Techniques (OST) among scientific disciplines described in the following table.

Table 2: Scientific disciplines according to the OST nomenclature.

Discipline-3	Discipline-10
1. Life Sciences	1. Fundamental Biology
	2. Medical Research
	3. Applied Biology
2. Matter Sciences	4. Chemistry
	5. Physics
	6. Space Science
	7. Engineering
	8. Mathematics
3. Social & Human Sciences	9. Human Science
	10. Social Science

Over the three different periods covered by the ministry surveys we can observe the following distribution of the university research groups according to the tables presented here after, with an important representation of medical research and fundamental biology research groups in the DISC-10 notation, and life science research groups in the DISC-3.

Tables 3 & 4: Percentage of laboratories per discipline according to the scientific disciplines:

Discipline-10	Survey 1996	Survey 2000	Survey 2004	Survey 2008
1. Fundamental Biology	26.25	24.69	24.66	29.27
2. Medical Research	32.50	30.86	30.14	12.20
3. Applied Biology	5.00	3.70	4.11	2.44
4. Chemistry	11.25	12.35	13.70	12.20
5. Physics	5.00	4.94	5.48	7.32
6. Space Science	3.75	3.70	4.11	7.32
7. Engineering	5.00	7.41	8.22	9.76
8. Mathematics	1.25	1.23	1.37	2.44
9. Human Science	3.75	3.70	1.37	2.44
10. Social Science	6.25	7.41	6.85	14.63
Total	100	100	100	100

Discipline-3	Survey 1996	Survey 2000	Survey 2004	Survey 2008
1. Life Sciences	63.75	59.26	58.90	43.90
2. Matter Sciences	26.25	29.63	32.88	39.02
3. Social & Human Sciences	10.00	11.11	8.22	17.07
Total	100	100	100	100

Moreover, both researchers and personnel's are counted in part of time activity following the methodology of the OST (report 2008, page...), therefore those individuals strictly performing research activities count as 1, and those performing teaching and research activities count as 0.5 per type of activity adding to 1, while those performing other administrative activities may show a different fraction. The following two tables represent the distribution of personnel across disciplines; they show that in average, fundamental biology, physics, and space science laboratories employ the highest quantities of personnel in the university.

Tables 5 & 6: Distribution of personnel across disciplines.

Discipline-10	Min	Survey 1996 Mean	Max	Min	Survey 2000 Mean	Max	Min	Survey 2004 Mean	Max	Min	Survey 2008 Mean	Max
1. Fundamental Biology	1	18.58	209.4	2	21.64	233	2	22.93	237	16	64.64	305
2. Medical Research	0	5.90	16	0	6.18	19	0	7.22	16.4	7	15.97	31
3. Applied Biology	1	5.70	14	0	4.43	11.8	1.8	15.77	28.5	55	55.00	55
4. Chemistry	3.5	8.16	20.9	0	6.89	17.1	0	9.65	33.5	5	29.20	87
5. Physics	1	76.03	200.7	1	64.53	161.6	1	66.00	165	74	128.70	212
6. Space Science	13.2	20.13	29.2	14	19.60	25.8	14	21.00	25	31.6	35.37	41.7
7. Engineering	2.5	9.51	19.9	0	5.48	25	0	10.83	27.1	3	16.13	23.6
8. Mathematics	16	16.00	16	13.25	13.25	13.25	13.25	13.25	13.25	26	26.00	26
9. Human Science	0	6.13	13.4	0	3.40	10.2	0	0.00	0	14	14.00	14
10. Social Science	0.5	4.21	9.3	0.5	3.92	9	2	4.66	7.8	2	6.38	13.3
Total	0	13.72	209.4	0	13.16	233	0	15.67	237	2	41.26	305

Discipline-3	Min	Survey 1996 Mean	Max	Min	Survey 2000 Mean	Max	Min	Survey 2004 Mean	Max	Min	Survey 2008 Mean	Max
1. Life Sciences	0	11.11	209.4	0	12.51	233	0	14.39	237	7	50.59	305
2. Matter Sciences	1	23.43	200.7	0	18.00	161.6	0	20.91	165	3	45.54	212
3. Social & Human Sciences	0	4.93	13.4	0	3.74	10.2	0	3.88	7.8	2	7.47	14
Total	0	13.72	209.4	0	13.16	233	0	15.67	237	2	41.26	305

On the other hand, we present the distribution of researchers across the scientific discipline that confirms the importance of fundamental biology, physics, and space science laboratories at the heart of the university, employing the highest average of researchers.

