UNIVERSITE DE STRASBOURG
FACULTE DES SCIENCES ECONOMIQUES ET DE GESTION

ECOLE DOCTORALE AUGUSTIN COURNOT

Thèse de Doctorat en Sciences Économiques

ANALYSE DE LA PRODUCTION SCIENTIFIQUE :
Rôle de la complémentarité des facteurs

AN ANALYSIS OF SCIENTIFIC PRODUCTION:
The complementary role of factors

Christian MARTINEZ DIAZ

Thèse dirigée par **Patrick LLERENA**, professeur
Université de Strasbourg

Jury
Guido BUENSTORF
Universität Kassel

Francesco LISSONI
Université Montesquieu, Bordeaux IV

Julien PENIN
Université de Strasbourg

Présentée et soutenue publiquement, le 21 Février 2014

Dissertation Committee

Professor Guido BUENSTORF

Professor Francesco LISSONI

Professor Julien PENIN

Professor Patrick LLERENA

L’Université de Strasbourg n’entend donner aucune approbation ou improbation aux opinions émises dans les thèses. Ces opinions doivent être considérées comme propres à leurs auteurs.

A mi Abuelita, Doña Barbara Fontalvo
A mi Madre, Aurora Díaz
À ma Compagne, Fanny Bertrand
A mi Hermana, Fabiana Sofia Martínez

Acknowledgements

I would like to thank Professor Patrick Llerena for giving me the opportunity to carry out this research work, for his counsel and his advices along these years. I thank colleagues at Beta, the Doctoral School and the Faculty for having provided me with their help and support throughout this time, especially Danielle who has always helped us with a smile since the beginning.

I would like to thank my friends and colleagues from the doctoral school for all our discussions and exchanges of opinion and points of view which over the past years helped shape my thoughts. A special thought goes to Faustine, Mickaël and Xi, who were there, either just across the desk, the hall or just at the other end of the line, always ready to help, have a nice chat, smile, and have good times together.

I would like to thank my partner Fanny for always being there in the moments I needed and my family in law for accepting me and providing us with the moral support we've needed through the latter years.

Finally, I have a special thought for my little sister, Fabiana, whose condition during the latter years has reminded me that our time is limited and that sometimes we take for granted important things in daily life and pay too much attention to things that may not be as "important" as we may think.

Table of Contents

List of Figures

List of Tables

Résumé en Français

Résumé en Français

La compréhension par les économistes des modes d'organisation et de production des connaissances scientifiques reste limitée malgré les développements récents de l'économie de la science (Dasgupta et David, 1994, Nelson, 2004, Stephan, 2008). Or la recherche et l'innovation sont devenues des enjeux majeurs pour la compétitivité des économies européennes et leur croissance (Meyer-Krahmer, 1989).

La littérature, essentiellement anglo-saxonne, s'attache à l'étude d'universités en tenant compte de l'individu/chercheur comme unité d'analyse et s'intéresse beaucoup plus rarement aux organismes publics de recherche et la recherche collective qui en résulte de la collaboration entre différents chercheurs dont les caractéristiques sont hétérogènes. Ces travaux de thèse portent une attention particulière non seulement au rôle de l'individu/chercheur au sein du processus de production scientifique mais aussi à celui des laboratoires de recherche en tant qu'organisations composés des groupes de chercheurs.

Le cas du système de recherche académique Français, dans lequel nous trouvons une structure mixte au sein des laboratoires de recherche académique, nous permet également de distinguer le rôle des organismes de recherche publique comme le CNRS ou l'INSERM dans cette organisation et production collective de connaissances. Ces organismes étant consacrés à la recherche de base ou fondamentale, leur apport à la croissance économique est indiscutable (Stephan, 1996).

Plus particulièrement, l'objectif de ces travaux est de mettre en lumière les caractéristiques propres aux phénomènes de production et de qualité scientifique. Cette thèse s'intéresse aux questions relatives aux caractéristiques de la production et à la qualité de la recherche scientifique, d'une part aux déterminants en matière des ressources humaines et financières de ces phénomènes, et d'autre part aux relations existantes entre différents éléments issus de la décomposition de ces deux types de ressources. L'intérêt porté à ces questions permettra ainsi d'effectuer des recommandations en termes de gestion institutionnelle d'équipes de chercheurs et des ressources financières favorisant la production publique des connaissances issue de ces deux organismes publics.

Ainsi, ces travaux sont centrés sur des questions relatives à la composition des équipes et aux modes de financement des laboratoires de recherche en tant que déterminants de la production et de la qualité scientifique, à la collaboration entre différents types de chercheurs, et aux relations de complémentarité entre différents types de financement de la recherche publique. Les paragraphes suivants fournissent une brève description des différents chapitres composant ces travaux.

Chapitre 1 : Revue de littérature

Dans le premier chapitre de ces travaux, nous effectuons un rappel de la littérature en économie de la science. Nous nous servons de différents concepts et résultats des travaux existants afin de contextualiser nos études sur la production et qualité scientifique, ainsi que l'existence de liens de complémentarité entre chercheurs de différents statuts et entre différentes sources de financement.

Ainsi, nous commençons par distinguer le fait que la production de la recherche scientifique est asymétrique dans sa distribution, la plupart des publications scientifiques étant produites par des chercheurs très expérimentés, dits seniors, alors que des travaux récents montrent que cette même production est assurée par des collaborations entre chercheurs. Nous plaçons ainsi nos travaux de recherche sous le cadre d'un processus de production et de qualité scientifique interactif et collectif.

Ce cadre de travail est délimité par différents thèmes revus dans cette littérature tels que, d'une part, les différents environnements de recherche, le rôle de la structure hiérarchique au sein des groupes de recherche et l'organisation sociale des laboratoires de recherche, et d'autre part, les comportements des agences de financement de la recherche ou la structure de financement des laboratoires et son effet sur la production scientifique.

Chapitre 2 : Ensembles de données

Nous avons travaillé sur les informations fournies par la base de données concernant les laboratoires publics de recherche au sein de l'ancienne Université Louis Pasteur. Ces informations portent sur la composition et la production des laboratoires de recherche de l'université au cours de la période 1996 – 2008.

Dans le cadre de ces travaux de thèse, cette base a été élargie au niveau des données concernant la production scientifique des laboratoires de recherche de l'université. La base originale s'arrêtant en l'année 2005, nous avons dû récupérer les informations manquantes jusque 2010 afin de pouvoir exploiter les informations présentes dans le contrat quadriennal de 2008, élargissant ainsi notre période d'analyse à 4 quadriennaux. Par ailleurs, on a également élargit la base aux informations concernant les facteurs d'impact des journaux scientifiques dans lesquels les chercheurs affiliés à l'université ont publié au cours de la période 1996 – 2010 permettant ainsi la dernière analyse. Finalement, on a obtenu des informations concernant les financements des laboratoires au cours de la période 2001 – 2008 qui nous ont permis d'effectuer les analyses portant sur l'impact des ressources financières.

L'ensemble des données peut ainsi être décomposé en quatre grands ensembles d'information : les laboratoires, le personnel, les publications et les financements ; ces ensembles ont été fusionnés à l'aide des identificateurs propres aux laboratoires et aux individus/chercheurs, ce qui nous a permis de procéder aux différentes études en tenant compte non seulement de l'individu mais aussi le laboratoire comme unité d'analyse.

Chapitre 3 : Déterminants de la production et de la qualité scientifique

En construisant sur la base des travaux de Carayol et Matt (2006), une première analyse est effectuée afin d'évaluer l'impact des ressources humaines et financières sur la production et sur la qualité de la recherche publique dans les laboratoires de recherche. Cette analyse propose une décomposition de ces deux types de ressources (selon le rang des chercheurs au sein du système académique Français et selon le type/provenance des fonds de financement). Cette décomposition nous a permis de mieux comprendre le phénomène de la production scientifique collective au sein des laboratoires de recherche publique financée par des capitaux publics ou privés.

Dans ce chapitre nous analysons la production scientifique collective des laboratoires mesurée par le nombre total des contributions aux articles scientifiques et par le nombre total des publications fractionnaires produites par l'ensemble de chercheurs associés à un laboratoire de recherche ; nous analysons également la qualité de la production scientifique mesurée par le facteur d'impact moyen et médian des revues scientifiques dans lesquelles les chercheurs associés à un laboratoire publient. Ces phénomènes de production et de qualité scientifiques sont analysés selon différents modèles de régression tenant compte de la composition des ressources humaines et des modes de financement en tant que variables explicatives des deux phénomènes. Finalement, les mêmes analyses sont effectuées en termes de recherche individuelle, prenant cette fois ci le chercheur individuel comme unité d'analyse.

Les résultats concernant les analyses dans ce chapitre montrent que la production scientifique des laboratoires est en effet affectée par une décomposition des ressources humaines et que la production scientifique individuelle est affectée par le statut des chercheurs individuels, la classe de chercheurs seniors étant déterminante de cette production. Cependant, la qualité de la recherche scientifique des laboratoires et des individus n'est pas nécessairement expliquée par une décomposition de ces ressources.

En termes d'influence des différents types de financement, les résultats montrent qu'au niveau des laboratoires, tant la production comme la qualité scientifique peuvent être expliquées par la disponibilité des sources de financement publiques, régionales ou privés.

Chapitre 4 : Complémentarité entre différents éléments des ressources humaines et financières dans la recherche publique

À partir de la décomposition des ressources humaines et financières des laboratoires de recherche et l'analyse de leur impact sur la production scientifique, on s'aperçoit que ces catégories de chercheurs peuvent être complémentaires au sein des laboratoires de recherche. Ces complémentarités peuvent donc définir la performance de production scientifique des laboratoires et organismes publics de recherche. Nous tentons ainsi de définir la complémentarité entre différents types de chercheurs en tant que facteur déterminant de cette production.

Afin de mener nos analyses sur les relations de complémentarité entre les différents types de personnel scientifique, et entre les différents types de ressources de capital des laboratoires, nous adaptons la notion de la supermodularité d'une fonction à valeurs réelles à notre champ d'étude afin d'établir les liens de complémentarité existants entre différents éléments de la décomposition des ressources humaines et des modes de financement. Cette méthode d'analyse est peu courante dans la littérature actuelle, mais peut s'avérer très utile lors des analyses de liens de complémentarité entre différents arguments d'une fonction. Ainsi, une analyse empirique sur la supermodularité d'une fonction peut être effectuée selon une méthode de régression paramétrique (Athey et Stern, 1996) dans laquelle on établit l'impact des états complémentaires et des états intermédiaires des arguments sur les variables dépendantes, ou bien selon une méthode non paramétrique (Beresteanu, 2005) dans laquelle on mesure la distance entre des estimateurs non paramétriques des variables dépendantes obtenus sous les différents états complémentaires et intermédiaires des arguments.

Les résultats des analyses contenues dans ce chapitre nous montrent que les collaborations entre chercheurs de différent statut constituent des déterminants de la production scientifique des laboratoires comme des individus, tandis qu'elles ne révèlent pas des effets sur la qualité de celle-ci. Nous trouvons ainsi des liens de complémentarité entre chercheurs de type senior et de type junior, comme des liens de complémentarité entre chercheurs seniors et assistants.

Ces liens de complémentarité sont obtenus grâce à des analyses de type supermodulaire, qui évaluent la condition de différences accrues de la fonction de production scientifique dans un couple d'arguments explicatifs, en l'occurrence, les paires de statuts des chercheurs basés sur la décomposition de ces ressources.

General Introduction

General Introduction

The understanding of organization and production modes of scientific knowledge by economists remains limited despite recent developments in the economics of science (Dasgupta and David, 1994, Nelson, 2004, Stephan, 2008). However, research and innovation have become major issues for the competitiveness of European economies and their growth (Meyer-Krahmer, 1989).

The literature base mostly focuses on the study of universities taking into account the individual researcher as the unit of analysis, while only a few studies focus in public research organizations and the collective research output resulting from the collaboration between different researchers, whose characteristics are heterogeneous. This thesis draws the public's attention not only towards the role of the individual researcher in the scientific production process but also towards the role of research laboratories and their composition in terms of human resources and types of funding. The case of the French academic research system, in which we find a mixed structure between the academic and institutional laboratories, also allows us to distinguish the role of public research organizations such as CNRS[1] or INSERM[2] in the organization and production of collective knowledge. These organizations, which are devoted to basic or fundamental research, provide an undeniable contribution to economic growth (Stephan, 1996).

More specifically, the objective of this work is to highlight the characteristics of the scientific production and quality processes. It focuses on issues related to the characteristics of scientific production and quality namely the determinants of human and financial resources of these phenomena on the one hand and the existing relationships between different elements resulting from the decomposition of these two types of resources on the other hand. The interest of studying these questions is to provide insights in terms of institutional management of research teams and financial resources to promote the production of knowledge and property based output issued from the public research system.

[1] Centre National pour la Recherche Scientifique
[2] Institut National de la Santé de de la Recherche Médicale

In summary, this work focuses on issues related to team and fund composition within research laboratories as determinants of scientific production and quality, on the collaboration between different types of researchers and complementary relationships between different types of funding for public research.

The present thesis is composed of four main chapters. In the first chapter I provide a survey of the literature by recalling some relevant works on the economics of science. The survey is structured by seven different bodies of concepts that help contextualize and set the grounds of my studies on the scientific production and quality. The themes building the literature base provide information on scientific production as a collective process; they focus on hierarchic and reputational issues within organisations as well as the incentives and funding structures in public organisations. Particular interest in is shown on the existence of complementary links between researchers of different status and between different sources of funding and their effects on the output and quality of scientific research laboratories and individual researchers.

I begin by distinguishing the fact that the distribution of the scientific production is positively skewed, in fact, most scientific publications are produced by highly experienced researchers or seniors whereas recent studies show that significant amounts of scientific output comes from collaborations between researchers. Thus, the objective of my research is placed under the frame of and interactive and collective production process.

This framework is defined by several themes reviewed in this literature such as the different research environments, the role of the hierarchical structure within the research groups and the social organization of research laboratories on the one hand and the behaviour of agencies funding research or funding structure laboratories and its effect on scientific production on the other.

The second chapter of the present thesis delivers highly detailed information on the data I used to study the complementary relationships between determinants playing a role on the scientific production and its associated quality. I worked with information provided by a database on public research laboratories within the University Louis Pasteur, a large and well ranked French university known for the excellence of its research.

Within the frame of the present work, I used information related to the composition and production the university research laboratories during the period from 1996 to 2008. In addition, I undertook the task of expanding the existing database at the level of research output variables since the original base only provided information up to 2005. For this purpose, I retrieved the missing information until 2010 in order to exploit the information corresponding to the four-year ministry-university contract of 2008. I also retrieved missing information regarding the impact factors of journals in which researchers affiliated with the University published during the period 1996 - 2010. Finally, I used in-house information on funding laboratories during the period 2001 - 2008, which enabled us to perform the analysis on the impact of financial resources.

The chapter explains the structure of the data, how it is originally broken down into four main sets of information: laboratories, personnel, publications and funding and broadens the picture by providing details on the operations performed on it to obtain merged datasets linking individual publications to personnel and laboratories allowing to study not only the individual, but more important, the laboratory as unit of analysis.

Chapter 3 builds on the work of Carayol and Matt (2006) by studying the determinants of scientific production. They provide further insight on the collective scientific production by focusing on the effects of laboratory composition and funding structure on the output and its quality of researchers not only at the individual level but also and more important the research laboratory as a whole. A first analysis is carried out to assess the impact of human and financial resources on the production and quality of public research in the university laboratories. A second analysis provides a decomposition of these two types of resources according to the rank of researchers within the French academic system and to the type or funding source of funding. This decomposition allows a better understanding of the collective scientific production within public research laboratories funded by public or private capital.

Collective scientific production of research laboratories is measured by the total number of contributions to scientific articles and by the total number of fractional publications produced by all affiliated researchers while quality of the scientific output is measured by the average and the median impact factor of the journals in which affiliated researchers publish.

Results from the analyses carried out in this chapter show that the scientific production in research laboratories is indeed affected by a decomposition of human resources: not only is the class of senior researchers critical of this production but also, and most important, is the class of assistant researchers. This is the main result of my work, which comforts the idea that science is not only collaborative but is stimulated by collaboration between researchers of different levels of experience. On the other hand results at individual level show that scientific production is affected by the status of individual researchers, with senior researchers producing large volumes of output although not necessarily of the always-outstanding quality, that is young researchers are associated with increased levels of scientific quality. In terms of the influence of different types of funding sources, the results show that at the laboratory level, the structure of the available funding does have an impact on the both the scientific production and its quality, with a remarkable finding showing that while public funding stimulates the production of large volumes of scientific publications, it is the private funding that stimulates the production of high-quality publications.

Finally, chapter 4 is based on the study of pairwise complementary relationships between determinants of scientific output and quality. For this purpose it provides a novel application to take account of such relationships based on the theory of supermodularity. Having showed in chapter 3 that the decomposition of human and financial resources for research laboratories has impact on scientific production and quality I proceeded to study the existence of complementary links between the different arguments. The existence of such links may thus define the performance of scientific laboratories and public research organizations and with this idea in mind I adapted the notion of supermodularity of a real-valued function (scientific output and quality) with the objective of revealing the existence of complementarity links between different elements of the decomposition of human resources and funding sources.

The use of the supermodular theory and its application to study the scientific output and quality is uncommon in the current literature but can be useful if the objective of the analysis is to reveal complementary links between different determinants of the function. Thus, an empirical analysis of the supermodular characteristics of the scientific output and quality was performed using two different approaches, the first one a parametric regression method (Athey and Stern, 1996) in which I established the impact of complementary and intermediate states of specific pairs of arguments explaining the this output and its quality; and a second approach, a non-parametric method (Beresteanu, 2005) in which the distance between the different estimators of the complementary and intermediate states of specific pairs of arguments is measured.

The results of the different analyses carried in this chapter show that the scientific production does present complementary relationship between researchers from different classes. For instance, I find there are complementarities between senior and junior researchers as well as between senior researchers and assistants. This result is of particular interest since it points out the existence of a cascade of complementary relationships across researchers of different levels of experience, which is evidence of the existence of research delegation within public research organisations.

Chapter 1

Literature Review

Chapter 1: Literature Review

1 Introduction

Scientific production as a process is characterized by a skewed distribution of output among researchers, with principal investigators producing important amounts of scientific publications and less experienced researchers producing lower amounts of publications. However, given that it may be hard for single individuals to master all the knowledge and information required to produce relevant scientific output, one may wonder whether a complementary link between principal investigators and other types of scientific personnel, such as fellow investigators, assistants, PhD students, and administrative personnel exists and may be accountable for the asymmetries in the distribution of scientific production.

Lotka (1926) showed evidence on this phenomenon when he performed a study on the publication counts of scientists in the fields of physics and chemistry observing that the distribution of scientific publications is positively skewed with the number of scientists signing n contributions equal to $1/n^2$ of those signing only one publication; his main results revealed that 60% of the scientific researchers only account for one publication. This phenomenon is known as the Lotka law on scientific distribution that shows evidence on the irregularities the scientific production.

Moreover, Bradford (1934) studied the citations in peer reviewed journals, and concluded there were “exponentially diminishing returns to extending a search of information in scientific journals” in order to find relevant scientific contributions to a specific topic confirming the existence irregularities in the scientific productivity distribution, where only a few scientists produce high quantities of contributions.

His results showed that as a search on a scientific topic involved additional journals, the probability of finding relevant publications becomes smaller. The combination of Lotka’s law and Bradford’s results provide some insight on the fact that as a special scientific topic is investigated by researchers, only a few of them will produce important amounts of relevant publications and it will always become harder to find new researchers with such high publication counts relevant to that topic.

Today it is commonly acknowledged that the skewedness of scientific production is attributed to differences of scientific fields, age of scientists, cohorts and other characteristics of the scientific personnel, which raises our interests on important issues such as the interactions driving it.

In the present work I study the process of scientific production and the skewedness of its distribution, which combined with the notion of collective scientific efforts among researchers, allows me to believe that there is a relationship between different sectors of the distribution of scientific production has an effect on the scientific output. In fact, those star scientists 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 a “research delegation” or a “collective scientific effort” in scientific production.

I study the notion of complementarities between principal investigators and other members of their research group. Throughout an empirical analysis of the scientific production of research laboratories and individual researchers using data from a set of public research laboratories affiliated to the University Louis Pasteur (today a branch of the University of Strasbourg). I pursue the objective of revealing the effects of the research laboratory composition in terms of human and financial resources on the scientific output.

Based on the theory of supermodularity (Topkis, 1998), I propose an alternative approach to verify the existence of such complementary relationships among different types of scientific personnel. The results provide evidence that supports the existence of such 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.

The present analysis is important because it helps to understand that science is produced through collective efforts. Indeed, research laboratories in our setting may be comparable to Alchian and Demsetz’s (1972) efficient firms in which the products of joint input efforts are higher than the product of separate ones, and where it is possible to measure the marginal contribution of each input. Studying complementarities between different types of researchers may help us develop our comprehension of a scientific production process generated by teams (research laboratories, coauthoring researchers) whose performance depends on the interactions between the individuals in the group.