Tables 7 & 8: Distribution of researchers across disciplines.

Discipline-10	Min	Survey 1996 Mean	Max	Min	Survey 2000 Mean	Max	Min	Survey 2004 Mean	Max	Min	Survey 2008 Mean	Max
1. Fundamental Biology	4	17.81	75	5	20.50	68	5	22.72	93	9	37.08	89
2. Medical Research	2	8.88	17	0	9.68	19	2	8.91	18	4	35.82	146
3. Applied Biology	5	8.75	12	3	6.50	13	10	14.00	19	32	32.00	32
4. Chemistry	7	17.33	31	0	13.50	34	0	12.80	25	9	32.40	103
5. Physics	11	56.25	79	13	55.50	81	9	48.00	72	55	77.67	106
6. Space Science	25	32.00	39	20	28.67	37	20	29.00	39	28	38.00	56
7. Engineering	12	23.75	38	0	18.33	51	8	24.33	61	15	40.00	80
8. Mathematics	79	79.00	79	78	78.00	78	77	77.00	77	92	92.00	92
9. Human Science	6	13.33	24	0	9.67	20	22	22.00	22	33	33.00	33
10. Social Science	6	17.40	41	1	17.00	47	10	22.90	50.5	14	39.17	104
Total	2	17.73	79	0	17.70	81	0	19.36	93	4	41.10	146

Discipline-3	Min	Survey 1996 Mean	Max	Min	Survey 2000 Mean	Max	Min	Survey 2004 Mean	Max	Min	Survey 2008 Mean	Max
1. Life Sciences	2	12.55	75	0	13.99	68	2	15.05	93	4	36.45	146
2. Matter Sciences	7	31.00	79	0	26.29	81	0	26.25	77	9	47.56	106
3. Social & Human Sciences	6	15.88	41	0	14.56	47	10	22.75	50.5	14	38.29	104
Total	2	17.73	79	0	17.70	81	0	19.36	93	4	41.10	146

To summarize the information obtained from this dataset, we present other statistics on the variables concerning the human resources of the university and its research laboratories.

Table 9: Summary statistics from the dataset "Laboratories"

Survey	Variable	Obs	Mean	Standard Err	Min	Max
1996	Researchers	83	17.40964	17.55829	2	79
	Personnel	83	13.37349	31.84602	0	209.4
	Theses	80	12.85	14.7529	1	60
	HDR	83	11.44578	12.81886	1	65
	PhD	83	14.55422	15.9054	2	82
	Post Doc	74	9.5	23.35858	1	188
2000	Researchers	82	17.79878	17.98248	0	81
	Personnel	82	13.44207	31.51415	0	233
	Theses	77	13.64935	13.88654	1	68
	HDR	82	11.2561	12.4383	0	54
	PhD	78	13.03846	13.8127	1	89
	Post Doc	59	8.084746	14.79665	1	107
2004	Researchers	74	19.25	18.52012	0	93
	Personnel	74	15.45878	33.41349	0	237
	Theses	73	13.13699	15.7174	0	105
	HDR	74	12.43919	12.82761	0	60
	PhD	73	15.9589	17.91371	0	120
	Post Doc	48	8.78125	12.03448	1	72
2008	Researchers	46	36.87174	33.47492	0	146
	Personnel	46	37.0337	55.5296	0	305
	Theses	42	26.09524	26.59251	1	117
	HDR	46	21.58696	22.26365	0	99
	PhD	42	32.16667	32.86774	0	164
	Post Doc	31	19.03226	18.012	1	63

4.1.2 Dataset "Personnel"

The second data list in the structure, the Personnel dataset, covers the period 1996-2008 and provides extensive information on the ULP's scientific personnel which accounts for 1454 individuals in 1996, 1433 in 2000, 1442 in 2004, and 2644 in 2008. A better comprehension of the Personnel dataset may be obtained through a brief description of some important variables it contains:

- The research group code, which allows establishing a link between the individual researcher and the research group he is affiliated to.
- The hierarchic status of the researcher, which reflects the administrative status of the researcher in the database as stated in the ministry survey. It also conveys the notion of a juridical status regarding the qualifications and position of the researcher within the French academic system.
- The host institution to which the researcher is affiliated; this variable may correspond either to the University Louis Pasteur, the CNRS[2], the INSERM[3], the INRA[4], the IRD, other universities in Strasbourg, other universities in Alsace, or other companies. We may notice that in France, an important amount of university research units are mixed, benefiting from different participating research institutions.

[2] Centre National de la Recherche Scientifique.
[3] Institut National de la Santé et de la Recherche Médicale.
[4] Institut National de la Recherche Agronomique.

- The CNU[5] code, which points out the disciplinary field in which the research group of affiliation is active. For the purpose of clarity, this codification can be easily translated into the OST codification of scientific disciplines.
- The habilitation to conduct research, which is an official state qualification, passed by the researcher in order to have the right to be the head of a research project.
- The researcher's date of birth.
- The researcher's date of entry in the research group.

We were also able to extract which types of researchers were declared by research group directors for the ministry survey, and relevant information on their age and period of time span of affiliation to the lab. We must notice here that a small percent of the population within this list (1.5%, 104 observations out of 6973 – all periods included) shows a negative time of affiliation to the lab possibly indicating that the position was declared but unfilled at the moment of the survey.

With the information on the different ranks of researchers within this data list we were also able to establish a correspondence between the status of the individual researchers and the hierarchical structure of the French research system, therefore obtaining information on the senior, junior and assistant researchers at the individual level.

These individuals are the same ones declared in the research group data list, and their sum over a specific research group should match the number of researchers indicated in the research group list. However, even if these two variables are not the same they correlate at the level of 0.9697; therefore we assume we can use the information coming from the personnel's data list at the research group level and benefit from the possibility of separation of researcher types it provides, which would have not been possible otherwise.

The observed distribution of researchers according to their host institution, their rank, and their category shows that the biggest shares of researchers are affiliated either to the university itself performing teaching and research tasks, or to the CNRS performing research only tasks.

Table 10: Percentage of researchers across the different host institutions:

	Survey 1996	Survey 2000	Survey 2004
CNRS	41.42	38.41	35.44
INSERM	5.88	6.77	5.62
ULP	49.03	48.04	45.84
AUTRES	4.22	6.84	13.11
Total	100	100	100

Furthermore, we present the distribution of researchers according to their rank within the scientific laboratory. It suggests that the most important shares of researchers fall in the most experienced

[5] Conseil National des Universités.

scientific categories: senior and junior researchers, with the ranks of directors, professors, and associates.

Table 11 & 12: Percentage of researchers according to their rank and category.

	Survey 1996	Survey 2000	Survey 2004	Survey 2008
Research Director	20.26	19.13	17.82	15.62
University Professor	17.91	16.41	17.41	18.46
Medical Professor and Hospital Practitioner	3.39	4.19	4.09	4.22
Associate Professor (MC)	24.41	25.77	29.47	32.96
Associate Hospital Practitioner	2.97	3.49	4.23	3.10
Associate Researcher	27.18	26.82	24.27	23.47
Engineers and Technicians	0.90	0.84	0.42	0.07
Other Hospital Staff	0.76	0.91	0.07	
Space Observation	1.52	1.47		1.45
Other Assistants	0.35	0.28		
Private Sector	0.14		2.22	0.07
Highschool Teachers	0.21	0.70		0.59
Total	100	100	100	100

	Survey 1996	Survey 2000	Survey 2004	Survey 2008
Senior	41.56	39.73	39.32	38.30
Junior	54.56	56.08	57.98	59.53
Research Support Staff	3.87	4.19	2.70	2.18
Total	100	100	100	100

4.1.3 Dataset "Publications"

The third part of the structure is the Publications data list, which provides information on more than 40000 publications done by the permanent scientific personnel of the university research groups during the period 1990-2005. This information was drawn from the Science Citation Index and Social Science Citation Index (ISI Web of Science) through several requests on the articles published in the journals listed within ISI, pointing out regional addresses and research groups, and contains information on the year of publication, main author and co-authors. In addition, since the publications are identified for each individual personnel, it becomes straightforward to establish a link between the publications and the personnel according to an individual-time dimension. Some variables we may cite are:

- The identifier of a publication done by an individual researcher.
- The identification code of the research group.
- The year of publication.
- The number of co-authors (internal and external to the ULP).