2 Scientific production as a collective process

It is of general acceptance that the scientific researchers are rewarded in the form of peer recognition according to the magnitude and importance of their research works. Studies have developed the idea that recognition is a requirement for appropriate professional development of scientists, and that highly productive scientists in major universities receive recognition more often than equally productive scientists in less visible universities.

Merton, 1968, conceptualized the process of allocating rewards to scientists for their contribution and analysed how those rewards affect future flows of ideas in their respective scientific field. He studied reward structures and communication systems in science and found out that a principle of cumulative advantages in social sciences concentrates the allocation of resources and rewards around centres of scientific excellence. This cumulative advantage is represented by a St. Matthew effect in the distribution of scientific talent.

These commonly accepted ideas serve as support for the notion that principal investigators who sign several scientific publications and receive important citation counts hold the scientific production process on their shoulders. This issue is the main research problem addressed in the present thesis, where I assume the scientific production process does not only rely on star scientists, but rather on complex and entangled relationships among researchers of different status. I base my assumption on the characteristics of the scientific production, in particular its skewedness and its collective nature.

Studies have confirmed that the distribution of scientific and innovative production is positively skewed. As an example, results obtained by a research conducted by Arora and David (1996) on the determinants of scientific productivity suggest that aggregate publication output of a research unit may vary with the distribution of research grants; additional results suggest that this distribution is skewed to the right since the elasticity of quality-adjusted publications with respect to the budget for a large number of research groups has a value of around 0.6 whereas for some individual researchers, this elasticity approaches the unity. In addition, for a very small number of researchers this elasticity is greater than 1, which implies that the past performance of these individual researchers influences their own future performance and given that these individuals are usually group leaders, their superior performance 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 may be misleading due to the fact that distribution of professors over the number of publications is skewed to the right. Given that scientific productivity follows a non normal distribution, they investigate the differences between academic inventors and other researchers through yearly publication data, and improve it by weighting the citations in the publications.

Further evidence shows that output distributions across scientists are highly skewed; this feature can be associated to a cumulative advantage process raised in the works of Allison and Stewart (1974) who assume that the distribution of scientific productivity becomes increasingly unequal, as a cohort of scientists grows older. Their analysis focuses on the assessment of whether or not the distribution becomes increasingly dispersed as time passes by.

Through simulated time series data stratified according to career and age, such that each strata represents a cohort of scientists across time, Allison and Stewart (1974) studied the life differences and measured scientific productivity in terms of publication counts on 5-year periods and citations on the whole published output. These results suggest there is a general increase in citation inequality as scientists grow older; the fit between scientist's resources and productivity improves over time.

Furthermore, tools have been developed for the purpose of evaluating the asymmetry and skewedness of the scientific production. As example, The *H*-index (Hirsch, 2005) ranks the publications of an author in decreasing order with respect to the number of citations a particular publication has received, then it positions the author in the rank *h* 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 shortfall in the fact that it only values the first *h* citations of the publications; he formulates an extension to the *h* index, the *G*-index, which represents the highest rank *g* for which the articles in the *1* to *g* ranks have at least *g* at the power of two citations.

But scientific research is not only a skewed and distributed process but also an interactive and collective one if it is assumed that innovation processes are based on distributed collective efforts. Based on this assumption I study the performance of scientific groups not within the scope of technological innovation, but rather the scope of scientific production one through the assessment of complementarities between different categories of researchers who perform an integrated and collective work. This study may help understand how science is produced collectively even if it may be hard to disentangle isolated performances within a scientific production process due to the increasing degree of interactions among actors.

Technological paradigms are said 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 progress along a technological trajectory and the emergence of new paradigms. Among different results of his research I may highlight that for current technological paradigms the interactions between researchers and inventors can be represented as an embedded set of relationships and links for which any group performance can hardly be studied in an isolated manner.

I rely on the literature on innovation economics and recall that an innovation process is the result of interactions between firms and organizations that are mutually influenced by pre-existing relations in a coordinated network of agents. Recently, there has been growing interest in joint production processes influenced by an increasing collaboration and the spur of firm alliances. These interactions affect the nature of productive and innovative activities, and generate a loop of knowledge creation that affects the same distributed process at the origin. As a consequence, innovation acts as both an input and an output with a process of distributed productive and innovative activities among several actors.

Coombs and Harvey (2003) studied the fact that innovation processes are often issued form coordinated network relationships between firms and organizations. The authors 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 relationships between firms are formed, stabilized and disrupted.

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

They performed their study based on key issues that are observed in the process of innovation driven by relationships between organizations. These issues deal with 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).

Publication counts are the base of the analysis of scientific research required to understand and evaluate the determinants of scientific production of individual researchers; studies within this frame are a matter of growing interest among economists and policy makers, who usually assess or evaluate the scientific production process based on the achievements of those few scientists who produce considerable amounts of publications. In other words, those star scientists who are located at the right tail of the scientific productivity curve are regarded as being more important than less experienced ones. However, given the nature of scientific production, this process should rather be evaluated based on the collective efforts of different scientists.

Since production and collaboration are strongly related, the question of whether the scientific production process is well understood in the context of a complex and ever growing accumulation of knowledge is raised. Are the connections and collaborations within the process of scientific production the base needed for understanding it better? If it is the case, then we need to better understand the circumstances under which researchers collaborate with each other in order to complement their work and produce collective science.

Several institutions have devoted resources to boost collaboration given the general perception according to which 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 and assess changes in scientific productivity when scientists collaborate. Their headline results show that a substantial change in the regime of collaboration between scientists has taken place around the 1990's, translating into the rapid increase of the average number of coauthors signing scientific publications.

It is widely assumed that collaboration is good per se; several policy programmes have been recently developed with the aim of creating networks of excellence and boosting industry university links. These programmes are supported by the general idea that research collaboration is perfectly understood, behaves in the same manner for any kind of individuals and institutions and can be properly measured and controlled to increase knowledge creation.

Katz and Martin (1997) define, characterize, and classify research collaboration according to three different dimensions:

- The individual;
- The institution;
- The international setting.

They obtained their results by studying and defining research collaborations according to motivations, actors, costs and benefits. Results from their analysis suggests that research collaboration is hard to define, mainly because its definition obeys to social relationships among scientists, leading to difficulties in the identification of the frontier between collaboration and informational links, which is usually blurred.

In addition, the authors state that measuring collaboration merely through coauthor indicators can be misleading since there are cases where close collaboration does not end in a publication and *vice versa* (where weak interactions between scientists result in one of them). They also highlight that due to conceptual problems, a differentiation between inter-individual and inter-institutional collaboration is advisable.

In summary, the discussion about research collaboration connects the notions of skewedness 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 production within research groups and the effects of complementarities between several researchers on the scientific production of a research laboratory.

3 Conceptualizing collective science

The process of scientific production can be explained by several concepts describing the effects of socioeconomic relations and interactions among agents working on a base and a body of knowledge that allows the movement of the scientific, innovative, and technological frontiers.

Among these concepts are found the national innovation systems developed by Lundvall (1992), where the elements and relationships interacting within in the production, diffusion and use of new and economically useful knowledge, are rooted inside the borders of a nation or state. Moreover, a second concept is the Mode2 developed by Gibbons (1994) who puts an emphasis on the applicability of scientific research and stresses the importance of interaction between multidisciplinary groups of researchers in the scientific production process. In addition, the concept of the triple helix developed by Etzkowitz and Leydesdorff (2001) represents three strands of a helix (academic, industrial, governmental) which relate to each other and develops a reflexive overlay of communications, networks, and organizations; as consequence of these imbrications, societal development no longer requires central alignment by a central government.

Finally, an additional concept that attempts to explain these social interactions is the creative knowledge environments. Under this concept, it is the context and the surroundings stimulating the interaction among individuals what generates changes in the institutional barriers on scientific production rather than a differentiation among governmental, academic, and industrial axis, Hemlin (2008). This is due to the fact that interactions among individuals engaged in the production of knowledge and innovations within an organisation are increasing in complexity given the cumulative knowledge needed to produce modern science.

Following the line of the environmental context in which research is produced, it is important to notice the existence of group, size and departmental effects on the scientific productivity. These effects suggest that scientific productivity increases with the size of the research team up to a certain threshold of 4 to 6 members, after which the effect turns over and becomes negative, Andrews (1979).

According to Andrews (1979) research groups must be developed within an environment where clear objectives are established in a coordinated manner. This requires the construction of a genuine research culture with a positive group environment, ensuring the active participation of members of the group within a flat and decentralized organizational structure. In order to develop this research culture, several characteristics such as structured communications, diversity, motivation, skills and reputation are needed.

On the other hand, according to Unsworth and Parker (2003) the factors influencing research and innovation creativity within a collective environment depend on the task and work design defining the optimality of the group structure. The social characteristics of the research group, such as collegial communication, team working, and leadership, as well as the organizational characteristics referring to the culture, human resource practices, and organizational design play an important role in the production of creative complex science.

Within an institutional environment, research organizations shape the capabilities of research groups throughout a relationship between productivity, and reward structures. Research organizations ensure the production of modern and complex science thanks to different factors such as the leadership and autonomy of researchers and the access to appropriate facilities, resources and complementary assets.

Heinze and Shapira (2009) studied the factors that influence the production of research teams by observing the characteristics and patterns of highly creative laboratories in Europe and the USA in the fields of nanotechnology and human genetics. They carried out a survey among 400 European and American experts in the fields of human genetics and nanotechnology with the objective of studying and providing and insight on whether or not there is a predominant contextual pattern for creative events in the scientific research.

Five categories of researchers were studied:

- highly cited scientists;
- academically active scientists;
- industry active scientists;
- journal editors;
- and research programme managers.

Creative activities were categorized in:

- the formulation of new ideas;
- the discovery of natural phenomena;
- the development of methodologies;
- the invention and development of instruments and tools;
- and the synthesis of former dispersed ideas.

Contextual factors were evaluated at different stages in the active life of a researcher emphasising on the strategic choices he/she may make along a scientific career. The creative events in the study were selected upon a parameter basis regarding their research field, organization, geographic location, documentation of the creative event (validated historically and technically), documentation of the preparatory phase (prior to the creative event) and factors related to the research team (institutional and organizational) that performed them.

The headline results of this study show that at the team (research groups around those nominated scientists in the survey), the parameters determining scientific creativity are the size of the group, its composition, its communication patterns, the quality of leadership, and access to external resources. At the organizational level the results showed that the determinants of scientific creativity are the organizational structure, size, centralisation of decision-making, clarity of research goals, funding structure, reputation and visibility of the group.

Moreover, human and financial resources, recruitment processes and leadership also have positive effects on the overall performance of research laboratories. In fact, importance is attached to the recruitment of talented scientists given their impact on the leadership and the trajectory of the laboratory.

In addition, it is worth noticing from this study that small group sizes allow the principal investigator to be active and efficient enough to stimulate members of the team; though on the other hand, larger groups are unable to unleash the whole creative potential of its members because principal investigators spend more time on bureaucracy.

Small groups usually present a flat decision making structure with no hierarchy, therefore stimulating the dynamics of the group. However, negative effects from group size are due to the related increasing hierarchy, bureaucracy, and the reduced leadership. In fact as funding and opportunities appear after major creative events take place negatives impact on the creative process are revealed.

The literature on scientific production has often focused on the analysis of faculty members with the purpose of studying the individual determinants that explain individual productivity in research laboratories. Scholars have focused their interest on studying how those determinants can explain collective scientific productivity in terms of intensity and quality.

Intensity and quality of research activities of peers and close colleagues are beneficial for individual research; not only output and quality of colleagues' research would have a positive impact on individual productivity, but would also reinforce the quality of these same colleagues' future research in a virtuous cycle.

Under the assumption that virtuous cycles in research collaboration exist, to what extent would the performance of a research group be affected by its own characteristics such as the presence of foreign researchers, the access to public and private funding, the channels used to interact, the size of the group, average age, or discipline field.

Answer to these questions are found in Carayol and Matt (2006) who show that full time researchers increase the intensity of their work (frequency of publication) when they are promoted towards a higher rank or status, perhaps due to a greater access to human resources. On the other hand, they are more productive in small labs, perhaps due to lower coordination costs, quicker decision making processes, and less bureaucracy. Their results also point out that the publication intensity of colleagues has a positive impact on the overall group, further results also show there are positive effects from the presence of foreign researchers and public contractual funding in the laboratory.

4 The role of hierarchic structures in research organizations

Shallow or deep organizational forms exist within research institutions. These forms depend on whether researchers work directly under or at the same level of the principal investigator. This assertion, questions whether the optimal organization structure is defined by how well investigators can adapt and articulate their abilities towards a common objective.

The success of research institutions does not only depend on human resources but also on organizational factors. According to Lons and McGinnis (1981, as cited in Leilich, 2005), reseaerch research institutes present pyramidal organizational structures where the principal investigator or director is found at the summit, advanced young researchers at an intermediate trier, and young researchers (PhDs) at the bottom.

The question that rises is whether such an organization may be optimal or not, and whether there are any cases where a flat organization is desired given that different types of organizational structures can be identified and can be judged to be better or worse based on their characteristics regarding the technical and organizational knowledge they embed. According to Mayntz (1985, as cited in Leilich, 2005), there is no optimal organizational form for research institutes.

An nterest of the literature on research management has been the parameters affecting the structure and research performance of research organisations (Laudel, 1999; Fuchs and Oehler, 1994; Mayntz, 1985, as cited in Leilich, 2005), while organisational economics has focused on how such institutions provide optimal incentives for specific investments on semi-public goods (results of the public research) which are subject to transfer, and property rights.

Assuming that assets used within the scientific production process are defined by the experience, network, reputation and access to intellectual resources; the question that rises is to what extent will an individual researcher work with another specific researcher and what would be the optimal incentives that would ensure their collective scientific production.

Across different fields of research, specific investments may affect the organisational structure; for instance, in natural sciences, there are complementary capabilities between production factors, multidisciplinary institutes and little training. These characteristics imply there is a high probability that such organizations present steep hierarchic structures. On the other hand, in the fields of humanities, there are additive capabilities, projects likely to be mono-disciplinary and high training with young researchers often leaving after the completion of their doctoral programs, which implies there is a high probability that these organizations present shallow hierarchic structures.

A study conducted to bring light to these questions was performed by Leilich (2005) and carried out a database of Max Planck research institutes. Within this study, the dependent variable pointed out whether laboratories were subject to steep of shallow hierarchies.

The analysis grouped the institutes according to their field of research in two categories: (natural sciences and humanities) while the characteristic of multidisciplinary was sorted out according to descriptions of the institutes. Training was determined by two variables: the existence of a research school involved with the institution and the percentage of young researchers compared to the total employees of the institute. The results from the study showed that the organizational structure of the research institutes is influenced by:

- the size in terms of employees, with large headcounts implying a steep hierarchy;
- their segmentation into departments, avoiding the need for an intermediate tier and imply a shallow hierarchy;
- and the existence of third-party project funding, which also implies shallow hierarchies.

Furthermore, the analysis finds that the probability of having a steep hierarchy in natural sciences is above average according to the descriptive statistics; In fact, the probability of observing a steep hierarchy structure within an organization increases when the institutes conduct research in the natural sciences with large numbers of trainees. The need for coordination is therefore translated into steep hierarchy structures or externalised towards partner or third triers.

As an overall conclusion of this analysis there is no optimal organizational structure for research institutes, but one can identify structural types that provide an optimal solution according to the environment surrounding them.

Harper (1992) performs an examination of the social organization throughout an investigation of the nature of the working environment in research laboratories in continental Europe, England and the USA. He describes the working environment and the social process that takes place allowing the relations between individuals. His description is based on the values and moral orders allowing the individual to behave correctly within the organisation in order to test ideas socially.

His main results show that common characteristics appear in the hierarchical structure of these laboratories; he observed that some researchers from a specific groups within the laboratories have rights and privileges of mobility and autonomy. These privileges reflect shallow hierarchies resistant to change perhaps due to personal esteem and reputation, hence differentiating research laboratories from other work places.

5 Organizational reputation

According to Fombrun (1996) “reputation is seen as a perceptual representation of the organization's overall appeal compared to other leading organizations”. In general, a firm's reputation is influenced by several factors such as financial performance, size, exposure to media, advertising expenditures and type of industry (Cable and Graham, 2000; Fombrun, 1996). It becomes clear that positive reputation is highly valuable since it sends an informational signal that appeals to highly skilled human resources.

The social identity theory tells us that individual's self-concepts are influenced by the attributes that others may infer about them based on their organizational membership. These attributes allow individuals to be classified into social categories based on group membership. Moreover, the signalling theory tells us that individuals have incomplete information about an organization, and therefore they interpret organizational characteristics, such as reputation, as signals about the organization's working conditions; according to Turban and Cable (2003) it is important to understand how reputation influences decision making given that a positive reputation is a rare, valuable, inimitable, and non-substitutable resource (Barney, 1991) that provides a competitive advantage in terms of attracting talents.

Cravens and Oliver (2003) created of a reputation index that captures some of the organizational components of a firm. They argue that corporate financial statements provide focused data on tangible assets, which ensures reliability but fail to provide information on important intangible assets such as corporate strategy, financial strength and viability, organizational culture, ethics and integrity, governance structures, alliances and innovation; these assets may be significant market-value drivers for firms and among these, the most important assets are perhaps trust and reputation.

Their composite index relies on different measures that allow a representation of the firm in terms of the products, employees, external relationships, innovation and value creation, financial strength, and organizational culture. This representation is in accord with the beliefs key groups such as customers, suppliers, employees and partners have about the firm. As a consequence, the index is suggested as a tool to assess as reputation audit.

Within the literature of human resource management there has been interest on how individuals develop intentions to join specific firms. According to Turban and Cable (2003) a firm's reputation and attractiveness provide competitive advantages in the acquisition of talented human resources. The authors carried out a study that addresses the question of whether the reputation of a given organization depends on the number and quality of applicants who are actually seeking positions. Their main results show that attracting top quality collaborators through competitive recruitment processes is important to any organization given that the attraction of talents increases the utility of the selection process and generates a competitive advantage relative to other organizations.

A firm's reputation is therefore defined as its public evaluation relatively to other firms; this public evaluation of the organisation influences the success in attracting high quality applicants who decide to pursue a job within the organisation based on the overall perception of its reputation. This process reveals a looped mechanism of organisational reputation that influences the intention to apply for a job, which in turn boosts the reputation capital once more.

Two independent studies in different locations of campus job fairs were carried out to measure the firms' reputation in the one hand and the characteristics of the pool of applicant on the other. Controls focused on industry sectors, interviewing dates, and the number of positions available in the firms to test the following two assumptions concerning the close relationship between firm-reputation and the quality of applicants who are interviewed.

- Organizational reputation influences the quality of applicants.
- Organizational reputation is positively related to the number of applicants seeking to be part of the organization.

Headline results indicate that firm reputation is not only correlated with the size of its pool of applicants, but also with the applicants' actual job pursuit behaviour.

According to Rynes (1989) applicants are more likely to pursue job alternatives when the job is perceived positively (high valence) and when the job is seen as attainable (expectancy). Therefore, only highly skilled applicants whose expectations on integrating the organization are important will actually spend time and effort on pursuing the application process in the top rated organizations, and so forth. These asymmetries raise the question of whether the reputation rate of the organization is positively related to the quality of the applicants looking forward to integrate it given that organizations with large pools of applicants can be more selective and therefore the results is the attraction of talented applicant pools.

6 Laboratory funding and its influence on scientific research

Crow and Bozeman (1987) developed a conceptual typology for R&D laboratory classification, and the evaluation of its implications. They highlighted the stereotypes concerning the environmental setting of the laboratory in terms of governmental influence resulting from government funding and they stressed in particular the assumptions about differences between public and private laboratories.

They realized that a classification of R&D organizations according to a triple helix conceptualization of industrial, governmental and academic axes followed by an assumption that their behavioural characteristics are based on their field of research is a limited approach to back science policy implications. For instance, the current connective nature of research and the evolution of laboratories into linked entities render the analysis more complex given that several changes in the organizational forms and environments have taken place and it is highly unlikely that laboratories have been immune to such new shapes of institutional structure.

As a consequence of their argument, they proposed an environmental context presented according to two dimensions: conceptualization and classification. In the first one, laboratories are classified according to ownership, public openness, and R&D market-orientation; in the second one, the laboratories are classified according to the conceptual framework. Their results show there are significant behavioural differences among laboratories in different categories.

Arora and David (1994) focused on the estimation of the production function for scientific research and proposed a structural model for the resource allocation process based on the performance of publicly funded research units. They also estimated the effects of past performance on publication rates to assess how budget allocation influences the reputation of principal investigators and their future levels of public funding.

They stated that even if economists have focused their attention on the links between scientific research and technological progress, little attention has been paid to the determinants of productivity in scientific research: how shifts in inputs and marginal research expenditures across research groups with different characteristics changes the overall scientific output.

Their analysis addresses the determinants of scientific productivity through the development of a structural model that represents the process by which research units receive funding, the model then estimates the corresponding production function of the scientific output, and is implemented on a data set issued form an Italian program in biotechnology and bioinstrumentation of the CNR[3] that took place between 1989 and 1993.