The following tables provide the distribution of total publications and weighted publications across time and disciplines. They show the average total publications and the average weighted[6] publications of research groups, and reveal fundamental biology, chemistry, physics, space science, and mathematics research groups as the most performing ones in the university.

Tables 13 & 14: Mean total publications across disciplines by year:

Discipline-10	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1. Fundamental Biology	29.38	31.57	30.76	31.24	34.95	34.30	35.00	31.65	37.89	15.88
2. Medical Research	16.37	17.85	18.59	18.56	22.52	20.00	20.56	18.96	18.73	8.19
3. Applied Biology	7.25	10.33	9.00	9.50	10.33	9.33	14.67	17.00	14.33	8.00
4. Chemistry	39.44	38.44	37.67	40.89	36.90	34.90	35.70	35.20	35.60	13.50
5. Physics	94.00	102.25	96.50	99.25	103.00	107.75	106.00	106.00	113.00	44.00
6. Space Science	32.00	35.33	37.00	39.00	36.00	28.00	34.67	40.33	42.00	15.33
7. Engineering	14.20	12.60	14.00	19.60	10.50	11.00	11.00	11.86	14.50	7.17
8. Mathematics	48.00	40.00	42.00	37.00	38.00	30.00	46.00	37.00	45.00	21.00
9. Human Science	5.00	3.33	5.50	5.00	15.00	4.00	9.50	10.00		
10. Social Science	7.00	4.67	3.00	4.00	3.75	7.50	1.33	2.00	1.00	2.00
Total	26.11	27.39	27.47	27.99	30.04	28.58	29.35	29.32	32.41	13.55

Discipline-3	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1. Life Sciences	20.92	23.06	23.04	22.98	26.94	25.29	26.21	24.28	26.44	11.45
2. Matter Sciences	43.00	43.82	43.09	46.23	41.25	40.00	40.32	40.68	44.42	17.54
3. Social & Human Sciences	5.80	4.00	4.00	4.50	6.00	5.40	4.60	6.00	1.00	2.00
Total	26.11	27.39	27.47	27.99	30.04	28.58	29.35	29.32	32.41	13.55

Tables 15 & 16: Mean weighted publications across disciplines by year.

Discipline-10	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1. Fundamental Biology	2.60	2.52	2.74	2.51	2.88	2.70	2.64	2.29	2.47	1.70
2. Medical Research	1.03	1.10	1.17	1.01	1.27	1.12	1.15	1.07	1.11	0.91
3. Applied Biology	1.13	1.19	1.40	1.10	1.24	1.07	1.08	1.69	1.68	1.01
4. Chemistry	2.96	2.87	2.92	2.82	2.53	2.44	2.60	2.44	2.60	1.60
5. Physics	5.94	5.93	5.61	5.88	5.78	4.50	4.07	3.52	3.65	2.89
6. Space Science	4.11	4.02	4.94	4.96	3.99	3.00	3.23	3.62	3.31	2.22
7. Engineering	2.12	2.16	2.42	2.23	1.57	1.37	1.64	1.48	2.74	1.44
8. Mathematics	16.42	17.06	15.27	9.96	10.50	10.65	14.15	11.46	12.64	9.69
9. Human Science	1.52	0.81	1.45	0.94	2.89	1.34	1.81	2.00		
10. Social Science	2.19	2.00	1.44	2.17	1.50	4.04	1.00	0.67	0.08	0.33
Total	2.35	2.33	2.45	2.23	2.37	2.18	2.17	2.03	2.26	1.58