Their results show that 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 the aggregate publication output of the research unit may vary with the distribution of research grants, and is skewed to the right.

During the last couple of decades public research has suffered from cuts in public funding, changes in rationales and an evolving regulatory environment. However, during recent years, public research has also been “rescued” by a growing share of private funding and resources given that a large part of grants and financial resources comes from the private sector in the form of competitive grants and remuneration from science commercialisation. The question this observation raises concerns the determinants that stimulate the attractiveness of public research for private funding.

A study carried by Bouhmadi, Carayol, Llerena (2008) on the scientific production based on the assumption focusing on the laboratory level provides relevant insights on the analysis of the private funding in scientific production allowed the analysis of three principal axis:

[3] Consiglio Nazionale delle Ricerche (CNR); Italy’s National Research Agency.

- The scientific production of research laboratories can evaluated through publication counts.
- Public funding crowds in simultaneously with private funding.
- The characteristics of research laboratories may influence managers and vice-versa.

Their results showed that as public funding does crowd-in private funding in research laboratories, the publication counts of these laboratories have a negative effect on the access to private funding. For instance, an explanation for this result may be that scientific publications tend to be too theoretical while private investors seek applied science and that private contracting incurs in opportunity costs given that researchers need to invest effort and time towards applied research. In conclusion, contractual funding crowds in simultaneous private funding, with a larger share of private funds in active and applied laboratories, and a lesser one in large publishing laboratories.

Nowadays, public research funding is allocated through competitive processes, which stimulate scientific collaboration with the objective to produce modern and complex science through organized research centres, networks of excellence, and interdisciplinary teams (Heinze and Shapira, 2009).

Since funds are usually awarded to scientists with long records of publications, there must be a certain capital of reputation before a creative event can take place. Usually preliminary results are required in order to get complementary funds, while the scientist must cope with funding agencies that operate hot capital by jumping on the "bandwagon" once there is enough attention on a field or line of research.

Moreover, agencies tend to ask for targets or expectations on the results, which are nearly impossible to forecast unless the research isn't that radical in novelty. Such practices jeopardize the renewal of funds, and increase administrative burdens due to higher accountability, hindering flexibility.

Following the initiatives of the Lisbon treaty and its ratification in Barcelona two years later, the EU set the target of raising R&D spending to 3% of the GDP by 2010 with two thirds of this expenditures being made by the business sector. Within this context, Czarnitzki and Frier (2003) investigated R&D collaborations between firms and public research institutions focusing on patent application at the firm level in order to assess whether public R&D spending enlarges the welfare of societies. The analysis required the identification of input-output relationships of publicly funded R&D activities such as patent outcome, product and services commercialization and cost reducing processes.

In their analysis, they proceeded to distinguish three groups of firms: non-collaborating firms, firms collaborating in publicly funded consortiums and firms collaborating with privately funded R&D institutions. They analysed a sample of German firms and estimated their propensity to patent.

Their findings show that collaborating firms are more likely to patent than non-collaborating firms, which implies the existence of knowledge flows generating a positive spill over effect among the partners. Within this group, firms within publicly funded consortia had a higher probability to patent compared to privately funded ones.

7 Incentive structures in public organisations

Bureau and Mougeot (2007) recall there is a link between performance and quality of incentives within the framework of the public administration. This link is quite particular and specific to public organisations given their important degree of multiplicity of their tasks and agents. Weak incentives in the public sector, augmented by the choice of a relatively secure professional career in terms of employment and revenue can be related to a higher degree of risk aversion. Personal objectives of civil servants may differ from those of the general interest because of such incentive structures in public research organisations.

Civil servants have very heterogeneous preferences; the most altruists do not need incentives to excel; they provide the most important part of their effort for the common welfare, while some others will only perform if they are incited by career promotions and monetary rewards. In addition, they evolve within a statutory framework that makes it difficult to set in place a set of good incentives due to administrative, juridical, and informational constraints that slow down the overall performance of public organizations.

All these constraints make it difficult for the policy maker to distinguish the performance among civil servants; therefore, a focused and deep analysis on the incentive structures within the public sector is important in the research agenda.

The study of this problem necessarily implies a characterization of the principal-agent relations in the public sector where the information asymmetries have a dominating role; as an example, problems of moral hazard, anti-selection and exogenous incertitude make it difficult to evaluate the performance of multi task civil servants who may use an informational advantage with the characteristics of a rent in order to reduce their efforts to increase the output and quality of their services.

In addition, the informational asymmetry between the principal and the agents is increased if collusion between agents takes place due to the hierarchic structures where employees and supervisors may share this informational advantage. Hence, according to the authors, the main issue within this framework is to develop effective incentive structures based on improved evaluations of individual agent performance, therefore the public sector, acting as the principal agent, needs to recognize the weakness of monetary incentives and career perspective in public organizations. Stock options and dividends are hardly applicable in the public sector, and the structure should focus on the probabilities of career advancement of civil servants based on their performance.

This implies an incentive structure in the form of future career advancements (Holmstrom, 1982) based on the observable part of the agent’s performance, and the signalling of his future performance by his current efforts, which remains difficult to evaluate due to the multiplicity of tasks these agents perform.

It is also important to recognize the role of group relations in the definition of the incentive structure; in fact, due to complementarities in the production of public goods and services, the individual incentives may lead to free ride problems. Peer evaluation and monetary rewards linked to the performance of the whole group may be the start of a plausible incentive structure.

Dixit (2002) established a link between general theories and empirical case studies on the incentive structures in the public sector related to the theory. He reviewed different issues supervisors and workers face when they develop a professional career in public organizations given the specific organisational characteristics such as the multiplicity of tasks, particular shareholders, and conflicting interests in terms of ends and means across supervisors and workers.

Public organizations present particular characteristics that are not suited by traditional incentive structures such as competition and performance based monetary rewards. In fact, such structures would only be useful in the case individual performance could actually be identified and measured with clarity.

His study was conducted on the education sector through the analysis of an agency in the administration of job training programs. His results show that complex interactions among agents and the multiplicity of tasks in public organizations should be dealt with regarding the organization as a whole instead of a set of several individuals.

Aghion and Tirole (1994) study property rights in an R&D process from an organizational perspective where neither the nature nor the production of the innovation can be agreed upon ex-ante, and therefore the relationships between a research organization and customers who benefit from the innovation are built around the uncertainty of success and delivery of the innovation. These relationships are subject to a particular design of property rights and sharing rules that protects the investments and optimizes the efforts of each party.

Managing innovation processes is an important challenge; theories on endogenous economic growth and intellectual property rights usually assume that the innovation process is ensured by an aggregate agent such as a financer, an innovator, and a user; however, the innovation process takes place as interactions among a multitude of independent agents with different interests that may conflict at the moment of recognition of the effort furnished.

The results from the model show that in the case of single innovations, when the share of capital inputs required by the research organization is larger than the share of intellectual inputs, the R&D process will be ensured by an integrated structure where the control belongs to the customer that uses the innovation, whereas when the share of intellectual capital is larger, the R&D process will be ensured by independent research units implying competition among different research teams for funding.

Scientists produce patents and publications in response to monetary and career advancement incentives, and by producing under incentives they help improve the performance of technology-transfer-offices. Scientific production is therefore an important element towards commercialization of science, local development, and income generation. Within this reasoning, Belenzon (2007) studies how the performance of technology-transfer-offices is affected by monetary incentives.

They argue that technology-transfer-offices allocate their effort into two main tasks: identifying and selecting inventions with commercial potential, and negotiating deals on these inventions. This behaviour introduces an arbitrage between license income and local development given that it is more costly to license in a national or international market than locally, although it provides higher income. Their results state that efforts to produce and commercialize science incur in high costs, and therefore, both, universities with greater local development objectives and universities that are subject to legal and structural constraints are less likely to adopt incentive structures for scientists due to the smaller magnitude of the expected returns.

Moreover, government restrictions and local development objectives hinder the adoption of those incentive payments that can be important for the improvement of technology transfer performance in both private and public institutions. Observations show that stronger local development objectives are costly in terms of forgone income, and that private ownership has a large positive effect on the adoption of incentive structures.

8 Scientific trajectories, the choice to perform science

According to Stern (2004) who carried an empirical study on the relationship between the wage level of research organisations and their field of scientific research, one may define scientists who publish their research agenda as having a preference for science that represents their possibility of accessing recognition for their scientific discoveries and commercial applications.

The relationship between scientists and research organizations is therefore driven by the preference for science on the scientists’ side, and the preference for productivity on the organization’s side. Indeed, since knowledge is a public good, monitoring its production may be costly and as a consequence the reward structure representing the organisation’s preference for productivity is based on priority races and recognition.

According to the extensive literature of the new economics of science, the combination of science and technology may be beneficial to the extent that engaging research on the solution of practical problems raises the quest for new fundamental inquires, Mansfield, (1995). For instance, it is suggested that when technical solutions precede theoretical understanding, technology offers an enormous amount of empirical evidence awaiting scientific explanation. In fact after the opportunity of doing industrial development, scientists who evolve along applied research trajectories find it easier to produce industrial applications than those on focused on fundamental science.

Nevertheless, mingling science and technology may also be potentially dangerous at least for two reasons. First, because the success of a technology depends on the short-term demand and the pursuit of market goals may provide an incentive for the rearrangement of academic research agendas in favour of short-term exploitable trajectories of research. Second, the rules of market competition may not be compatible with the social norms of priority and free circulation of knowledge within the scientific community (Dasgupta and David, 1985; Heller and Eisenberg, 1998).

The traditional assumption that basic and applied sciences are at opposite ends of the agenda is questionable. Shifting the efforts towards merely practical ends would loosen the scientific reward system, and would cause damage to the scientific agenda, goals, and procedural rules. However, the same involvement in practical ends may raise new questions for which fundamental research is necessary.

Different cross sectional studies of the scientific productivity based on scientometric indicators show that the most productive researchers are actually those who patent. Fabrizio and DiMinin (2005) show a positive correlation between actual and lagged numbers of publications and patents. Breschi and Lissoni (2005) show a positive relation between publications and patents by comparing groups of academic inventors and other researchers. Stephan, 2007, performs an analysis on PhD recipients for which the publication and the patent counts are positively related.

In addition, other studies on the influence of patent behaviour on productivity indicate that the event of a patent modifies the natural flow of publications of individual researchers (Azoulay, 2005, Breschi, 2005, Fabrizio, 2005).

Further observations show that a development of sub-disciplines is growing, pushing scientists to choose their fields of interest and specialization at an early age. This choice is often irreversible, and shapes the space of opportunities and costs the scientists faces when engaging in applied research.

Calderini and Franzoni (2007), performed a study to assess how the qualitative characteristics of the work done by a scientist influences his/her patenting behaviour. Using a sample of Italian researchers in the material science field throughout a network of 1276 scientists they control for individual, institutional and environmental characteristics and argue that publications provide reputation, while patents do not provide such rewards given that they lack visibility given that they take place on the closest date to the invention (priority).

In their study, they explain the patent event by the scientific productivity over a three-year moving average of past publications and the basicness of the research composed of the average rank of journal based on citation patterns of the journal.

From the results, the authors state that researchers are at risk of patenting from the very beginning of their career with a change in probability occurring around their 13th year within in the academy (around 36 to 38 year old) implying an inversed u-shaped curve relating patent events to aging effects.

The hypothesis is that scientists face different regimes of opportunities and costs when incurring in applied development depending on the characteristics and trajectories of their career. Two major characteristics are studied: the degree of basicness vs. the degree of appliedness of their research and the importance given by the scientific community in terms of potential impact on future research.

In general, the model evaluated individual hazard to patent depending on both quantity of publications and other institutional characteristics, going from scientific discovery to applied development and vice versa. The results showed that every increase in the degree of applied science increased the probability of patenting, while for those scientists increasing their productivity in high impact fundamental science, the probability of patenting decreased

In terms of policy-making it is important to understand that industrial applications rise from very productive researchers, therefore there is a need for a critical mass of research to sustain future generation of industrial applications. In fact, good applications rise from high impact science, therefore incentives should be specific to the discipline fields.

Breschi and Lissoni (2007) studied whether there is a trade-off between patenting and publishing (basic or applied research), and whether productivity and quality change accordingly. Their results show that although patenting per se does not enhance productivity, it enhances the relations with the industrial world.

There used to be a general perception that commercialization restricts the production and flow of ideas and therefore, the authors question whether patenting activities distract scientists from doing research. They studied scientific productivity along the professional career of scientists by assessing the impact of each of their patents on their publication record and according to their results; there are publication delay effects for as long as the patent has not been filed.

These delays are imposed by the secret nature of discoveries until the patent is approved and may represent a distraction from basic to applied research. Although, by focusing on applied research, the publication rates of individual scientists in refereed journals are subject to delay effects until the application is filed. The two incentive structures: Rewards and IPR's imply a level of secrecy that distracts researchers from performing basic science. These two incentive structures present diverging interests; in fact, rewards encourage the disclosure of data and codified information while IPR's encourage secrecy.

According to Dasgupta and David (1994) IPR's encourage incomplete or selective disclosure. Patent intensive firms tend to rely on secrecy in order to capture the returns on non-patentable assets acquired along the development process, which constraints scientists devoted to applied research to keep secrecy on an important part of the results.

Patenting activity also has a positive impact on scientific resources and individual productivity; it increases the publishing rates around the patent application due to additional resources devoted to the specific patentable research. However, increased productivity effects can be difficult to differentiate from delay effects since there are increasing publication rates after the patent application. Some institutions may be willing to delay their publications in exchange of additional resources.

Among these observations on patenting and publishing one may realize that some inventors are productive scientists; their productivity increases in the years immediately after a non-occasional patent and in some cases a few years before. In addition, patenting activity generates resource effects and stimulates individual productivity throughout links with the publishing activity, while persistent academic inventors appear to be more productive than occasional ones.

These results suggest that it is not the patenting activity per se that stimulates productivity, but rather the links it generates with the industry; in fact the expected returns from licensing may be less than the opportunity costs of putting relations with the industry at risk, therefore, legislators should encourage joint projects with the industry instead of pushing scientists and academy towards IPR's.

Societies expect useful fundamental science that can be applied in the private sector; in order to ensure this transfer of knowledge, IPR structures are set in place to enable the commercialization of knowledge in exchange of financial resources. Developing technology transfer structures does not diminish the importance of the scientist in the process of commercialization.

Taking a closer look at the process, one may realize that scientists are usually involved in entrepreneurial activities such as spin-offs and start-ups. Lowe and Gonzalez-Brambila (2007) carried out some research on the topic of technology transfer and scientific output of spin-off founders that questions whether these activities create a distraction in the process of scientific production or if they increase secrecy. In addition, a study carried by Buenstorf (2009) takes a look at whether research and commercialization are competing or complementary. He studies the commercialization level, academic inventors and spin-off creators rather than merely inventors for the academy. A first observation is that patent-publication pairs are not uncommon, which contradicts the misconceived general *apriori* that fundamental research and commercialization are competing activities, whereas in fact, these activities are complementary and not substitutes as Fabrizio and DiMinin (2008) had stated based on their analysis on a panel data set on academic researchers.

Academic inventions per se are not very valuable in the market; they depend heavily on the tacit knowledge of the inventor who must engage in entrepreneurial activities to see his invention succeed. This process generally brings issues such as delays in publication, incentives for secrecy, and divergence from the initial interest in basic science towards applied and industrial science.

The links and interactions between the academic and the industrial world represent the positive effects from the commercialisation activity (Stephan et al, 2007), although when taking a closer look at the role of talent and ambition, one may observe that interactions between researchers and the private sector provide them with learning opportunities, which in turn may have an increasing effect on their own research output. In addition, financial payoffs and income flows are generated, not only for the scientist but also for the laboratory.

According to Agrawal and Henderson (2002) prior empirical studies on academic patenting and its relation to individual scientific output based on US data suggests that the scientific output of academic inventors does not increase in quantity but quality and relevance, while Stephan (2007) states the inverse, meaning that academic inventors publish more and get more citations implying an increase in quantity and relevance of research. Other similar conclusions have been obtained for European studies that show that patent ownership is a weak measure of academic patenting (Breschi-Lissoni-2007).

A study was carried out matching spin-offs created by Max Planck Institute directors with the list of inventions made within the institution. Publication and citation counts were established to measure personal research output (which are usually highly skewed), while cross sectional heterogeneity amongst researchers was tested through a fixed effect specification. The cross sectional analysis suggests that inventions and commercialization activities are associated to above average scientific performance.

The econometric analysis models out a researcher's output as a function of his invention and commercialization activities in previous periods. The results show that publication figures after the invention takes place are significantly higher, whereas citation figures are not; in fact, these figures fall when the professor is listed as a spin-off founder. Therefore, the entrepreneurial activity appears to come with a cost in terms of publications and quality since the perceived resources from commercialization do not boost output.

In summary, the analysis carried on the relation between invention and commercialization of scientific output on the one hand and research productivity on the other shows that invention activities do not incur in a decrease of scientific productivity. The crowding out hypothesis on the scientific resources is rejected and the compatibility between these different activities is proven. In addition, spin-off activities are not as clear in their implications, since permanent treatment specifications suggest that in the long run spin-off activity may incur in a loss of performance, this result being counter intuitive, contrasts with the earlier results of Gonzalez-Brambila (2007).

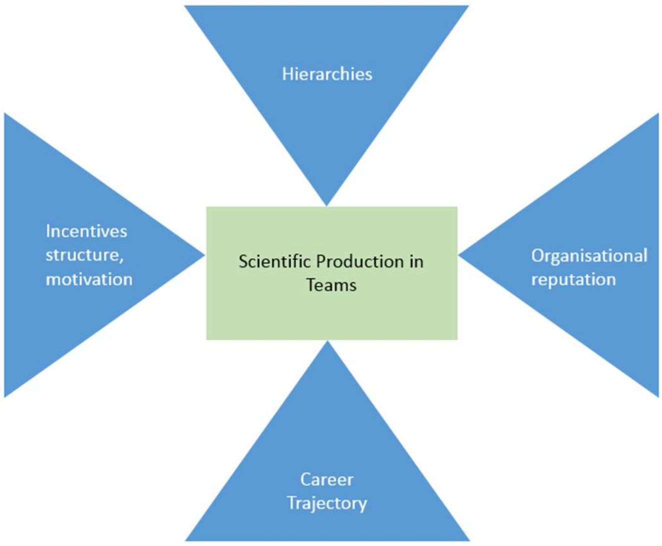
The closer scientists are to the private sector, the more efforts are required by their entrepreneurial activity, the more negative effects on the research output. In fact learning effects do not account for performance.

9 Conclusion

The present chapter provided a broad literature review that gathers elements from different fields studied within the production of science, knowledge and innovation. These elements constitute a set of tools that allows understanding better the production process of collective science in research laboratories and in research teams.

The main elements described throughout this literature review form an interactive system. This system may be interpreted as a particular case of the theory of the firm framework developed by Alchian and Demsetz (1972) in which cooperation between agents and production in teams is studied highlighting the roles of incentive structures, hierarchies, motivation and control structures.

Figure 1 Representation of elements interacting in the production of collective science



However, the difference between the firm and the research laboratory is that in the case of scientific production in teams within research laboratories, the actual contribution of members of the team can be distinguished and observed, which is rather difficult in the case of production in teams within the firm.

In the next chapters, the present research work will focus on the collective scientific production in research laboratories (green box in the previous figure) with particular interest in their composition, funding and interactions between researchers and types of funding.

Chapter 2

Data: Evidence from the University Louis Pasteur

Chapter 2: Data, evidence from the University Louis Pasteur

1 Introduction

The former University Louis Pasteur (ULP), is an important actor among European higher education institutions (HEI's). Described as the 11th European university in terms of impact by the third European report on science and technology indicators (2003), it presents a long dating tradition of excellence in both fundamental and applied research focused on three different axes of life sciences, matter sciences and humanities[4] building on an industrial tissue composed of a network of research and technology transfer. Nowadays part of the second most international university in France (University of Strasbourg), the ULP also benefits from the important international visibility since several thousands of its students are foreign (19.8%), and has several research laureates among which two Nobel prizes working in the scientific disciplines.

For the purpose of the present research on scientific production, we gathered an important amount of information by means of a documentary research axed around different and reliable sources[5]. In the first place, we used an existing database providing detailed information on the research activities of the ULP. This information is mainly built around two important axes:

- Detailed information on the research personnel of the university.
- Detailed information on its research laboratories.

The information recorded in this database are part of the 4-year ministry surveys[6] that academic laboratories 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 laboratories within the academy, and ensure their

[4] This disciplinary field decomposition follows the nomenclature used by the Observatoire des Sciences et des Techniques (OST).
[5] These sources were institutional surveys on the research activity of the university laboratories, public records on scientific publications, and former studies on the funding sources of research carried at BETA (Bureau d'Economie Théorique et Appliquée), Strasbourg.
[6] Nowadays the ministry carries the compulsory survey on a 5-year basis starting on 2012.

survival subject to the ministry's decision making; given this fact, we may regard the information provided by this database as a reliable source of institutional data on the scientific research groups of the university.

On the second place, we used information on the scientific output of researchers associated to the university over a long period of time (1994 – 2010); this information was collected using ISI Web Of Knowledge records on scientific publications.