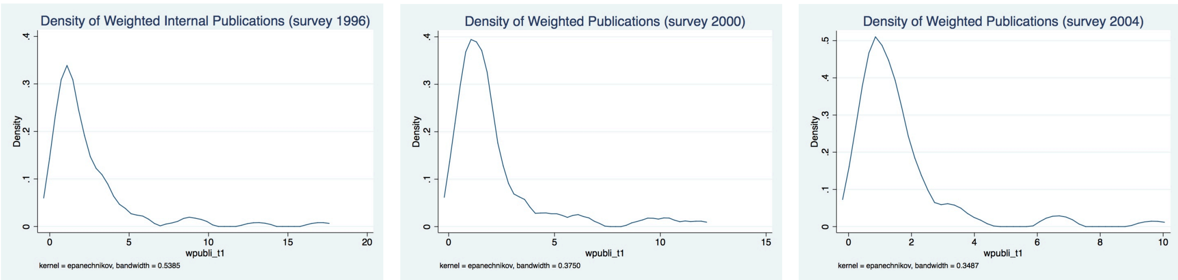
Discipline-3	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1. Life Sciences	1.67	1.69	1.83	1.62	1.94	1.77	1.77	1.62	1.72	1.25
2. Matter Sciences	4.08	4.07	4.13	3.86	3.35	2.93	3.10	2.84	3.32	2.19
3. Social & Human Sciences	1.79	1.40	1.44	1.56	1.78	2.42	1.33	1.33	0.08	0.33
Total	2.35	2.33	2.45	2.23	2.37	2.18	2.17	2.03	2.26	1.58

4.2 Institutional and intellectual assets as determinants of scientific production

We use as dependent variable of our analysis the total part of internal contributions to a set of publications in peer reviewed journals done by members of a scientific laboratory or research group in the year following the ministry survey so that we are sure that the personnel declared in the survey is responsible for those contributions.

According to the density function of these internal research group publications, we learn that the dependent variable is highly non-normal and positively skewed, the following figures provide some insights on the shape of the density function for the total publications of the research groups following each ministry survey.

Figures 1, 2, & 3: Density of weighted publications in the year following the ministry survey.



Based on the observations described above and using short and imbalanced panel of data composed of the first 3 surveys (1996, 2000, and 2004), we perform a pooled panel regression based on an exponential mean model for the total internal contributions to publications $y_i = exp(x_i'\beta)\varepsilon_i$, and by defining $\varepsilon_i = exp(u_i)$ and taking the natural logarithm we fit the log-linear model $ln(y_i) = x_i'\beta + u_i$ with the objective to elicit the relationship between the internal scientific output of the research group and its composition.

The first phase in the analysis on was to establish a relationship between the publications and the intellectual capital of our research groups in the dataset through the following specifications of the model.

Specification 1:

$ln(Internal\ Publications_i) = \beta_1\ Researcher\ Teachers_i + \beta_2\ Researchers\ CNRS_i + \beta_3\ Researchers\ INSERM_i + \beta_4\ Other\ Researchers_i + \beta_5\ University\ Personnel_i + \beta_6\ Personnel\ CNRS_i + \beta_7\ Personnel\ INSERM_i + \beta_8\ Other\ Personnel_i + u_i$; controlling for the discipline field of the research group, the average age of its researchers, and their average career in the laboratory.

Specification 2:

$ln(Internal\ Publications_i) = \beta_1\ Senior\ Researchers_i + \beta_2\ Junior\ Researchers_i + \beta_3\ Doctoral\ and\ Postdoctoral\ Researchers_i + \beta_4\ Staff\ Personnel_i + u_i$; controlling for the discipline field of the research group, the average age of its researchers, and their average career in the laboratory.

The first model establishes a link between the internal publications performed by a research group in the year t+1 following the ministry surveys and the institutional identity of the researchers working in the group; it estimates the influence of the total number of researchers and administrative personnel according to the institution they belong to, and addresses our interest of understanding how French scientific laboratories are influenced by the mixed unit structure that combines employees from different institutional hosts such as the academia and national research centers.

The second model establishes a link between the internal publications and the different categories of researches working within the group at the moment of the ministry survey, therefore estimating the influence of the intellectual assets of the group on its output. This model is of particular interest because it addresses the issue of the influence of different categories of researchers defined according to their rank within the French research system, and serves to predict the estimated output upon which our analysis of complementarities takes place.

Both models include a control for the main disciplinary field in which the research group is active (notations DISC-3 of the OST, life science, matter science, and social and human science), the average of its members and their average career in the laboratory.

The results obtained from the first specification show the weighted publications are significantly and positively influenced at the 1% level by the presence of those researchers affiliated to the university and to the CNRS, although the presence of certain categories of administrative personnel (CNRS and others) seem to penalize this output.

The results obtained from the second model show the weighted publications are significantly and positively influenced by the presence of senior researchers as well as post doctors and PhD students at the 1% level, and junior researchers at the 1% level, although we were surprised by the fact that the administrative support and assistants turned out to have a negative influence on the output at the 1% level. In addition, we may notice there are positive and significant average age effects telling us that older research groups (in average) are likely to contribute more.