In addition, we used information on the research funding sources of the ULP, which was originally gathered at BETA by means of a survey on a population of all research laboratories of the ULP performed with the objective to observe every single funding endowment received over a the period 2001 – 2008.

Carrying on with our description, these sources of information upon which the database holds represent detailed records on the research activity of the university. They were matched in order to establish a relationship between the resources of the laboratories (human and financial) and their scientific output (publications, patents and industrial contracts).

Additional sources of information were two databases from the statistics office of the university concerning researchers habilitated to direct research on the one hand, and the doctoral students on the other, as well as complementary information drawn from the Internet and other research reports.

Using the information collected from the different data sources described above, we proceeded to merge it 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 laboratories on the one hand, and their scientific output (publications) on the other.

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.

We may state now that these variables used in the analysis obey to the broad categorization found in Slipersaeter (2007), illustrating different sets of variables found in the analyses of higher education institutions (HEI). Projects AQUAMETH and CHINC[7] reflect a base for the analysis of the landscape of HEI's in Europe, focusing on information regarding the strategies of governance, efficiency and productivity of HEI's during a 10 year period in 11 to 22 countries.

The analyses of variables used in these projects revealed six broad areas of variables used in such analyses. We believe we can cite some these categories as representation of the information used in our analysis of public research institutions.

- General information on the institution: Year of foundation, address, composition, disciplinary field, governance, and other historical information.
- Revenues: Total resources, types of funding.
- Expenditures Total operability expenditures.
- Personnel: Total staff (academic and administrative).
- Educational production: Ratios of total students to total awarded degrees.
- Research and technology production: ISI publications, technological indicators, granted and filed patents.

2 Datasets

2.1 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).

[7] Project AQUAMETH was developed by the PRIME network of excellence (see http://www.concours-urbanartparty.fr) under the European Commission 6th Framework Program. It built a comparative database of 487 universities in 11 European countries. The CHINC project on the changes in institutional funding and their consequences was developed under contract with the European Commission, and studied a feasible data collection system with the objective of monitoring European higher education institution.

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 laboratory (in full time activity ratios).
- The number of researchers affiliated to several non-academic public research 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 all other administrative staff, executing non–researcher tasks) per laboratory and institutional affiliation.
- The total number of researchers holding the *habilitation* or right to direct scientific research per research group and institutional affiliation.

These research laboratories are distributed across different disciplinary fields following the notation of the Observatoire des Sciences et des Techniques (OST); the three main scientific disciplines are life science, matter science, and humanities. According to the OST, these disciplines are divided into subfields corresponding to fundamental biology, medical research and applied biology for the life sciences; chemistry, physics, space science, engineering and mathematics for the matter sciences; and finally social and human sciences for humanities.

The four different periods covered by the ministry surveys show a distribution of university research groups across disciplines as follows: in the life sciences, which represent 63.7% of total laboratories in 1996, and decrease to 42.2% of them in 2008, we find the subfields of fundamental biology accounting for around 26% of the laboratories with little variation in time (in tenths of percent), medical research is steady around 30% of them between 1996 and 2004, then decreases to 11.1% in 2008, and applied biology, which oscillates around 4% from 1996 to 2004, and then drops to 2.2% of laboratories in 2008. We may notice here the important weight of life science laboratories in the university, and providing insights on the determinant role of laboratories in this discipline and their influence on the university's scientific output.

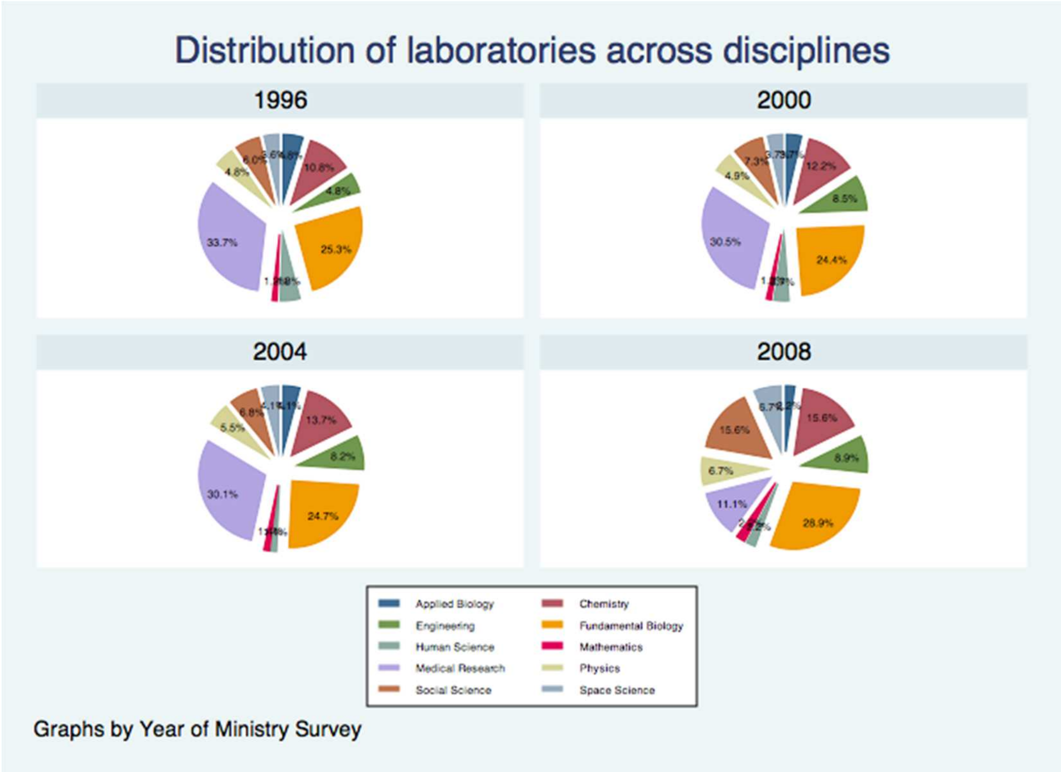


Figure 2 Distribution of laboratories

Laboratories in the matter sciences represent 25.3% of the university in 1996, and grow up to 30.5% in 2000, 32.9% in 2004 and 40% in 2008, within this class of laboratories we find the chemistry field steadily increasing from 10.8% of the labs in 1996 up 15.5% in 2008 with a variation of nearly 2% from period to period. We may notice that from the general field of matter sciences, chemistry is the subfield that presents the highest weight of laboratories. We then find the subfields of physics represented by around between 5% and 6% of the laboratories over the whole period with little variation in time, space science ranging from 3.6% to 6.7%, engineering starting from less than 5% of laboratories in 1996, and increasing up to almost 9% in 2008, and mathematics with the lowest share of laboratories going from 1.2% in 1996 to 2004, although almost doubling its share to 2.2% in 2008.

The share of social and human sciences represents around 10.8% of the laboratories in 1996 and 2000, 8.2% in 2004, and grow up to a share of 17.8% of laboratories in 2008 making up for the lowest weight of representation in the university, within this class we find the subfield of human science representing around 3% of laboratories in 1996 and 2000, decreasing to 1.3% and 2.2% in 2004 and 2008 respectively, and the subfield of social science accounting for 6 to 7% of the laboratories from 1996 to 2004, and 15.5% in 2008.

If we focus on the personnel working at the research laboratories, we may realize there are two different types of scientific workers, those executing research and/or teaching activities, and those executing administrative only tasks. The distribution of these personnel across disciplinary fields shows that those research laboratories in the field of life sciences gathered around 45% of the research/teaching personnel steadily from 1996 to 2004 and then dropped to 38% in 2008 while those laboratories in the matter sciences gathered as well around 45% of this personnel during the whole period of our documentary research. In addition, laboratories in the field of social and human sciences gathered around 9% of the research/teaching personnel from 1996 to 2004 increasing to 15% in 2008. If we look at this variable regardless of the disciplinary field, we may notice it presents an average of 17 to 19 and a median of 11 to 13 individuals per laboratory in the first three surveys, increasing to an average of 37 and a median of 30 individuals in the 2008 survey.

Moreover, the distribution of administrative personnel across disciplinary fields is nearly steady during all four periods (1996 – 2008) and for all three general fields with 52 to 54% of the administrative staff working in those laboratories in the life sciences, 42 to 44% working in matter science laboratories and finally 2 to 3.5% of them working in social and human science laboratories. The administrative staff presents an average of 13 to 15 and a median of 6 to 7 individuals per laboratory during the period 1996 – 2004; these figures increase by more than double attaining an average of 37 and a median of 23 individuals per laboratory in the 2008 survey.

Given the mixed nature of laboratories in the French research system where their structure mixes the academy with other public research institutions. Focusing on the major non-academic institution, the CNRS, we may comment hereafter the distribution of the institutional affiliation through the percentage of researchers affiliated to either of them. The information declared in our documentary surveys shows that among all those researchers affiliated to the CNRS, those laboratories in the field of life sciences held around 30 to 41% of them steadily during the whole period from 1996 to 2008. When looking at the field of matter sciences, we realize this percentage presents an slightly decreasing trend across the first three surveys, starting at 58% in 1996 and reaching 54% in 2004 before growing back to 56% in the 2008 survey; finally, in the laboratories in the field of human and social sciences held a share of 2.6% researchers affiliated to the CNRS in the 1996 survey which decreases down to 1.6% in the 2008 survey. Now if we consider those researchers affiliated to another important non-academic public research institution, the INSERM, we may realize that life sciences laboratories hold the biggest share of these researchers, around 100% in surveys 1996 2004 and 2008, while 89% in survey 2000). This is expected since the INSERM deals with medical research and therefore the associated laboratories may be those in the field of life sciences.

If we take a look at the distribution of total researchers affiliated to the main institutional research organisms without any disciplinary distinction we may find that CNRS holds an average of 20 researchers and a median of 10 researchers steadily from the surveys of 1996 to 2004 and then holds up to an average of 46 and a median of 33 researchers in the 2008 survey. On the other hand, the INSERM holds during all four surveys a median of 3 to 4 researchers and an average of 6 researchers during the first three surveys (1996 – 2004), growing up to an average of 17 in 2008.

Finally, if we take a look at those researchers with no affiliation to a non-academic research institution – or who are only affiliated to the university – we may find they were gathered at 42% by laboratories in the life sciences during the1996 survey, with that share fluctuating up to 44% in 2000, then 39% in 2004 and decreasing by almost 10% down to 39% in 2008. In addition, those laboratories in the matter sciences held a hare of nearly 43% of these academic researchers in 1996 with an increasing trend reaching 48% of them in 2008, while those laboratories in the social and human sciences, increased form 13% in 1996 to 20% of academic researchers in 2008.

The evolution of this variable shows there were 11 to 12 academic researchers per laboratory in average and a median of 8 to 9 declared in the surveys of 1996 and 2000, with an increasing trend in these figures showing a mean of 15 and a median of 11 academic researchers in 2004 and almost doubling up to a man of 31 and a median of 22 academic researchers in 2008.

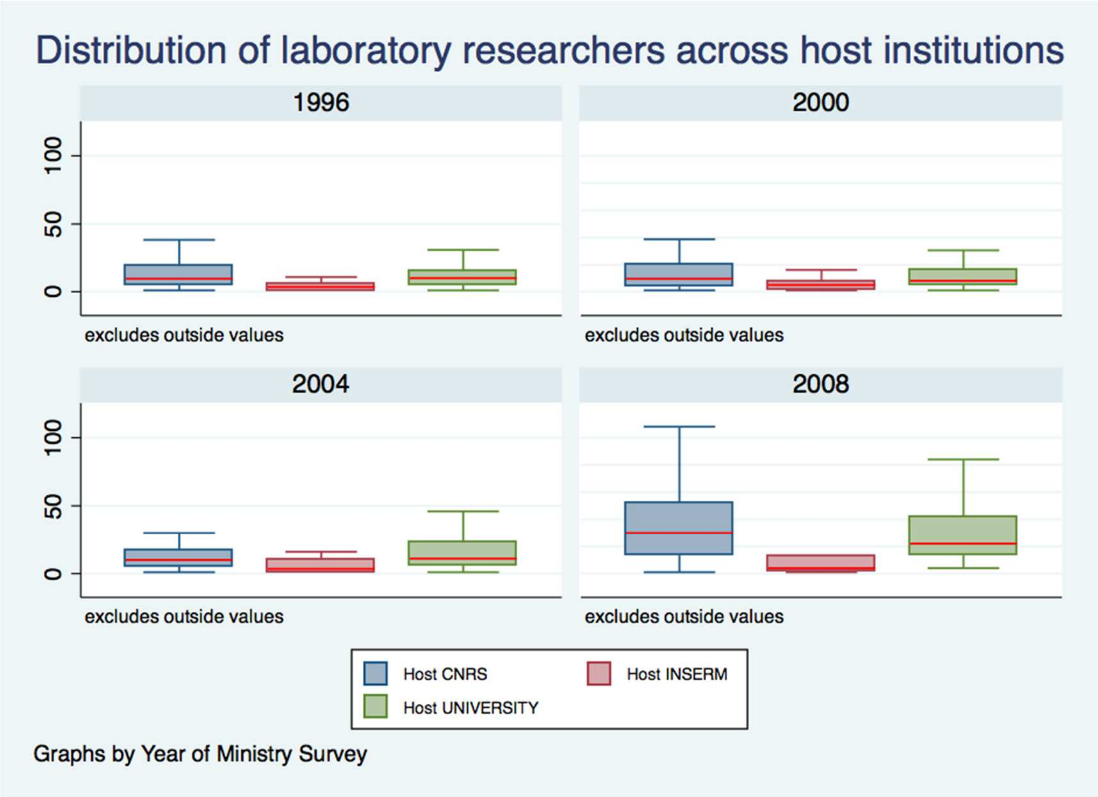


Figure 3 Distribution of institutional affiliation (in numbers of researchers)

The present dataset also provides information on the number of PhD candidates associated to the laboratory at the moment of the surveys. Their distribution across disciplinary fields shows similar figures as in the case of other researchers described above. Laboratories in the field of life sciences gather 39% of all PhD candidates associated to the university in 1996 increasing up to 42% in 2000 and 43% in 2004 before dropping back to 38% in 2008. The field of matter sciences presents a decreasing then increasing trend, with nearly 46% of PhD candidates in 1996 down to about 42% in 2000 and 2004 before settling at 49% in the 2008. Finally, laboratories in the social and human sciences only represent around 15 to 16% of PhD candidates in the first two surveys, and 12 to 13% in the latter two surveys.

In addition, we may also describe the distribution of the number of defended PhD theses during a survey period. Research laboratories in the life sciences gathered 40% of them in 1996 increasing up to 44% in 2004 and dropping back to 42% in 2008, while the laboratories in the matter sciences gathered 53% of the defended theses in 1996, decreasing to 46% of them in 2008; the social and human sciences gathered on the other hand 6% of the defended theses in 1996, fluctuating until reaching 10% in 2008. If consider the distribution in time of these two variables, regardless of the disciplinary field of the laboratories, we find in the first place, that during the first three surveys (1996 to 2004) there is an average of 13 to 15 PhD candidates per laboratory and a median of 8 to 10 of them while those figures almost double in the survey 2008 growing up to an average of 32 and a median of 23 PhD candidates per laboratory. In the second place, we find that the number of defended theses also represent an average of 13 to 14 and a median of 6 to 8 of theses in the first three surveys, and grow up to a mean of 26 and a median of 18 theses in 2008. The behaviour of these variables is expected, and we may underline the fact the median of defended theses is lower than the number of PhD candidates during the first three surveys.

This observation led us to use those two variables in the construction of an indicator translating the notion of the capacity of research laboratories to transform simple PhD candidates into actual doctors during a given period of time. Such indicator is defined by the ratio of number defended theses declared in each of the surveys over the number of local PhD candidates declared in the same survey. We may suppose that the higher ratio of theses per PhD is, the more likely the laboratory is capable of turning students into researchers.

2.2 Dataset “Personnel”

The second data list in the structure deals with information about the personnel of the university. It covers the period 1996-2004 and provides extensive information on the ULP’s scientific personnel which accounts for 1451 individuals in 1996, 1426 in 2000, 1434 in 2004, and 1637 in 2008. The following is a brief description of some of the important variables it contains:

- The research group code, which allows us to establish a link between the individual researcher and the laboratory he is affiliated to.
- The hierarchic rank of the researcher, which reflects the administrative status of the researcher in the database as stated in the ministry survey. This variable serves as an indicator related to a notion of “juridical status” regarding the qualifications and position of the researcher within the French academic system.

- The research institution to which the researcher is affiliated; this variable may correspond either to the University Louis Pasteur, the CNRS[8], the INSERM[9], the INRA[10], 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.
- The CNU[11] 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 direct 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 laboratory directors for the ministry surveys, as well as relevant information on their age and their experience at the laboratory.

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

These individuals are the same ones declared in the laboratory data list, and their sum over a specific research group should match the number of researchers indicated in that data list. However, taking a closer look at the data these two variables are not the same, although 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 the decomposing these researcher into different “researcher types”, which would have not been possible otherwise.

[8] Centre National de la Recherche Scientifique.
[9] Institut National de la Santé et de la Recherche Médicale.
[10] Institut National de la Recherche Agronomique.
[11] Conseil National des Universités.

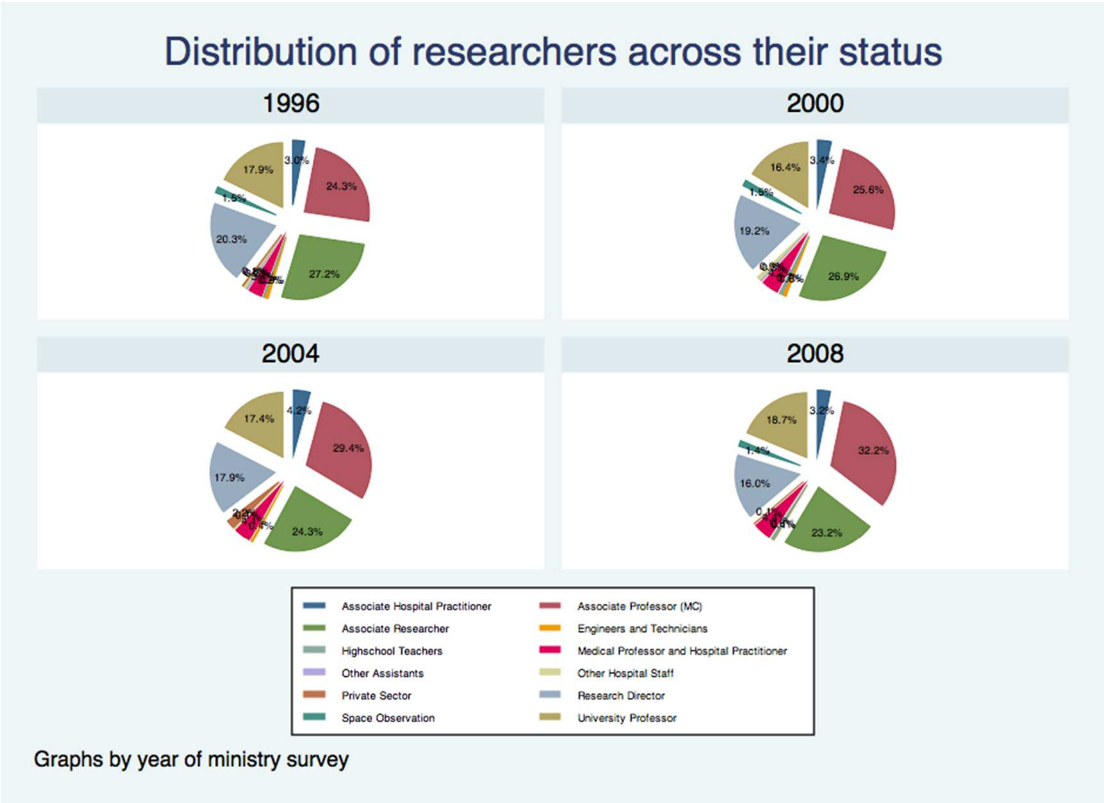


Figure 4 Distribution of researchers

As depicted in Figure 4, the distribution of researchers according to their rank or status within the scientific laboratory suggests that an important percentage of them falls into the most experienced scientific categories, senior and junior researchers, composed of the following ranks: directors, professors, and associates. In fact, the share of research directors ranges from 20.2% of the personnel of the university in 1996 to 16.5% in 2008 slowly decreasing of nearly 1.5% from period to period, the share of university professors is steady at around 17% across all 4 surveys, while the share of medical professors and practitioners ranges from 3.4% to 4.3%. These three ranks account for the senior personnel of the university, which represents around 40% of it along all four surveys.

In addition, the share of associate professors represents 24.4% of the personnel in 1996 and grows to 32.2% in 2008; associate practitioners oscillate around 3% and 4% of the personnel during the same period, and associate researchers start at 27.2% in 1996 and decrease to 23.2% in 2008. These three ranks constitute the category of junior researchers whose share ranges from 54.5% of the personnel in 1996 to 58.6% in 2008.

Finally, the remaining ranks: space observation assistants, assistants, private sector staff, and schoolteachers, account for around 4% of the personnel in 1996 and 2000, and drop to around 2.4% in the following periods; they constitute the less represented category of scientific personnel affiliated to the university.