Table 17: Results obtained from the econometric regressions.

Specification 1 (Host Institutions)			Specification 2 (Categories of Researchers)		
Explanatory Variable	Log Weighted Publications	Log Total Publications	Explanatory Variable	Log Weighted Publications	Log Total Publications
Researcher Teachers	0.0303***	0.0149*	Senior Researchers	0.0451***	0.0517***
	(0.0066)	(0.0076)		(0.0108)	(0.0155)
Researcher CNRS	0.0588***	0.0773***	Junior Researchers	0.0195**	-0.0040
	(0.0106)	(0.0123)		(0.0088)	(0.0126)
Researcher INSERM	-0.1623	-0.0253	PhD's and Post Doctors	0.0161***	0.0085***
	(0.1019)	(0.1186)		(0.0021)	(0.0030)
Researcher OTHER	0.0030	0.1317**	Staff	-0.0152***	0.0003
	(0.0482)	(0.0562)		(0.0019)	(0.0027)
Personnel University	0.0024	0.0294*	Squared Average Age	0.0003**	0.0004**
	(0.0131)	(0.0153)		(0.0001)	(0.0002)
Personnel CNRS	-0.0169***	-0.0060	Squared Average Career	-0.0001	-0.0003
	(0.0043)	(0.0050)		(0.0004)	(0.0006)
Personnel INSERM	0.1671**	0.1055	Matter Science	0.1224	0.0661
	(0.0677)	(0.0788)		(0.0971)	(0.1390)
Personnel OTHER	-0.0231*	-0.0468***	Human and Social Science	-0.0634	-1.5537***
	(0.0116)	(0.0135)		(0.2484)	(0.3553)
Squared Average Age	0.0001	0.0004	Constant (baseline Life Science)	-0.9852***	1.3012***
	(0.0003)	(0.0003)		(0.3088)	(0.4417)
Squared Average Career	-0.0003	-0.0015			
	(0.0010)	(0.0011)			
Matter Science	0.1115	0.2725			
	(0.1799)	(0.2094)			
Human and Social Science	-0.9302*	-1.1216*			
	(0.4937)	(0.5749)			
Constant (baseline Life Science)	-0.6535	0.5389			
	(0.6305)	(0.7341)			
N	50	50	N	165	165
R2	0.7551	0.7491	R2	0.6567	0.4875
F	9.5087	9.2072	F	37.3041	18.5469
RMSE	0.4794	0.5582	RMSE	0.5253	0.7513
Standard errors in parentheses			Standard errors in parentheses		
* p<.10 ** p<.05 *** p<.01			* p<.10 ** p<.05 *** p<.01		

4.3 Supermodularity of the estimated scientific production

Using a supermodular approach to verify the existence of complementarities between pairs of researcher categories requires testing whether the real valued production function of research groups fulfills the increasing differences condition of supermodular functions. We perform this analysis on the estimated value of the internal publications of research groups; therefore we use the vector of predicted values for the publications generated from by the parametric regressions carried in the previous section.

We perform series of evaluations of the number of scientific publications; for each one of the survey periods we proceeded with several random extractions of four different research groups at a time, each one fulfilling a position within the structure of the increasing differences inequality given the lowest and highest elements in a couple of arguments (researcher categories).

As explained in section *3.2*, we ranked each research group in a random combination of four according to two of its categories of researchers, with rank 1 being the research group with the highest counts of researchers in each category, rank 2 being the research group with the lowest counts of researchers in each category, and ranks 2 and 3 being the other two research groups with intermediary counts of researchers; as an example take four research groups randomly drawn from the sample (alpha, beta, gamma, and delta) and rank them according to a pair of categories of researchers (tall and short), the first research group, alpha, has 2 tall and 2 short researchers, therefore it holds rank 1, the second research group, beta, has 1 tall and 1 short researchers, therefore it holds rank 2, they are defined as the complementary research groups in the combination; the other two research groups, gamma and delta, have 1 tall and 2 short, and 2 tall and 1 short researchers respectively, and thus they hold ranks 3 and 4 respectively, and are defined as the intermediate research groups in the combination.