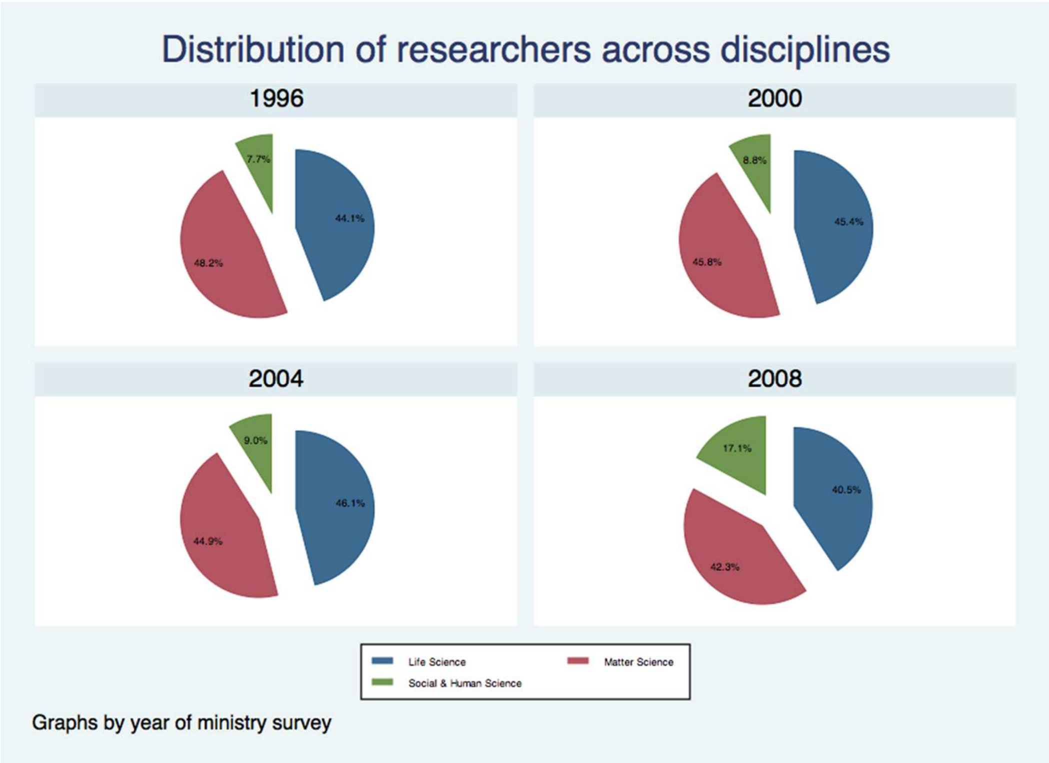


Figure 5 Distribution of researchers across main disciplines

The distribution of researchers across disciplines shows that the largest shares of them work in the life and the matter sciences. As a matter of fact, researchers in the life sciences account for around 45% of the population from 1996 to 2004, and then drop to 40% in 2008; researchers in the matter sciences start at 48% of the population in 1996, and drop to 42% in 2008, while those researchers in human and social sciences account for around 8% of the population in 1996, and grow up to 17% of the population in 2008 recording the highest share increase over the whole period.

A similar pattern of figures is observed when we decompose our researcher into three main categories, senior, junior and assistant, and when we take a deeper look at their distribution across the general disciplinary field, we observe that 45% of all senior researchers were affiliated to life science laboratories in 1996, growing up to 48% in 2004 before settling back at 44% in 2008 while slightly higher shares, 47% to 49% of them were affiliated to matter science laboratories during the period. As expected, due to the low representation of social and human science laboratories at the ULP, we observe that 7% of senior researcher worked for these laboratories, fluctuating across time and settling at 5.4% in 2008. In addition, we may state here that the distribution of the junior researchers across disciplinary fields in time presents a rather similar pattern than that of senior researchers.

Let us now describe the personnel dataset in terms of the most important institutional filiations of researchers working at a laboratory of the university. The most represented institutions are the CNRS and the University, which confirms our previous observation in the laboratory dataset. These two institutions fluctuate over the period 1996 – 2008 within range of 36% to 41% for the CNRS and 46% to 56% of the population of researchers in the case of the University.

In more detail, most of the CNRS researchers work in the matter science laboratories, with a share fluctuating in time within the range of 46% to 53%, while the second largest share in time (43% to 52%) works in the life science laboratories. Only a small share of researchers affiliated to the CNRS work in the social and human science laboratories (1.2% to 2.7%). In addition, among those researchers solely affiliated to the University, we observe that most of them work in the matter science laboratories, with a share fluctuating in time within the range of 47.7% to 58% while the share of those working in the life sciences fluctuates between 29% and 38%. As expected, due to the low quantity of laboratories in the life sciences, only a range 12% to 14.5% of these researchers works in this academic field. Finally, looking at the institutional affiliation to the INSERM, we observe that the share of total researchers affiliated to this organism fluctuates within the range of 5.6% to 6.6%, with the majority of them working in life science laboratories as expected due to the mission of the INSERM, which is dealing with medical and health research.

After discussing the representation of different ranks/types of researchers and their institutional affiliation, we may describe their age and experience. This information is provided in the personnel dataset in the form of date of birth and date of entry in the laboratory for all researchers declared in each of the surveys. In average, the age of researchers working at the laboratory varies between 45 and 47 years during the period covered by our surveys, with a median age of 46 to 47 over the same period, which denotes some homogeneity of age among researchers. These figures remain almost unchanged when we look at the average age and the median age of researchers across disciplinary fields, both ranging among 45 and 47 years for all three of them. In the case of individual experience, we may observe an average experience in the laboratory fluctuating in time between 8 and 10 years, and a median experience fluctuating between 6 and 8 years; these figures also remain almost unchanged when we take a look at the experience in the laboratory across disciplinary fields, which ranges between an average of 8 to 10 years; however, we may note that for the survey of 2008, we observe a median experience of 3 years in life science laboratories while all other disciplines and periods present a median of 5 to 8 years of experience. We may therefore believe that life science laboratories went through a process of renewal of researchers along the period 2004 – 2008.

Regarding the evolution of affiliated researchers over the 4-year periods marked by the ministry surveys, we observed small changes during the period 1996 – 2004 corresponding to the first three surveys, while a higher variation was observed between the last two surveys corresponding to the period 2004 – 2008. In fact, from the 1451 researchers affiliated in 1996, 317 individuals were no longer listed as part of the personnel in 2000, while 292 new entries appeared indicating a net decrease of 25 affiliated researchers, or a net research personnel variation rate of -1.72% in between the two ministry surveys of 1996 and 2000. Between the period 2000 and 2004 there is a net research personnel variation rate of +0.56% accounting for a total of 398 leaves and 406 new entries. Finally, the change in between the 2004 and 2008 periods takes into account a net research personnel variation rate of +14.15% representing a total of 346 leaves and 549 new entries of affiliated researchers. This last observation confirms our previous belief on the renewal of research personnel inferred from our observation of a low median experience in life science laboratories revealed in the survey of 2008.

In addition to the variables concerning researchers and administrative personnel at the laboratory level, we carried our decomposition of scientific personnel into five types 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. This decomposition will later be useful for our analyses.

2.3 Dataset "Publications"

The third part of the structure is the Publications data list, which provides 81412 contributions to publications made by personnel attached to the university. These contributions cover the period 1988 – 2008, and were gathered through at least 3 different waves of data extractions from the Web Of Knowledge (formerly ISI Web Of Science). The first two waves of data collection were done according to a first method of extraction, matching ULP authors to their contributions (period 1988 – 2005).

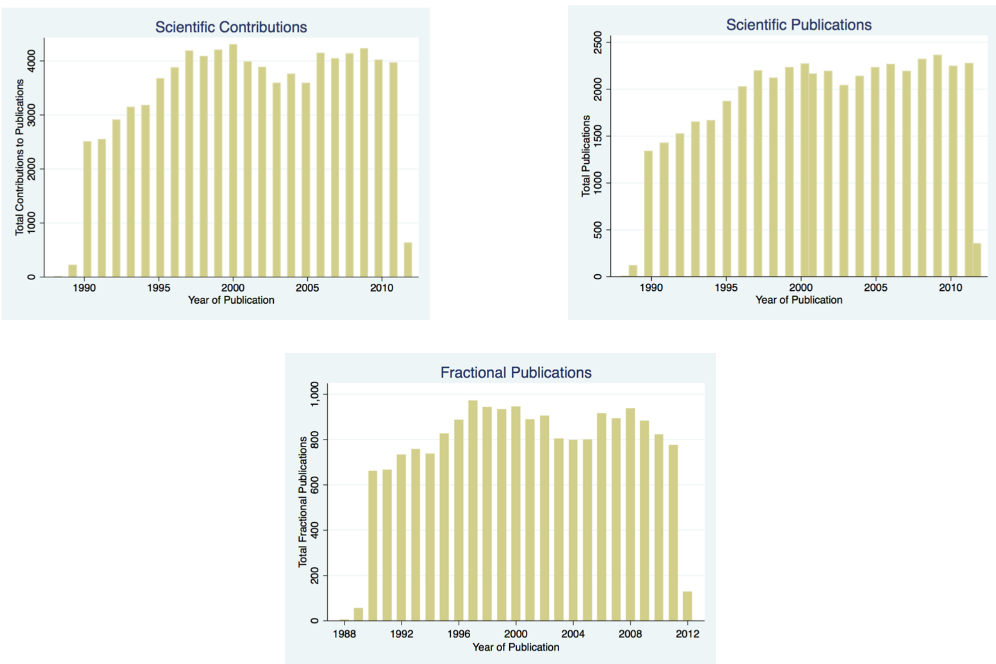
This first and most exhaustive extraction covers the period January 1990 – July 2002 (although some entries from 1988 and 1989 are present), and contains 43241 single contributions to published articles (one entry for each coauthor attached to the ULP), while the second wave cover the period August 2002 – mid 2005, and contains 12108 single contributions. The data was gathered by searching inside the ISI Web Of Science; each researcher member of the ULP as recorded in the Personnel data list following the structure *SURNAME + I** (where "I" denotes the initial of the name) in the author-field and controlling his association to the university by searching *STRASBOURG* or ULP or ILLKIRCH* or SCHILTIGHEIM or WISSEMBOURG or CRONENBOURG or COLMAR or HAGENAU* in the address-field.

The third wave of data collection used a second method of extraction matching contributions from a site of the university to authors recorded in the ULP personnel database. This wave of extraction answered the need to complete the information on the ULP publications to match the latest ministry survey of 2008.

To complete the Publications dataset for the period of mid 2005 – 2008 we built a sample of contributions directly extracted from the Web Of Knowledge, and based on general searches pointing out a site of the university using the keywords *STRASBOURG* or ULP or ILLKIRCH* or SCHILTIGHEIM or WISSEMBOURG or CRONENBOURG or COLMAR or HAGENAU* in the address-field of the articles for the period of mid 2005 – 2012; all these keywords representing the cities in which the university has established a site.

We were able to gather 29086 entries for which all coauthors were mentioned, and then match them with the members of the university stated in the Personnel data list by ministry survey up to 2008. This was possible given that the personnel from the university is recorded in the form *SURNAME_NAME* and can be associated to a certain laboratory during a certain period of time covered by the ministry surveys, while the authors in the records extracted from Web Of Science can only be expressed according to the form *SURNAME_I* (where "I" denotes the initial of the name) and are associated to a certain year.

Once the newly collected data was appended to the previous set, we obtained a total of list 84435 individual contributions to scientific peer reviewed publications performed by researchers affiliated to a research laboratory of the University Louis Pasteur. With this complete set of individual contributions we were able to establish a count for the total publications by collapsing those individual contributions over their publication identification code, therefore obtaining a count of 45196 publications in which at least one or more members of the university contributed to. We also established number of fractional publications by weighing the publications with their total share of ULP contributing authors accounting for a total amount of 12090,9 fractional publications associated to the university over the period.



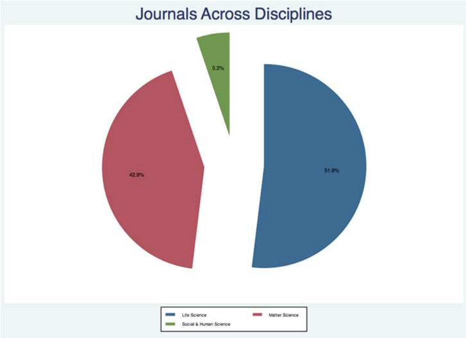
Set of Figures 6 Contributions, Publications and Fractional Publications

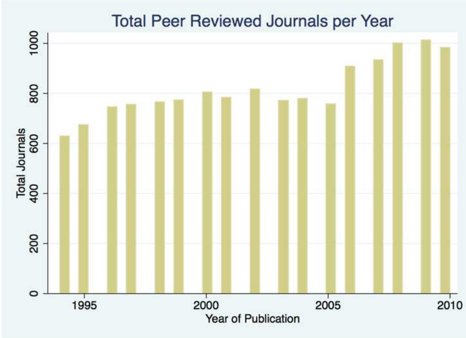
The following figure presents the distribution of both contributions and publications of the whole university across time. It reveals a growing trend from 1990 that stabilizes around 4000 contributions or 2000 publications in the year 1996.

In addition to the data on the number of publications, we were also able to gather information on their quality using as proxy the impact factors of the journal in which they were published. The information on these impact factors was collected from the ISI Web Of Knowledge – Journal Citation Reports (JCR), covering the period 1997 – 2010 for science journals, and the period 2002 – 2010 for both science and humanities journals. It accounts for a total of 12003 different scientific journals identified by their ISSN code, which allowed us to match the publications of the university with the information on quality of the journals they were published in.

Over the period 1988 – 2012, the scientific research of the university was published in at least 3886 different peer reviewed journals for which an identification ISSN code was associated to the publication. This period covers all waves of data collection, although for the years previous to 1997 and after 2010, we have no information on the impact factors and other related quality information on these journals. The lack of information for those periods is due to the fact that journal quality data was gathered during the 3rd wave of data collection (late 2011) from the available online ISI Journal Citation Reports, which only covered the period 2000 – 2010, and the archives for the period 1997 – 1999. Since our different analyses take into account the publications around (two years before) the 1996 ministry survey, we assume equality between the available journal impact factors in year 1997 and those 1994, 1995 and 1996.

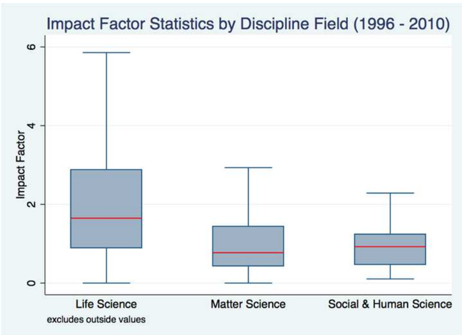
This data indicates that over the period of our interest (1994 – 2010, covering all 4 available ministry surveys), the researchers of the university published their work in 2265 different journals, with 51.9% of these journals published the research issued from the life science laboratories, 42.9% from matter science laboratories, and 5.2% published research from social science and humanities laboratories from the university as illustrated in the following figures.

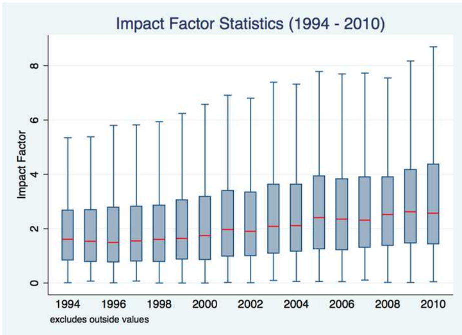




Set of Figures 7 Percentage of peer reviewed journals by discipline and total journals by year

In addition, the data on the journals in which the university members publish indicates the existence of a trend of increasing median impact factors over the whole period, as shown in the Set of Figures 8, which comforts the idea that the overall quality of our university's research has improved during this period, with those publications in the field of life sciences showing the highest median impact factor.





Set of Figures 8 Distribution of impact factors by discipline, and by year

2.4 Dataset "Funding"

The final dataset we have access to concerning the University Louis Pasteur gathers information on different funding sources of its research laboratories. We observe among others, public and private funding, as well as recurrent funding from new agencies such as the National Agency for Research[12] (ANR) and different regional and European financing.

As highlighted by Llerena and Benaim (2010), information on recurrent funding is difficult to gather due to a lack of centralization or even homogeneity of research funding data in a given region, or in our case, the laboratories of the university. In a study on the evolution of research funding in French laboratories, BETA was able to build a significant dataset of single endowments for research funding based on two ministry surveys, covering the period between 2001 and 2008, for the whole Strasbourg site – meaning all laboratories affiliated to all three historic universities of the city.

The different types of finding present in the dataset gather information on funding received by the research laboratories from:

- Regional collectivities: Covering funding from regional organisms, and representing a small share of the total research funding.
- Public recurrent funding: Covering the actual main source of laboratory funding (about 40% of their total resources), essentially coming from public organisms associated to the labs such as CNRS and INSERM. This type of funding is the one negotiated every 4 years during the ministry survey, with the objective to ensure the operability of the research laboratories.
- European Union: Covering funding from European structural programs regarding scientific research, which reflect the objectives of the Lisbon treaty and the “Horizon 2020” program. This type of funding represents about 10% of the total laboratory funding observed during the period 2001 - 2008.
- National Agency for Research: Covering project-based funding, it represents about 13% of total resources of the laboratories. Since the creation of the agency in 2005, this type of funding has changed the behaviour in research laboratories, and while still representing a relatively small share of funding; it has evolved progressively gaining in importance.

[12] The “Agence Nationale de la Recherche” is an agency in charge of increasing French research and innovation through specific project funding. The objective of the institution is to act in favor of economical and social priorities, intensify interactions between the private and the public sector, and help develop international partnership; all this by focusing on the improvement of the knowledge and technological base of the nation.

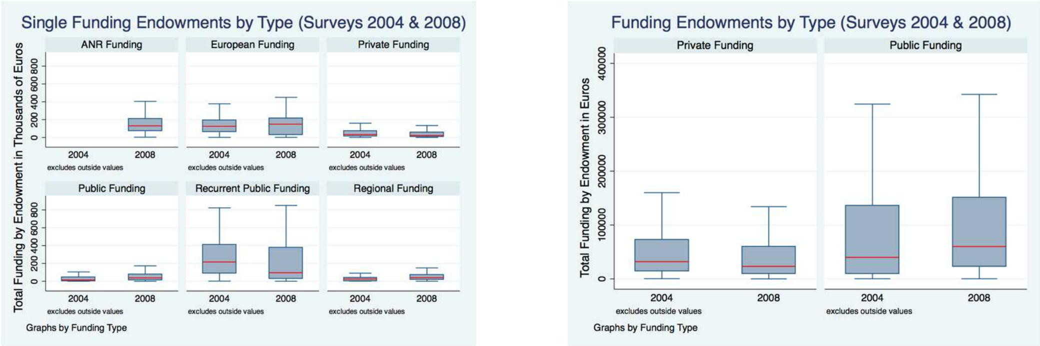
- Other public funding: Covering funding from other public institutions such as French ministries, Public investment banks, other CNRS project-based funding etc. This type of funding represents about 15% of total research funding.
- National and international private funding: Covering funding received from contracts with different private organism (either national or international), licensing, and patent exploitation. This type of funding represents about 20% of the total resources, with particular importance in those laboratories dealing with applied science near the market.

Using this information, and for the purpose of our analyses, we were able to aggregate these single endowments according to two different systems of categorisation:

- Private and public funding, as a general characterisation.
- General private and public funding, European and regional funding, recurrent public funding, and finally, ANR funding.

Since most of these single endowments are identified and the linked to the research laboratory they were affected is reliable based on the responses obtained by these surveys, were able to aggregate them according to these categories and obtain an observation for each laboratory belonging to the ULP, therefore creating a set of variables representing the funding structure of those laboratories.

We may note that this information on funding sources covers the period 2001 – 2008, which following the logic of the other scientific resource information (personnel) implies a correspondence to the 2004 and 2008 surveys; the aggregation of funding types from 2001 to 2004 representing an index of the funding received by the laboratories somewhat declared in the 2004 survey, and the aggregation from 2005 to 2008 representing that of the 2008 survey.



Set of Figures 9 Distribution of aggregated single endowments across periods of time corresponding to surveys 2004 and 2008

3 Merging datasets

Based on the information gathered from the different datasets described in the previous section, we were able to perform several merges of data with the objective of building a thorough body of information linking laboratory characteristics, personnel composition, funding structure and aggregated scientific production across two different levels: institutional and individual, using respectively the laboratory and the researcher as the unit of analysis.

These merges were possible given the presence of key ID variables such as the laboratory code or the researcher's identity (both by name and code within the university) in several datasets. For instance, during a first stage we performed the following merges:

- "Laboratory" and "Personnel" datasets.
- "Personnel" and "Publications" datasets.
- "Laboratory" and "Funding" datasets.

We used the laboratory code present in the personnel dataset to merge these two sets of information with the objective of obtaining the characteristics of the laboratory of affiliation for each researcher in the personnel dataset, and in addition, we collapsed the characteristics of researchers affiliated across their laboratory of affiliation with the objective of obtaining a detailed composition of research laboratories in terms of human resources for the different periods of our analysis; this two merging directions illustrate both analytical units, the laboratory and the researcher.

Moreover, we used the researchers ID code declared in the personnel dataset for the first three surveys (1996, 2000 and 2004), also present in the publications dataset until the first half of the year 2005; for the survey 2008, we performed the match using the surname and name initial – *SURNAME_I*, where *"I"* denotes the name initial – as an identifier of the publications gathered from the second half of the year 2005 on in the publications dataset, and the same structure drawn from the identity of the researcher in the personnel dataset given in the form *SURNAME_NAME*. This match allowed us to obtain a dataset associating the overall characteristics of each single scientific publication to each one of the individual researchers of the university appearing in the list of authors. By collapsing the number of publications and their characteristics over their associated authors by year we obtained a dataset that provides information on the publications of the researchers of the ULP.

Finally, the first merging stage involved a third match linking laboratory and the funding datasets. We used the laboratory code present in the funding dataset for each single endowment during the period 2001 – 2008 with the objective of obtaining a dataset representing the funding structure of laboratories during that period which is covered by the surveys 2004 and 2008.

During a second merging stage, we used the two of the datasets issued from the first wave of merges, the "researcher-laboratory" and the "researcher-publications" datasets. In this case we used the individual researcher ID code, now present in both of these datasets, with the objective of obtaining a complete dataset on associating the publications signed/authored by the individual researcher, his/her personal characteristics and the overall characteristics of the laboratory he/she works at, with the individual researcher being the unit of analysis. Furthermore, by collapsing all characteristics of individual researchers and their publications, we obtained a complete dataset providing information on the laboratory composition in terms of human resources and its overall scientific production in terms of quantity and quality, which reflects the laboratory as the analytical unit.

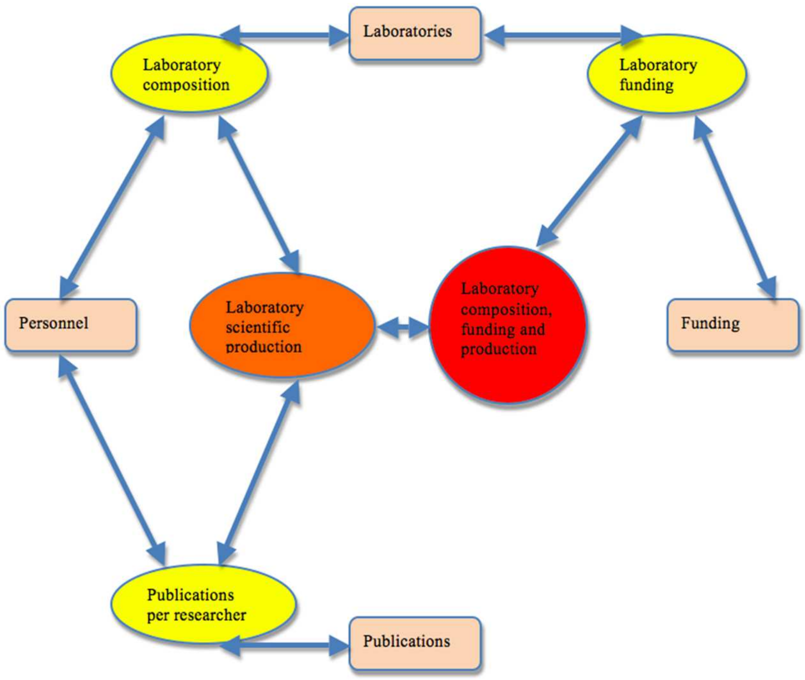


Figure 10 Data structure

During our third and last merging stage, we used the laboratory code to match the dataset "laboratory-personnel-publications", issued from the second merging stage, with the "laboratory-funding" dataset obtained earlier during the first wave of merging. As a result, we obtained two datasets (one for each analytical unit, the laboratory and the individual researcher) representing the overall characteristics of the laboratories, their scientific personnel, their scientific production and their funding structure for the period covered by the ministry surveys of 2004 and 2008. These datasets were limited to the last two surveys given the lack of information on the funding structures of the scientific research laboratories of the university prior to the year 2001.

Taking a deeper look at the data displayed in the datasets issued from our different merges we may observe other laboratory and individual factors of the scientific production process that we were not able to observe before. Some of these variables may provide some insights on different questions such as what the characteristics of laboratories in terms of composition and funding are? What are the characteristics of their scientific production? What the characteristics of those researchers working at the ULP who to scientific publications are? What types of researchers do they usually co–author with?

3.1 Merging the Laboratory and Personnel datasets.

Describing those variables representing this additional information given the laboratory as the analytical unit we may state that in terms of laboratory composition, there were a total of 612 senior researchers at the university, declared in the survey of 1996; this figure decreased down to 586 senior researchers in 2000, 566 in 2004 and finally 526 declared in the survey of 2008, with an average of around 7 and a median of 5 senior researchers per laboratory. The distribution of the total senior researchers across disciplinary fields is consistent with the distribution of laboratories across these general disciplines, with the life sciences varying between 41% and 48% of total senior researchers during the overall period, the matter sciences varying between 44% and 51% of them, and the social and human sciences varying between 6% and 7.6% of them during the whole period of the analysis.

On the other hand, the total number of junior researchers added up to 798 individuals in 1996, slightly increasing to 809 in 2000 and 830 in 2004 before dropping back down to 774 individuals declared in the survey of 2008. These individuals represented an average of 9 to 10 junior researchers per laboratory during the period 1996 – 2008 and an average of 15 per laboratory in 2008, with a respective median of 7 to 10 individuals during the whole period. Junior researchers were distributed across disciplinary fields in similar shares across time although with slight differences with respect to senior researchers, with life science laboratories holding around 44% of them, matter science laboratories 46%, and social and human science laboratories holding around 10% of junior researchers, a higher share than in the case of senior researchers.

PhD candidates and post doctors added up to a total of 1911 individuals declared in the survey of 1996, decreasing down to 1494 in 2000, and then growing back to 1586 in 2004 and 1941 in 2008, they represent an average of around 18 to 20 and a respective median of 11 to 12 young researchers during the period 1996 – 2004, though due to the decrease in the number of laboratories in 2008, these figures increase to an average of 38 and a median of 28 individuals in 2008. We may say that their distribution across general disciplinary fields is very similar to that of junior researchers.

Finally, assistant researchers and academic staff represented a total of 1111 individuals in 1996, 1108 in 2000, 1181 in 2004 and 1707 in 2008, with an average of 13 and a median of 6 individuals per laboratory during the period 1996 – 2004 increasing to an average of 34 and a median of 20 individuals in 2008. Their distribution across disciplinary fields shows that life science laboratories held a steady share of about 54% of them, while matter science laboratories a share of about 43%. Social and human laboratories only held a share of about 3% of the staff during the period covered by the analysis, which is different from the distribution across disciplines for other ranks of scientific workers.

Now, in terms of laboratory funding during the period 2001 – 2008 we may describe the behaviour of public and private funding during this period, which covers the ministry surveys of 2004 and 2008. During the 4 years preceding the ministry survey of 2004, the total of endowments constituting the public funding of the laboratories of the ULP added up to around 96.6M, during this period, this amount was shared between disciplinary fields, with laboratories in the life sciences benefiting from 43.4% of it, matter science laboratories from 53.8% of the amount, and social and human science laboratories from only 2.7% of the total public funding. Over the next 4 years, the total public funding decreased to 93.5M, shared among life science laboratories (32.3%), matter science laboratories (64.1%) and social and human science laboratories (3.4%). On the other hand, the total endowments constituting the private funding of the laboratories added up to about 18.25M during the 4 years preceding the survey of 2004, distributed in shares of 50.5% for life science laboratories, 48.2% for matter science and 1.3% for social and human laboratories. During the next period (2005 – 2008) this amount decreased to 11.12M with life science laboratories receiving only 17.3% of the amount while matter science laboratories increased their share to 79% and social and human science laboratories increased theirs up to 2.8% of this amount.

Finally, speaking about the scientific production of ULP laboratories and its quality we observe over the period covered by the four ministry surveys (1996 – 2008), we focused on the total publication counts over a period of 2 years (before and after) surrounding each of the ministry surveys. The amount of publications covering the period around the survey of 1996 (1995 – 1998) adds up to a total of 16461 scientific publications, during the next period (1999 – 2002) this amount increased to 17494 publications signed by at least one researcher working at a laboratory of the university; the following this period, this amount decreased to 16284 publications during the period 2003 – 2006 and then down to 14395 total publications during the period 2007 – 2010 corresponding to the survey of 2008. These publications follow a distribution across disciplinary fields were a steady share in time of about 51% to 55% belongs to the life sciences, a share of ranging between 43% and 47% belongs to the matter science, and a very small share ranging between 0.46% and 0.74% of the total publications of the university belongs were signed by researchers working in the field of social and human sciences. In addition, we also observed as indicators of the quality of these scientific publications the average and median impact factor per laboratory of the peered reviewed journals in which these publications appeared during the span of four years around a ministry survey. The average impact factor of journals increased steadily from 2.9 points, for the journals in which the publications aggregated for the ministry survey of 1996 appeared, to 3.05 points for those journals corresponding to the aggregation for the survey of 2000, until reaching 3.57 points for those corresponding to the survey of 2004 and 4.1 points for the survey of 2008. The associated median impact factor per period also increased along the different surveys, from 2.21 points associated to the survey of 1996, to 2.52 points in 2000, 2.9 points in 2004 and 3.3 points in 2008.

3.2 Merging Publications and Personnel datasets

From the merging of the publications and personnel datasets described at the beginning of the present section we were able to observe the patterns of distribution of several interesting variables such as the total publications and their distribution across different types of researchers, as well as total internal coauthorship of a given rank that individual researchers worked with during the period covered by the surveys and the average age and experience of their coauthors.

In the first place, we observed that the total publications (on the basis of a 4 year span over each ministry survey) were distributed almost steadily across senior and junior researchers, with a range of 57.3% to 62% of these publications being signed by a senior researcher and a range of 35% to 39% being signed by junior researchers. The share of the publications being signed by an assistant researcher is highly variable over time, with 0.01% of them in the survey of 1996, 0.2% in the next survey, 1.6% during the survey of 2004 and finally 0.3% during the last survey. In addition, the distribution of these publications across the major public research organisms (CNRS, INSERM and University) reflects a decreasing share of publications signed by researchers associated to the CNRS (from 50% down to 42%), a share ranging between 35% and 40% signed by university only researchers, and finally a rather small share oscillating between 7.3% and 9.3% signed by researchers affiliated to the INSERM.

Moreover, we also observed the total internal coauthorship of researchers working at the ULP, with a total of 24604 co-signatures performed by senior researchers during the period corresponding to the 1996 survey; this number decreases down to 17118 co-signature for the survey of 2008. The level of coauthorship shows there were an average of 20 and median of 9 senior coauthors per individual researcher revealed in the survey of 1996, which decreased to an average of 15 and a median of 7 senior coauthors in the survey of 2008. In addition, the distribution of these co-signatures of senior researchers show that a fluctuating percent in the range of 51% to 60% of them occurred in the field of life sciences, a range of 39% to 46% in the field of matter sciences and a range of 0.3% to 0.9% in the field of social and human sciences.

The total co-signatures of junior researchers decreased from a total of 17815 during the 1996 survey down to 13652 for the 2008 survey reflecting a decreasing an average from 14 to 12 total junior coauthors and a rather stable median of 6 to 7 of them per individual researcher. Coauthorships of this type, contrary to senior co-signatures, mainly appeared in the field of matter science, with a decreasing share in time from 60% to 51% of the total junior co-signatures happening in laboratories of this field. The laboratories in the field of life sciences presented an increasing share of 38% to 45% of junior co-signatures, while they oscillated within a range of 0.5% to 1.8% in the case of laboratories in the social and human sciences.

Finally, the level of coauthorship by assistant researchers added up to total of 917 signatures corresponding to the survey of 1996 and fluctuated to 1401 signatures in 2000, 862 signatures in 2004 and 1083 signatures in 2008. There were an average of 1 and median of 0 assistant coauthors per individual researcher during the whole period. Their distribution across disciplinary fields was given by a range of 45% to 60% of them happening in laboratories in the life sciences, 41% to 48% in laboratories in the matter sciences, and 3.8% to 4.3% for laboratories in the social and human sciences during the periods corresponding to the surveys of 1996, 2000 and 2008. The survey of 2004 diverges in the sense that laboratories in the life sciences hold 80% of the assistant coauthors while laboratories in the matter science hold 18.9% of them.

4 Conclusion

In summary, this section provided information on the methodology utilised to build the merged datasets at both the laboratory and the individual level. These datasets take into account valuable and reliable information from the four-year ministry surveys on research laboratories within the French academic system, in addition they make use of valuable information from the ISI Web of Knowledge regarding the scientific publications of researchers of the University Louis Pasteur as well as information on project funding from internal surveys on research projects within laboratories at the University Louis Pasteur. The following table provides details about the datasets and variables in use highlighting the laboratory and the personal identification codes that allowed the merging process.

Table 1 Detailed datasets

FUNDING	LABORATORIES	PERSONNEL	PUBLICATIONS
Laboratory Identification Code	Laboratory Identification Code	Laboratory Identification Code	Laboratory Identification Code
Contract start date	Year of Ministry Survey	Personal Identification Code	Personal Identification Code
Contract end date	Researcher / Teacher	Civil Status	Year
Time period	Researcher CNRS	Year of Ministry Survey	Total Publications
Funding Source	Researcher INSERM	Date of Birth	Total Authors
Amount	Researcher OTHER	Age during Survey Year	Total Internal Authors
	PhD Students	Academic Rank	Laboratory Discipline Code
	Defended Theses	PRO affiliation	
	Personnel University	Laboratory Name	
	Personnel CNRS	Habilitation	
	Personnel INSERM	Year of Habilitation	
	Personnel OTHER	Date of Entrance in Laboratory	
	Laboratory Discipline Code	Personal Career in Laboratory at Moment of Survey	
	Researchers (full time equivalent)	Laboratory Discipline Code	
	Personnel (full time equivalent)	Researcher Category	
	Researchers with Habilitation	Average Age in Laboratory	
	Post Doctoral Researchers	Average Career in Laboratory	
		Year of Entrance in Laboratory	

Chapter 3

Determinants of Scientific Research

Chapter 3: Determinants of Scientific Research

1 Introduction

The present chapter studies the individual characteristics of researchers members of research laboratories and their impact not only on the individual scientific output and quality but also on that of the research laboratory as a whole. The interest of this study is to investigate the determinants that explain collective scientific research in terms of scientific publications and the quality of peer reviewed journals in which they are published.

This chapter is composed of a first section on the determinants of collective scientific performance in terms of publications and quality of publications for which empirical results show that the composition of research laboratories in terms of researcher status plays an important role on the output and quality of the laboratories as whole entities; in addition, results also show that the composition of the funding structure of these laboratories also play an important role on their collective performance. A second section on the determinants of individual scientific performance follows to present a complete picture of scientific production and quality.

Finally, the chapter ends with an empirical analysis of the characteristics of coauthors and their influence on the individual scientific production and quality. This section offers a first analysis of the complementarity of status among coauthors at the individual level, and represents a preamble to chapter 4 in which these types of complementarities are studied using a supermodular approach.

2 Determinants of collective scientific performance

The aim of the present chapter is to evaluate the role of laboratory composition and funding as determinants of the scientific output of research laboratories. For this purpose, I follow the empirical work of Carayol and Matt (2006) who analyzed the scientific production of faculty members of the Louis Pasteur University to understand which are the collective and the individual characteristics that explain individual scientific output and quality. The present analysis is carried out on both, the count of contributions to publications and the fractional publications performed by members of the laboratory and goes further by studying not only the individual output and quality but more important, the collective output and quality of the laboratory as a whole entity. The dependent variables I define reflect the total participation of research laboratories in scientific production trough contributions and fractional publications and the average and median quality of this output.

I measure the dependent variables according to three different assumptions:

1. The personnel declared in a given survey have had an important influence on the laboratory contributions and publications during the previous two years.
2. The personnel declared in a survey have an influence on the contributions and publications during the two years following it.
3. The personnel declared in a survey have an influence on the contributions and the publications of the laboratory during a range of two years before to two years after the survey.

The purpose of measuring the dependent variables according to these particular cases is to take into account the evolution of the scientific personnel from survey to survey according to periods of four years. In fact, the personnel declared during a given survey may not only be composed of researchers who were already affiliated to the laboratory during the previous period and are still affiliated during the period corresponding to the survey, but also researchers recently affiliated to the laboratory who become active only during the period following the ministry survey.

I illustrate this by recalling our figures on researcher's turnover between different surveys. From the period 1996 to 2000, the university laboratories lost a total of 317 out of 1451 individuals; which are those researchers declared in 1996 but who were no longer affiliated in 200. This turnover concerns about 22% of the total university researchers in 1996, with 52% of this loss of researchers occurring in matter science laboratories, 40% in life science laboratories and 8% in the social and human science. Furthermore, one may notice that the loss of researchers between the period 1996 and 2000 was compensated by a total gain of 292 researchers who were not listed in the 1996 survey and appeared as entrants in the following survey of 2000 that counted 1426 researchers; this entrants represented 20.5% of the university researchers in 2000 and were distributed according to a share of 45% entrants in life science laboratories, 40% in matter science laboratories and 15% in social and human science laboratories.

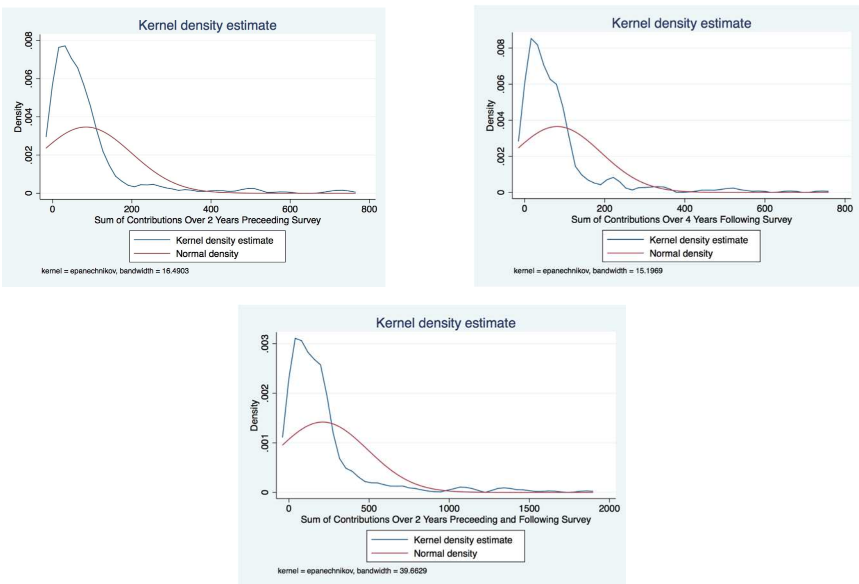
However, since more than two thirds of these entrant researchers who were only declared in the ministry survey of 2000 actually entered their university laboratory at some point in time between 1996 and 2000, we must take into account the fact that a few years are needed in order for scientific research to be published, especially for articles published in journals in technical sciences and mathematics (Luwel and Moed, 1998).

Given this fact, I formulate the following three assumptions on the aggregation of scientific output matched with the data particularly gathered in periods of four years, as it is the use within the French academic system:

- Those individuals who performed research during the period 1996 – 2000 may be accountable for the publications performed during a period of two years preceding the ministry survey of 2000 and 2 (model 1).
- They continued working for the laboratory during the period 2000 – 2004 and therefore may also be accountable for the laboratory publications signed during a period of two years following the ministry survey of 2000 (model 2).
- If the two previous assumptions hold, then we may also assume that researchers declared during this ministry survey may accountable for the laboratory publications signed over a period of four years around that survey (model 3).

In addition, we may also note that in between these two surveys (1996 and 2000) a total of 90 researchers transferred from one of the university laboratories to another; this means they were declared in both surveys but in different laboratories, with the transfer taking place at some point in between the surveys. The issue of transfers in between laboratories and their influence on scientific publications may be addressed by assumptions 1 and 2. Researchers declared during a given survey may be accountable for publications during years before and/or after the declaration, indeed, a researcher working in a given laboratory may be accountable for publications signed during a period of two years before the survey, then after his transfer, he may be accountable for publications signed at his new laboratory during a period of two years following the same survey.