We then proceed to test whether the difference between the sum of the estimated publications done by the two complementary groups (ranks 1 and 2) and the sum of the estimated publications done by the intermediate groups (ranks 3 and 4) is positive. In addition, we compute the following indicator to get an idea about the degree of the complementarities between the couple of researcher categories.

$$1 - \frac{[\hat{f}(x_i'', x_j') - \hat{f}(x_i', x_j')] + [\hat{f}(x_i', x_j'') - \hat{f}(x_i', x_j')]}{\hat{f}(x_i'', x_j'') - \hat{f}(x_i', x_j')} \geqslant 0$$

The process is repeated several times to obtain a sample of evaluations of the estimated internal publications for the comparison for a given couple of researchers.

Following the establishment of such samples, we performed a one sided t-test where the null hypothesis states the inequality is equal to zero against the alternative hypothesis stating the inequality is positive: *H_0: The main increasing difference of the sample equals zero, against H_1: the mean increasing difference of the sample is positive.*

A rejection of the null hypothesis implies there are of complementarities among the couple of researcher categories in question. This would allow us to conclude there is a certain "research delegation effect" between highly experienced researchers, less experienced ones and assistants or support staff. Such delegation would clearly imply a strong link between different segments of the distribution curve of the scientific productivity, which we know is positively skewed (Lotka, 1926).

The results obtained from the analysis of complementarities between types of researchers using the property of supermodularity of the estimated weighted publications of scientific research groups shows there are complementarities between all possible couples of researchers. In fact, for all samples of estimated weighted publications drawn from the possible combinations of research groups we obtained an positive average value for the increasing differences inequality; and once the one sided test was performed on them we were able to reject the null hypothesis. We conclude that the estimated weighted publications function is supermodular, and that there are complementarities between the categories of researchers; we may also notice that the indicators of complementarities are situated near the unit, suggesting benefits from equally proportional changes in the composition of the research groups.

Table 18: Results obtained from the analysis of complementarities and supermodularity.

Survey	Couple	Mean Increasing Differences	t Statistic MID = 0	Upper Pvalue for rejection of Null	Mean Indicator	Reject Null	SPM Complementarity
1996	Senior-Junior	37.282	12.932	4.24E-34	0.7955	Yes	Yes
1996	Senior-PhDs and Post Doctors	42.786	17.086	0	0.7886	Yes	Yes
1996	Senior-Staff	49.406	15.137	0	0.8283	Yes	Yes
1996	Junior-PhDs and Post Doctors	45.207	15.640	0	0.8672	Yes	Yes
1996	Junior-Staff	53.033	16.819	0	0.9366	Yes	Yes
1996	PhDs and Post Doctors-Staff	60.917	18.186	0	0.8907	Yes	Yes
2000	Senior-Junior	28.387	11.078	1.48E-23	0.8437	Yes	Yes
2000	Senior-PhDs and Post Doctors	31.277	12.371	2.96E-28	0.8277	Yes	Yes
2000	Senior-Staff	37.873	14.967	5.26E-37	0.8136	Yes	Yes
2000	Junior-PhDs and Post Doctors	36.894	14.025	1.02E-33	0.9045	Yes	Yes
2000	Junior-Staff	38.485	15.660	6.67E-39	0.8983	Yes	Yes
2000	PhDs and Post Doctors-Staff	37.371	14.161	4.15E-35	0.8135	Yes	Yes
2004	Senior-Junior	55.795	7.849	1.16E-13	0.8501	Yes	Yes
2004	Senior-PhDs and Post Doctors	74.953	9.798	1.05E-19	0.8119	Yes	Yes
2004	Senior-Staff	82.301	9.206	3.43E-17	0.8497	Yes	Yes
2004	Junior-PhDs and Post Doctors	63.682	8.145	1.53E-14	0.8167	Yes	Yes
2004	Junior-Staff	67.787	8.633	6.72E-16	0.8822	Yes	Yes
2004	PhDs and Post Doctors-Staff	82.407	8.147	3.41E-14	0.7714	Yes	Yes

5. Concluding remarks

In summary, the insights provided by the regressions we performed show significantly effects of senior and junior researchers, as well as post doctors and PhD students on the output of research groups, and that a relation of complementarities between these categories of researchers exist, allowing the research group to benefit from collective research efforts. Although the assistant and support staff seems to provide a negative effect on the output, we learn there is a complementary link between this category and all other types of researchers, implying benefits from the presence of such assisting personnel in the laboratory.