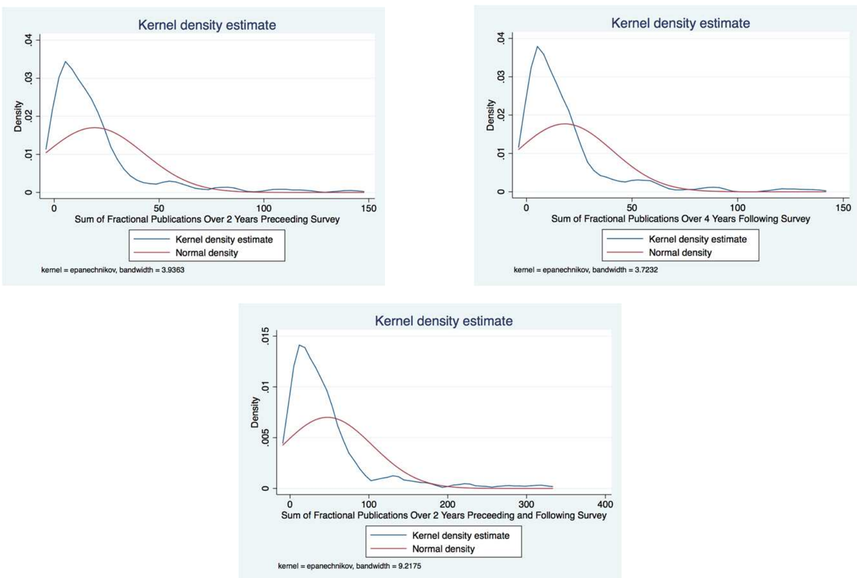
In addition, the turnover of individual researchers in the university laboratories represented a loss of 398 researchers between the surveys of 2000 and 2004, with a share of 43% of the exits taking place in life sciences laboratories, a share of 45% in matter science laboratories and a share of nearly 11% in the social and human science laboratories. During the same time, the university affiliated a total of 406 new individual researchers, with a share of 45% of these new entries occurring in life science laboratories, a share of 41% in the matter science laboratories and a share of 13% in the social and human science laboratories; during this same period some 119 researchers transferred from one laboratory to another.



Set of Figures 11 Density of total contributions against normal density

Finally, in between the last two surveys (2004 and 2008) these figures became somewhat different; there were a total of 399 researchers who exited the survey, most of them in matter science laboratories (47%) while a total of 396 entered the survey with a similar distribution as in the previous periods; although it is interesting to notice is that a total of 508 researchers appear to have changed laboratories in between these two surveys. This high figure of transfers may be explained by the fact that, in time, several laboratories within the university merged explaining the laboratory attrition on the ministry surveys, though reflecting the fact that the human resources of these labs became part of a bigger research unit within the university.

Having explained the assumptions on the output aggregation, Let us develop further the description of the dependent variables; these correspond to the sum of contributions – that is every time a researcher participates in a publication it counts as one event – and the sum of fractional publications – that is the actual share of the researcher's contribution among his coauthors –; these sums are calculated according to the three assumptions on output aggregation mentioned above. These dependent variables (contributions and fractional publications) are defined by the sets of observations in the publications dataset; they take on positive values, are highly non-normal and are skewed to the right as shown in the Set of Figures 12



Set of Figures 12 Density of fractional publication counts against normal density

Scientific production is based on two important factors, the human factor based on the individual researcher's skills and competences, and the technical factor based on the financial resources used by the researchers as means to produce new science as it is witnessed by the important growth of universities towards the last quarter of the twentieth century in terms of the number of researchers and the level of financial commitment (Geuna, 1999). Within the frame or this research, my interest focuses the role these factors exercise on the scientific production process; in other words, how different characteristics may explain the scientific output of individual researchers and research laboratories.

Laboratory composition in terms of human resources ensures a key role in the process; in fact, human resources may be decomposed according to their rank/status. Within the particular frame of the French academic system as described in the previous chapter on data, the categories of ranks/status include senior or highly experienced researchers with a professional career longer than 8 years, junior researchers with about 5 years of experience, postdoctoral researchers and PhD candidates on well-defined research contracts and/or just working on their doctoral thesis, and finally assistants which include engineers, technicians, assistants, second-class professors, and other personnel.

Moreover, the financial resources exploited by researchers and other personnel as means to ensure operability and production also play a key role in the scientific production process. With the objective of understanding how different characteristics of funding resources may affect the scientific production of research laboratories, I performed my studies given a decomposition of funding resources as independent variables. These resources, as mentioned in the previous chapter, can be decomposed into general private and public funding, European and regional funding and finally recurrent public funding.

Since in the present case the studies are carried on the research laboratories of the University Louis Pasteur, one must keep in mind that most of them are composed of mixed structures where both the university and the public research institutions such as CNRS, INSERM etc. operate simultaneously. Researchers and other personnel working within the same laboratory or research groups may thus be affiliated either to the university itself or to one of these public institutions, therefore the allocation of their time between researching and teaching may vary. Given the existence of such mixed structures, we use the number of institutional researchers and the number of institutional personnel in the laboratory to emphasize the effects of such mixed research units.

We also look for the effects of the average age and experience of researchers in squared terms to take into account their non-linear relationship with the dependent variable and the number of defended PhD theses declared at the laboratory for each survey to capture the capacity of laboratories to consolidate knowledge by turning PhD students in to doctors. Finally, we control the effect of the discipline field and the method of data collection to capture the difference between the information gathered before and after mid 2005.

2.1 Effects of human resources on scientific research

Arranging the decomposition of human and funding resources plus all other variables capturing the laboratory characteristics in the form of a pooled panel dataset drawn from the ministry surveys of 1996, 2000, 2004, and 2008. I use a pooled panel data setting of the type $y_{i,t} = \alpha_{i,t} + x'\beta_{i,t} + u_{i,t}$ given that the data constitutes a short, unbalanced panel with regular intervals, with both T and N small; within this frame and based on Collin Cameron's explanation on panel data methods for microeconometrics (2007)[13] , I consider that the information in the panel is not subject to a comparison between fixed and random effects given that for such short panel data (with T and N small) the estimate of the constant term α_i cannot be consistently estimated and therefore I chose to use population average models where α_i is purely random. I performed two different types of analyses; the first is a count data analysis of the total contributions given its nature of count variable and the second is a log linear regression for the analysis of fractional publications given its nature of continuous variable. These two analyses are modelled by the following expression:

$$y_{i,t} = \mu_{i,t} = exp(x_{i,t}'ß)\varepsilon_{i,t}$$

Where y is the vector of total contributions of a given laboratory to publications in peered reviewed journals, x is the vector of total researchers in each category of scientific personnel and other control variables, i is an index denoting a specific research laboratory or a specific individual researcher and t denotes the period the observation belongs to.

The regression analysis on the role of human resources on collective scientific production is based on the decomposition of the scientific personnel according to their rank/status within the academic system, the institutional affiliation of researchers within the laboratory and other laboratory characteristics. We explain therefore both, contributions to publications and fractional publications, by the total number of researchers in each of the following categories: senior, junior, PhD candidate or post doctor, and assistants; which allow writing the regression as follows:

[13] Explanation based on: C. Cameron and P. Trivedi, *Microeconometrics Using Stata*, *Stata Press*, 2007.

$$\textit{Contributions and/or Fractional publications}_{i,t} =$$
$$\exp(\beta_1 + \beta_2 \textit{ Senior Researchers}_{i,t} + \beta_3 \textit{ Junior Researchers}_{i,t}$$
$$+ \beta_4 \textit{ Doctoral and Postdoctoral Researchers}_{i,t} + \beta_5 \textit{ Staff Personnel}_{i,t}$$
$$+ \beta_6 \textit{ Institutional Researchers}_{i,t} + \beta_7 \textit{ Institutional Personnel}_{i,t} + \beta_8 \textit{ Defended Theses}_{i,t}$$
$$+ \beta_9 \textit{ Age}^2_{i,t} + \beta_{10} \textit{ Career}^2_{i,t} + \beta_{11} \textit{ Matter Science}_{i,t} + \beta_{12} \textit{ Humanities}_{i,t}$$
$$+ \beta_{13} \textit{ OST data collection method}_{i,t}) \, \varepsilon_{i,t}$$

The correlations among laboratory composition variables, reported in Table 2, show positive linear relationships among most of them except the staff personnel, which displays a negative and weak relationship with all other independent variables. This observation is rather unexpected since one may suppose that as laboratories are bigger in terms of human resources, they would need to increase as well their supporting staff for organizational purposes in order to be more efficient, however, in the present case the inverse is observed, which suggests that young and experienced researchers tend to manage their own activities with little support from staff personnel; this addresses the current debate on whether nowadays, experienced researchers actually perform high quality research or trade their research activities for management activities. In addition, one may also note that the average age of researchers in the laboratory also presents a negative relationship with some variables such as the junior, PhD's and post doctors and staff personnel, which is expected since as researchers grow older they are promoted towards the status of senior researchers.

Table 2 Correlations among laboratory composition variables

Independent variables	Senior	Junior	PhD and Postdoc	Staff	Inst. Research.	Institutional Personnel	PhD Theses	Age	Career in Lab
Senior	1								
Junior	0.8828	1							
PhD/Postdoc	0.7183	0.7004	1						
Staff	-0.1214	-0.0802	-0.0585	1					
Inst.Research.	0.8361	0.7833	0.7292	-0.0763	1				
Inst.Pers.	0.6581	0.6504	0.7203	-0.0743	0.8242	1			
PhD Theses	0.8235	0.8331	0.8221	-0.0675	0.8081	0.6971	1		
Age	0.0367	-0.0594	-0.0548	-0.0662	0.0612	0.0411	-0.0175	1	
Career in Lab	0.2413	0.2153	0.1274	-0.0959	0.2252	0.2832	0.2026	0.2206	1

2.1.1 Influence of human resources on total contributions to publications

As the model explaining the contributions to publications deals with a positively skewed dependent variable defined by positive integers, the starting point of the analysis is a comparison between a Poisson and a Negative Binomial regression of the total contributions on the laboratory composition, followed by a test of over dispersion of the dependent variable to justify the choice between these two models.

Since the Poisson model assumes equidispersion[14] of the dependent variable, the next step of the analysis is to perform a test of over dispersion following Cameron and Trievdi (2005) to check whether or not the Poisson specification is appropriate.

The test consists on assessing whether the null hypothesis representing the equidispersion of the process assumed by the Poisson model holds against the alternative hypothesis: $Var(y|x) = E(y|x) + a^2E(y|x)$. Such test is implemented by creating an over-dispersion test statistic a, where:

$$H_0 : a=0 \text{ against } H_1 : a\neq 0$$

Through the auxiliary equation: $[(y_{it} - \mu_{it})^2 - y_{it}] / \mu_{it} = a[\mu_{it}^2 / \mu_{it}]$

Overdispersion test	Model1	Model 2	Model 3
	Sum of contributions over 2 years preceding the survey	**Sum of contributions over the 2 years following the survey**	**Sum of contributions from 2 years before to 2 years after survey**
Overdispersion Statistic	0.1131***	0.1284***	0.1120***
	(0.0164)	(0.0211)	(0.0170)
a is significantly positive at the 1% threshold therefore implying overdisperion			

From the over-dispersion test we learn that the coefficient a is significantly different that zero for all three models, implying that a Poisson regression is not appropriate since it does not take into account the over-dispersion of the dependent variable (in this case the contributions to publications). The commonly used alternative to the Poisson model for our data is a Negative Binomial model, which takes into account the over-dispersion of our dependent variables.

In the case of the first assumption I suppose that the laboratory composition has an influence on the articles published during the two years preceding each ministry survey. The results obtained from the negative binomial regression, detailed in Table 3, reveal that the scientific production in life science laboratories is mainly driven by senior researchers and we notice that the headcount of researchers affiliated to a research institution different than the university (CNRS, INSERM or other) and the average career

[14] Equidispersion is defined by an equality between the variance and the expectancy of the dependent variable given its explanatory variables: $Var(y|x) = E(y|x)$.

in the laboratory also present a significant influence on the expected log of the total contributions to publications.

The exponentiation of the coefficients provide the associate semi-elasticicities which gives us the incidence rate ratio of the explanatory variables on the expected outcome; in this case this ration indicates an a increase of +2.91% in the total expected contributions for each additional senior researcher in the laboratory –holding all other variables equal–; while in terms of institutional affiliation, a positive variation of one researcher affiliated to a research institution different than the university represents an increase of +3.61% and that of an additional institutional personnel represents a small decrease of -0.58% of the total expected contributions; moreover, the career effects represent a slight increase of +0.09% for each additional year in the average career of researchers in the laboratory. These results mean that within the mixed structure of research laboratories, increasing the presence of researchers with an institutional affiliation different that the university is associated with a higher scientific output. These results are coherent with previous works indicating that researchers with increased academic obligations produce less scientific output than researchers who do not have to split their time between pure scientific research and teaching activities.

In case of our second assumption, where the laboratory composition has an influence on the articles published during the two years following the ministry survey, I noticed that senior researchers, PhD’s and postdocs as well as assistant researchers have a significant and positive impact on the total count of contributions to publications performed by researchers working for a laboratory of the university.

Table 3 Regression results for the research output explained by laboratory composition

	Model 1 NB	Model 2 NB	Model 3 NB	Model 1 LogLinear	Model 2 LogLinear	Model 3 LogLinear
Explanatory Variables	Sum of contributions over 2 years preceding the survey	Sum of contributions over the 2 years following the survey	Sum of contributions from 2 years before to 2 years after survey	Sum of fractional publications over 2 years preceding the survey	Sum of fractional publications over the 2 years following the survey	Sum of fractional publications from 2 years before to 2 years after survey
Senior	0.0287**	0.0379***	0.0333***	0.0359***	0.0421***	0.0397***
	(0.0128)	(0.0129)	(0.0127)	(0.0112)	(0.0115)	(0.0118)
Junior	-0.0032	-0.0014	-0.0015	0.0071	0.0074	0.0088
	(0.0089)	(0.0081)	(0.0084)	(0.0082)	(0.0090)	(0.0096)
PhD and Postdoc	0.0023	0.0039**	0.0030*	0.0065**	0.0078***	0.0083***
	(0.0020)	(0.0017)	(0.0018)	(0.0027)	(0.0026)	(0.0028)
Staff	0.0934	0.1377*	0.124	0.0164	-0.0087	0.0271
	(0.0867)	(0.0770)	(0.0816)	(0.0787)	(0.0947)	(0.0832)
Inst. Researchers	0.0355***	0.0280***	0.0316***	0.0306***	0.0257***	0.0278***
	(0.0080)	(0.0069)	(0.0072)	(0.0072)	(0.0066)	(0.0069)
Inst. Personnel	-0.0058**	-0.0041*	-0.0050*	-0.0110***	-0.0092***	-0.0115***
	(0.0027)	(0.0024)	(0.0026)	(0.0026)	(0.0027)	(0.0028)
PhD Theses	0.0075	0.0001	0.004	0.0001	-0.0051	-0.0033
	(0.0069)	(0.0059)	(0.0063)	(0.0056)	(0.0057)	(0.0062)
Age	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	(0.0002)	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Career in Lab	0.0009**	0.0003	0.0006	0.0008**	0.0001	0.0008*
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0006)	(0.0004)
Data Method	0.3151**	0.2349**	0.2925***	0.1094	0.0147	0.172
	(0.1265)	(0.1057)	(0.1133)	(0.1431)	(0.1386)	(0.1496)
Matter Science	-0.0941	0.0444	-0.0344	0.0935	0.2679***	0.1256
	(0.1047)	(0.0933)	(0.0964)	(0.1092)	(0.1003)	(0.1009)
Humanities	-2.6731***	-2.0584***	-2.3563***	-1.7631***	-1.2960***	-2.0428***
	(0.2278)	(0.1886)	(0.1955)	(0.1892)	(0.2112)	(0.2685)
Life Science (constant)	3.2867***	3.4570***	4.2728***	1.6114***	1.6821***	2.5507***
	(0.3830)	(0.3360)	(0.3510)	(0.3697)	(0.3716)	(0.3969)
N	256	256	256	245	248	252
r2				0.6521	0.6142	0.6751
F				38.4084	34.2708	33.689
rmse				0.6623	0.6866	0.7019
chi2	401.7395	420.7193	425.6299			
ll	-1286.0299	-1319.0559	-1536.515	-239.9987	-251.9696	-261.6915
aic	2598.0598	2664.1118	3099.03	505.9973	529.9393	549.383
bic	2644.1471	2710.1991	3145.1173	551.5137	575.6138	595.2656
Standard errors in parentheses	* p < 0.1	** p < 0.05	*** p < 0.001			

The incidence rate ratios of these explanatory variables show respectively an increase of +3.86%, +0.39% and +14.76% in the expected contributions associated with the variation of one unit of these researchers. These results are of particular interest for the present work given that they provide evidence that scientific research is driven not only by senior researchers, but also by very young researchers and technical staff or personnel, which as we may recall is the core of our analysis. In terms of institutional affiliation, a positive change of unit in the count of institutional researchers represents an increase of +2.83% of the expected contributions, while a similar variation in the count institutional personnel is associated with a decrease of +0.41% expected contributions, which indicates that the presence of more institutional researchers is associated with higher scientific output.

In the case of our third assumption, where I suppose that the laboratory composition has an influence on the total contributions performed during a four year span around the ministry survey, I found out that only senior researchers and PhD's and post doctors explain the scientific output with positive and significant impacts associated with a respective increase of +3.38% and +0.3% in the expected outcome for each additional unit in their counts. . This main result corroborates our proposition that scientific production in research laboratories does not only rely on senior and experienced researchers, but also on younger and inexperienced ones.

The institutional affiliation of researchers also presents positive and significant effects, while the effects of the institutional affiliation of personnel are significant and negative with an associated increase of +3.2% and a decrease of -0.5% of the total contributions respectively for each additional unit in their headcounts.

Finally, one may notice that for all three different models, laboratories in the social and human sciences present lower counts of contributions with a decrease ranging from - 87.1% (model 3) to -93.3% (model 2) in the expected output with respect to those laboratories in the life sciences; these results are expected given that laboratories in the life sciences do present higher output volumes with respect to those in the social sciences, hence the negative semielasticities associated to laboratories in social and human sciences. Furthermore, our indicator for the change in the data collection method (for publications from mid 2005 on) does present a significant positive effect on the output indicating higher contribution counts after changes in the method. This particular shows that differences in the procedures used to match researchers and publications do influence the statistical analysis which calls for comparative research on the differences of analyses with same-period data collected using a researcher-to-publication matching direction on one hand and using a publication-to-researcher direction on the other; unfortunately, in our case data was collected according to a researcher-to-publication for the first three quarters of the period and to a publication-to-researcher for the latest quarter.

2.1.2 Influence of human resources on total fractional publication counts

The analysis of the total sums of fractional publication counts of the laboratories allows us to determine the impacts of the laboratory composition on the actual shares of publications attributed to members of the university laboratories instead of single contributions, which provides a more complete evaluation of the production capacity of research laboratories by the efforts of individual researchers. Since I deal in the present case with a continuous positively skewed variable, I perform a log linear regression with robust errors, hence modelling the effects of the explanatory variables on the expected logs of the sum of publications shares. The results from these regressions (models 1, 2 and 3) are reported in the second half of Table 3 along the results from the regression of contributions to publications.

In the case of our first assumption (model 1), these analyses show there are significant and positive effects from senior researchers and PhD’s and postdocs; the incidence rate ratio for these variables obtained through the exponentiation of their coefficients indicates the that an increase on one unit in either of these categories –holding all other variables equal– is associated with an increase of +3.65% and +0.65% in the expected total fractional publication counts of the university laboratories. Additional units of institutional researchers and institutional personnel are respectively associated with an increase of +3.1% and a decrease of -1.1% fractional publications. The average career of researchers in the laboratories also presents significant and positive effects for models 1 and 3.

In the case of our second assumption, we noticed very similar results concerning the effects of laboratory composition –positive and significant for senior researchers and PhD’s and post doctors, with respective increases of +4.3% and +0.78% in the expected outcome associated with variations of one unit in their headcounts–. Also similar results were found concerning the institutional affiliation of researchers and personnel with an increase of +2.8% in the expected outcome for each additional institutional researcher and a decrease of -1.14% in the expected outcome for each additional institutional personnel.

In the case of our third model, we also found positive and significant effects of senior researchers and PhD’s and postdocs in terms of laboratory composition (+4% and +0.88% associated increase in the expected outcome respectively), positive and significant effects for institutional researchers (+2.8% associated increase), negative and significant effects for institutional personnel (-1.14% decrease) and positive and significant effects for the average career of researchers in the laboratory.

We may also notice that for all three models, laboratories in the social and human sciences present statistically lower fractional publications with respect to those laboratories in the life sciences, ranging from -72.6% (model 2) to -86.9% (model 3) lower counts. Only in the case of our second model we found significant effects of laboratories in the matter sciences, with higher expected fractional publications counts of +30.7% with respect to laboratories in the life sciences.

As a conclusion of these two different analyses of the effects of laboratory composition on the expected scientific output we may distinguish senior researchers and PhD’s and post doctors as main determinants of this production. In terms of laboratory characteristics, we found that higher numbers researchers affiliated to a non-academic research institution like CNRS or INSERM are associated with a higher expected production while the same is not true for higher numbers of personnel affiliated to these institutions; in addition, we found there are significant career effects in the laboratory although no evidence of age effects were found and neither were found effects of the laboratories’ capacity to produce doctors or young researchers since the count of defended theses did not present significant effects.

The use of two different analyses (contributions to publications and fractional publications) allowed me to study the phenomena in a more robust manner and understand the process better since they rely on different mathematical process and different definitions of the dependent variable as follows:

- Contributions to publications as count data treated using a Negative Binomial process.
- Fractions of publications as continuous data treated using a Log Linear process.