In conclusion, not only principal investigators but also all other types of scientific personnel determine the scientific production of research groups throughout a relationship of complementarity. This implies the existence of link between the higher, mid, and lower sectors of the scientific productivity curve first depicted in the works of Lotka, 1926.

References

1. Allison, P. D. & Stewart, J. A. (1974) Productivity Differences Among Scientists: Evidence for Accumulative Advantage. *American Sociological Review* 39, 596 - 606.

2. Amir, R. (2005) Supermodularity and Complementarity in Economics: An Elementary Survey. *Southern Economic Journal* 71, 636 - 661.

3. Arora, A., David, P. A., & Gambardella, A. (1998) Reputation and competence in publicly funded science: estimating the effects on research group productivity. *Annales d'Economie et de Statistique* , 163 - 198.

4. A. Beresteanu. Nonparametric analysis of cost complementarities in the telecommunications industry. RAND Journal of Economics, pages 870–889, 2005.

5. Breschi, S., Lissoni, F., & Montobbio, F. (2007) The scientific productivity of academic inventors: new evidence from italian data. *Economics of Innovation \& New Technology* 16, 101 - 118.

6. Chambers, C. P. & Echenique, F. (2009) Supermodularity and preferences. *Journal of Economic Theory* 144, 1004 - 1014.

7. Coombs, R., Harvey, M., & Tether, B. (2001) Analysing distributed innovation processes: a CRIC position paper.

8. Diamond, A. (1986) What is a Citation Worth?. *The Journal of Human Resources* 21, 200 - 215.

9. Giovanni and Dosi. Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. Research Policy, 11(3):147 – 162, 1982.

10. Egghe, L. (2009) A rationale for the Hirsch-index rank-order distribution and a comparison with the impact factor rank-order distribution. *Journal of the American Society for Information Science and Technology* 60, 2142 - 2144.

11. Egghe, L. (2007) Distributions of the h-index and the g-index.

12. Galia, F. & Legros, D. (2004) Testing for complementarities between team, incentives, training and knowledge management: Evidence from France, *ERMES working paper*.

13. Gouriéroux, C., Holly, A., & Monfort, A. (1982) Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters. *Econometrica* 50, 63-80.

14. Guidetti, G., Mancinelli, S., & Mazzanti, M. (2009) Complementarity in Training Practices Methodological Notes and Empirical Evidence for a Local Economic System. *The IUP Journal of Applied Economics* 8, 39 - 56.

15. Hage, J. & Clignet, R. (1982) Coordination Styles and Economic Growth. *Annals of the American Academy of Political and Social Science* 459, 77 - 92.

16. Katz, J. S. & Martin, B. R. (1997) What Is Research Collaboration?. *Research Policy* 26, 1 - 18.

17. Kodde, D. A. & Palm, F. C. (1986) Wald Criteria for Jointly Testing Equality and Inequality Restrictions. *Econometrica* 54, pp. 1243 - 1248.

18. F. Lissoni, J. Mairesse, F. Montobbio, and M. Pezzoni. Scientific Productivity and Academic Promotion: A Study on French and Italian Physicists. *Industrial and Corporate Change*, 2010.
19. Love, J. H. & Roper, S. (2009) Organizing innovation: Complementarities between cross-functional teams. *Technovation* 29, 192 - 203.
20. MacRoberts, M. H. & MacRoberts, B. R. (1982) A Re-Evaluation of Lotka's Law of Scientific Productivity. *Social Studies of Science* 12, 443 - 450.
21. Merton, R. K. (1968) The Matthew effect in science: The reward and communication systems of science are considered. *Science* 159, 56.
22. Mohnen, P. & Röller, L.-H. (2005) Complementarities in innovation policy. *European Economic Review* 49, 1431 - 1450.
23. Radner, R. (1986) The Internal Economy of Large Firms. *The Economic Journal* 96, 1 - 22.
24. Rumelt, R. P. (1982) Diversification Strategy and Profitability. *Strategic Management Journal* 3, 359 - 369.
25. Strube, E. & Resende, M. (2009) Complementarity of Innovation Policies in the Brazilian Industry: An Econometric Study.
26. Topkis, D. M. (1998) Supermodularity and complementarity.
27. Vives, X. (2007) Supermodularity and Supermodular Games. *Occasional paper*.