These two different analyses provide very similar results, which indicates they are robust and reliable, shedding light on the influence of certain laboratory characteristics on the process of collective scientific production.

2.2 Effects of funding sources on scientific research

In the present subsection I switch my attention and focus on the influence that different funding sources of research laboratories may have on their scientific output. The objective here is to explain the scientific output of research laboratories not only by their characteristics but also by a decomposition of their funding structure.

For this purpose, I use as independent variables the amount of regional funding from regional organisms, European funding from European structural programs regarding scientific research, public funding from public institutions such as French ministries, public investment banks and some project-based allowances, private funding from contracts with private organisms, licensing and patent exploitation and finally, recurrent public funding which comes essentially from public organisms associated to the research laboratories such as CNRS and INSERM and is renegotiated every four years (nowadays five years) based on the ministry surveys.

In addition to the decomposition of the funding structure, I also use laboratory characteristics as explanatory variables in the present models such as the number of institutional researchers and personnel, the number of defended theses, the age and career of researchers and the total size of the laboratory in terms of human resources, controlling for the laboratory disciplinary field and the data collection method.

Let us recall that the dependent variables (contributions to publications and fractional publications) are non-normal and positively skewed, which suggest they should be modelled according to the equation $y_{i,t} = \mu_{i,t} = exp(x_{i,t}'ß)\varepsilon_{i,t}$, with the resulting model being represented as follows:

$$Contributions\ to\ publications_{i,t} =$$
$$exp\ (\ ß_1 + ß_2\ Regional\ Funding_{i,t} + ß_3\ European\ \ Funding_{i,t} + ß_4\ Private\ Funding_{i,t}$$
$$+\ ß_5\ Public\ \ Funding_{i,t} + ß_6\ Recurrent\ Public\ Funding_{i,t} + ß_7\ Size_{i,t}$$
$$+\ ß_8\ Institutional\ Researchers_{i,t} + ß_9\ Institutional\ Personnel_{i,t} + ß_{10}\ Defended\ Theses_{i,t}$$
$$+\ ß_{11}\ Age^2{}_i + ß_{12}\ Career^2{}_{i,t} + ß_{13}\ Matter\ Science_{i,t} + ß_{14}\ Humanities_{i,t}$$
$$+\ ß_{15}\ OST\ data\ collection\ method_{i,t}\ )\ \varepsilon_{i,t}$$

Table 4 Correlations among laboratory-funding and other characteristics

	Reg. Funds	Eur. Funds	Private Funds	Public Funds	Recurr. Public Funds	Size	Inst. Res.	Inst. Prs.	PhD Theses	Age	Car. in Lab
Regional Funds	1										
European Funds	0.1759	1									
Private Funds	0.1468	0.473	1								
Public Funds	0.5428	0.1361	0.2783	1							
Recurrent Public Funds	0.4607	0.0964	0.0345	0.3764	1						
Size	0.5124	0.0228	0.0575	0.4158	0.8708	1					
Inst. Research.	0.4479	-0.0355	0.0257	0.2622	0.8819	0.9309	1				
Ins. Personnel	0.3846	-0.0992	-0.0044	0.3284	0.7907	0.9212	0.8751	1			
PhD Theses	0.5385	0.0231	0.0911	0.4088	0.8199	0.9311	0.8715	0.7865	1		
Age	-0.1441	0.0654	-0.0033	-0.2585	0.077	0.0872	0.1756	0.0627	0.0053	1	
Career in Lab	-0.1425	-0.0935	-0.1587	-0.0572	0.053	0.2161	0.2195	0.2941	0.1812	0.163	1

The correlations among these variables reported in Table 4 show that relationships among different types of funding and among different laboratory characteristics are positive; what is interesting to point out is that funding may display negative relationships with some laboratory characteristics; indeed, regional, private and public funding decrease as the average age and career in the laboratory increases, suggesting that these types of funding serve teams of young researchers in priority. In addition we may notice that while both institutional researchers and personnel are highly correlated with recurrent public funding, their correlation with other types of funding, specially European funding, is negative suggesting that recurrent public funding may crowd-out other types of funding in research structures where the intensity of researchers with an institutional affiliation different than the university (CNRS, INSERM or other) is important.

2.2.1 Influence of different funding types on contributions to publications

The results from the analysis of the impact of funding on the scientific production shown in Table 5 indicate that in the case or our first model, there are positive and significant effects from the public funds obtained by research laboratories of the university. An additional resource of a thousand Euros of public funding is therefore associated with a slight increase of +0.05% of the expected contributions counts of the laboratory.

Regarding our second and third models, we found positive significant effects of public funding with an associated increase of +0.03% expected contributions to scientific publications, in both models, for each additional thousand Euros allocated to research projects a given laboratory.

In this analysis and across all three different models of aggregation it is important to recognize that, we found very similar and significant effects from other explanatory variables representing laboratory characteristics such as the total number of researchers in the lab (size), the total number of institutional researchers (CNRS, INSERM or other), the number of defended theses and the average age in the laboratory. In fact size effects are negative and represent a decrease of -1.26%, -0.65% and -0.89% expected contributions for models 1, 2 and 3 respectively; this findings indicate that the scientific production in research laboratories presents decreasing returns to scale, with bigger laboratories producing relatively less.

Positive effects were found in the case of the institutional affiliation of researchers in the laboratory, with an associated impact of +7.66%, +5.66% and +6.39% expected contributions for each additional institutional researcher in the laboratory for our three models respectively, as well as positive average laboratory age effects associated with an increase of +0.07%, +0.05% and +0.06% expected contributions. In addition, we found for the first time in our different analyses on the scientific productions significant effects from the number of defended theses during a given period with an associated increase of +3.2%, +1.25% and +2.2% expected contributions (models 1, 2 and 3). These results tell us that laboratories with more researchers devoted to conduct solely research activities perform better than those laboratories with more researchers splitting their time between teaching and researching; in addition, these results also suggest that those laboratories with a higher capacity to turn PhD students into young researchers perform better.

2.2.2 Influence of different funding types on fractional publications

We now focus our interest on the effects of different sorts of funding on the actual amount of publications shares produced by research laboratories of the university. The results form the log linear regressions reported in Table 5 show that in the case of our first assumption, public funding also plays an important role as determinant of the scientific research with positive significant effect and an associates increase of +0.06% expected publications shares for each additional thousand Euros available for scientific research in the form of public funding. Results are also similar in the case of our second and third assumptions on the aggregation of scientific output with an associated increase of +0.04% expected share in the case of model 2 and of +0.05% in the case of model 3 for each additional thousand Euros of public funding.

Table 5 Regression results for research output explained by laboratory funding

	Model 1 NB	**Model 2 NB**	**Model 3 NB**	**Model 1 LogLinear**	**Model 2 LogLinear**	**Model 3 LogLinear**
Explanatory Variables	**Sum of contributions over 2 years preceding the survey**	**Sum of contributions over the 2 years following the survey**	**Sum of contributions from 2 years before to 2 years after survey**	**Sum of fractional publications over 2 years preceding the survey**	**Sum of fractional publications over the 2 years following the survey**	**Sum of fractional publications from 2 years before to 2 years after survey**
Regional Funds	-0.0006	0	-0.0003	-0.0007	-0.0004	-0.0005
	(0.0007)	(0.0004)	(0.0005)	(0.0008)	(0.0007)	(0.0007)
European Funds	0.0002	0.0001	0.0001*	0.0002	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Private Funds	0	0.0001	0	0	0	0
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0001)
Public Funds	0.0005***	0.0003***	0.0003***	0.0006***	0.0004**	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0001)
Recurrent Public Funds	0	0	0	0	0	0
	(0.0001)	(0.0000)	(0.0000)	(0.0001)	(0.0001)	(0.0001)
Size	-0.0127***	-0.0065***	-0.0089***	-0.0104***	-0.0049	-0.0067**
	(0.0032)	(0.0018)	(0.0019)	(0.0037)	(0.0034)	(0.0030)
Inst. Researchers	0.0739***	0.0551***	0.0620***	0.0698***	0.0501***	0.0569***
	(0.0128)	(0.0071)	(0.0083)	(0.0120)	(0.0108)	(0.0093)
Inst. Personnel	0.0002	-0.0009	-0.0011	-0.0041	-0.0051	-0.006
	(0.0028)	(0.0024)	(0.0026)	(0.0044)	(0.0044)	(0.0042)
PhD Theses	0.0317***	0.0125***	0.0220***	0.0291***	0.0147*	0.0215***
	(0.0083)	(0.0044)	(0.0053)	(0.0088)	(0.0077)	(0.0069)
Age	0.0007***	0.0005***	0.0006***	0.0007**	0.0007*	0.0006**
	(0.0002)	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Career in Lab	0.0007	0.0003	0.0006	0.0017*	0.0015	0.0021**
	(0.0008)	(0.0008)	(0.0008)	(0.0009)	(0.0012)	(0.0009)
Data Method	0.6788***	0.3892***	0.5459***	0.6247***	0.3827**	0.5250***
	(0.1639)	(0.0836)	(0.0928)	(0.2212)	(0.1797)	(0.1701)
Matter Science	0.1436	0.4788***	0.2509*	0.3491*	0.7284***	0.4367**
	(0.1463)	(0.1307)	(0.1387)	(0.1835)	(0.2450)	(0.2025)
Humanities	-2.3544***	-1.0247***	-1.6975***	-1.6688***	-0.6192	-1.2649***
	(0.3909)	(0.3900)	(0.3315)	(0.3195)	(0.5808)	(0.3870)
Life Science	1.9451***	2.3935***	3.1730***	0.179	0.2679	1.3097
(constant)	(0.5846)	(0.5486)	(0.5870)	(0.7225)	(0.9823)	(0.8303)
N	40	40	40	40	40	40
r2				0.8967	0.8246	0.8748
F				24.9818	8.8952	22.0675
rmse				0.4474	0.5305	0.4585
chi2	771.8891	710.0882	1005.548			
ll	-221.8817	-227.2589	-261.7496	-15.1811	-22.0014	-16.1665
aic	473.7634	484.5178	553.4991	60.3621	74.0028	62.3331
bic	499.0966	509.851	578.8323	85.6953	99.336	87.6662
Standard errors in parentheses		* p < 0.1	** p < 0.05	*** p < 0.001		

Furthermore, one may also notice that in the present regressions there are significant effects from other variables such as size effects, which are negative with an associated decrease of -1.03% expected outcome in model 1 and -0.66% in model 3 (although no significant size effects are found in the case of model 2; this confirms the results from the previous subsection (funding on contributions to publications) where decreasing returns to scale were found; these findings are in accord with the notion that the smallest research laboratories are the most productive (Adams and Grilliches, 1996; Bonaccorsi and Daraio, 2003).

Moreover, positive effects were found for the institutional affiliation of researchers with an increase of +7.2%, +5.13% and +5.85% expected publications shares for models 1, 2 and 3 respectively. The overall defended theses over a period of four years declared in the ministry surveys also present significant positive effects with an increase of +29.5%, +14.8% and +21.7% expected shares across our three models.

In addition, one may notice that positive average age effects are found in all three models of aggregation, with associated increases in the expected outcome of +0.07%, +0.07% and +0.06% in the case of models 1, 2 and 3 respectively, as well as positive average career effects in the case of models 1 and 2 for which additional years of experience in the laboratory present a respective increase of +0.17% and +0.21% expected publications shares; which indicates that as researchers grow older and obtain more experience within a given laboratory, this laboratory benefits from their seniority with increased outcome.

These results are follow the idea that the average age of researchers in laboratories has an influence on the attractiveness of the institution and its scientific prestige which induces a virtuous cycle in which young researchers are attracted to the institution (Bonaccorsi and Daraio, 2003). Finally, as expected, laboratories in the human and social sciences present a much lower expected production with respect to those laboratories in the life sciences, while those in matter sciences present higher expected shares.

2.3 Influence of laboratory characteristics on research quality

The objective of the present section is to study the effects of laboratory composition and funding on the quality of the research output performed by researchers affiliated to laboratories of the university. For this purpose, based on Mairesse and Turner (2002) I used the average and the median impact factor of all peer-reviewed journals in which members of a laboratory published their research as a measure of research quality. These average and median impact factors were defined according to the three assumptions on publication aggregation I made based on the relationship between researcher-turnover in the laboratories and their output, which were declared at different points in time and needed certain assumptions on their accountability[15].

Similarly to the analysis of the determinants of laboratory output, the analysis of research quality in the laboratory is based on an empirical study on the effects of laboratory composition in terms of human resources, on the one hand, and of different types of funding on the other; although in addition to those explanatory variables mentioned in the analysis of scientific production, I decided to include in the present analysis the effects of an aggregated lagged scientific production of laboratories and the respective lagged quality to take into account the dependencies of research quality of these organizations on their past performance reflecting learning effects.

This lagged production is defined by the publications signed by members of the laboratory during a period covered in between five years to two years prior to the ministry surveys and the respective average and/or median impact factor of journals in which they were published, therefore creating an aggregation of research production and quality over a period of three years lagged by two years prior to the surveys.

In fact, a parallel may be established with Arthur (1994) who stresses the idea that technology and social context generate increasing returns. He states that knowledge gained in the operation of complex systems leads to higher returns from using it; therefore, I may state that as scientists deepen their research works, repetition and learning effects allow them to produce new output more effectively increasing the return of research laboratories.

The dependent variables in this analysis, the average/median impact factor of laboratory research, as depicted in the Set of Figures 13, present the characteristics of continuous non-normal variables[16], with a slight positive skewedness. These characteristics suggest

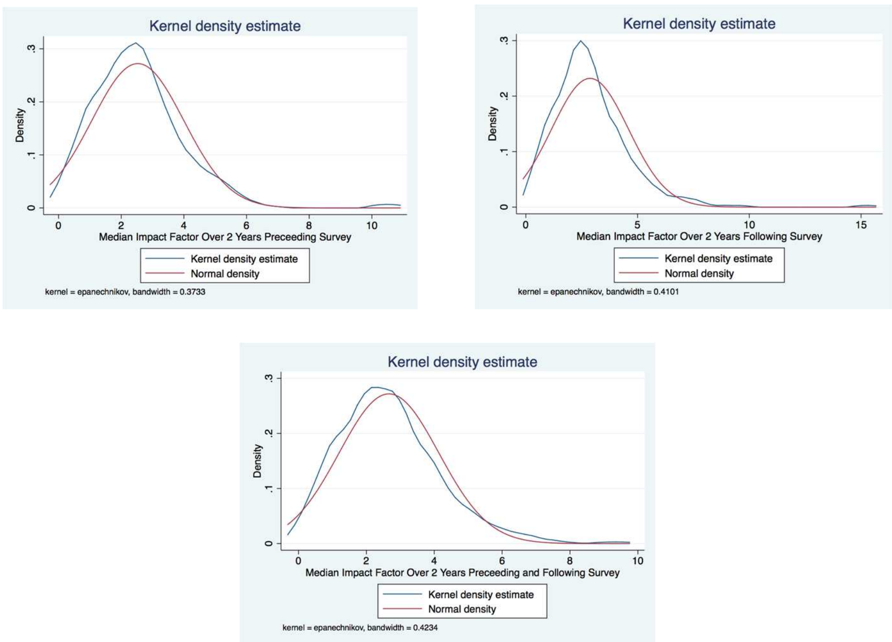
[15] Let us recall that independent variables were measured according to three different assumptions:

- The personnel declared in a given survey have had an important influence on the laboratory contributions and publications during the previous two years.
- The personnel declared in a survey have an influence on the contributions and publications during the two years following it.
- The personnel declared in a survey have an influence on the contributions and the publications of the laboratory during a range of two years before to two years after the survey.

[16] According to Shapiro-Wilk tests performed on these variables, the null hypothesis of normality is rejected for all of them at the 1% threshold.

that research quality should be regressed on the explanatory variables discussed above using a log-linear model of the type $y_{it} = \mu_{it} = exp(x_{it}'ß)\varepsilon_{it}$ defined as follows:

$$\begin{aligned}
&\textit{Average/Median Impact Factor}_{i,t} = \\
&exp\,(\,ß_1 + ß_2\,\textit{Senior Researchers}_{i,t} + ß_3\,\textit{Junior Researchers}_{i,t} \\
&+ ß_4\,\textit{Doctoral and Postdoctoral Researchers}_{i,t} + ß_5\,\textit{Staff Personnel}_{i,t} \\
&+ ß_6\,\textit{Institutional Researchers}_{i,t} + ß_7\,\textit{Institutional Personnel}_{i,t} + ß_8\,\textit{Defended Theses}_{i,t} \\
&+ ß_9\,\textit{Age}^2_{i,t} + ß_{10}\,\textit{Career}^2_{i,t} + ß_{11}\,\textit{Lagged Contributions}_{i,t} + ß_{12}\,\textit{Lagged Quality}_{i,t} \\
&+ ß_{13}\,\textit{Matter Science}_{i,t} + ß_{14}\,\textit{Humanities}_{i,t} + ß_{15}\,\textit{OST data collection method}_{i,t}\,)\,\varepsilon_{i,t}
\end{aligned}$$



Set of Figures 13 Density of median impact factors against normal density

Please note that the results discussed in this section will focus on the effects of laboratory composition and funding on the median impact factor of the research performed, although both average and median impact factors were modelled. For simplicity reasons I decided to focus on the results on the median impact factor for it seems to be a more informative measure; indeed in the present case, the aggregated impact factor of the publications in which the university researchers publish presents positive skewedness with its median being lower than its average. Therefore, the aggregated impact factor is asymmetric, with fewer publications being published in journals with higher impact factors.

2.3.1 Effects of laboratory composition on research quality

On the present subsection, I study the effects of laboratory composition on the median impact factor of publications; the results from the analysis, reported on Table 6, show that quality of research has not much to do with laboratory composition but rather laboratory past performance.

In the first model (publications 2 years prior to the survey), senior researchers display negative and significant effects with an associated decrease of -1.78% in the expected median impact factor of the publications in which a member of the laboratory participates, whereas no other effects from any other kind of researcher proves to be significant in this model. This result suggests that research quality in a laboratory presents decreasing returns to scale in the category of senior researchers, which is counterintuitive given that one may think that as more experienced researchers work together the expected quality should be greater. Compared with the results of the previous section, where I studied the effects of laboratory composition on collective output, this particular result opens the question of whether the scientific output in quantity is merely driven by senior researchers because of reputation and management features, while the scientific quality is actually driven by the efforts of technical staff (PhD’s and post doctors).

On the other hand, the quality of the publications to which members of the laboratory contribute are clearly influenced by the presence of institutional researchers in the laboratory with an associated increase of +1.67% expected median impact factor of the laboratory publications for each additional researcher affiliated to the CNRS, INSERM or other research institution different than the university; these results confirm that laboratories with more research oriented scientists perform better than laboratories with more scientists split between teaching and researching activities. In addition institutional non-research personnel present the opposite effects, with an associated slight decrease of -0.47% in the expected median impact factor for each of such additional staff. We also found positive and significant average age effects in the laboratory with an associated increase of +0.03% in the expected median impact factor for each additional year of average age, which means that as researchers forming laboratory teams grow older, the quality of the collective research increases.

It is important to highlight that research quality does depend on past performance. Indeed, we found in the present model significant effects from the past scientific production measured by the sum of contributions to scientific publications during the three years preceding a lag of two years counted from the year of ministry survey. This lagged sum of contributions presents positive effects with an associated increase of +0.05% in the current median impact factor of the laboratory for higher counts of contributions produced during the period *t-5* to *t-2* years counted from the ministry survey. Furthermore, the quality of these lagged contributions also presents positive and significant effects on the current median impact factors with an associated +29.04% higher current median impact factor for each point of past median factor reached. This particular result provides evidence that the trajectory of scientific quality, in our case measured by the impact factor of peer-reviewed journal in which the output is published, is path dependent. This can be interpreted as a process in which choices are made researchers regarding the journals in which they want to have their output published which boosts their efforts eventually succeeding in their objectives of publishing in journals with a given impact factor; if the process is dynamic we may infer that member of the laboratory may wish to maintain the level of quality of their research. The process we just described presents a sensitive dependence on initial conditions where information is imperfect, that is, the choice of journals may not always be optimal and it will be impossible to know if a chosen trajectory is inferior to another possibility. As choosing to publish in journals of higher impact factor is associated with increased future quality we may infer that scientific quality falls into the category of second degree path dependence (Liebowitz and Margolis, 1995) where sensitive dependence on initial conditions leads to outcome that may not be inefficient given the limitations on prior knowledge of choices.

Concerning our second model where the laboratory characteristics influences the scientific quality of output performed during the two years following the ministry survey we found significant effects from two kinds of scientific researchers. In fact, junior researchers present significant negative effects on research quality with an associated decrease of -1.58% median impact factor points for each additional junior researcher in the laboratory. This result is rather expected given that junior researchers rarely count with a strong capital of reputation, compared with senior researchers, and this may explain why increasing the number of juniors may affect the aggregated impact factor indicator of research laboratories.

On the other hand, it is interesting to find that Phd's and postdoctors present significant positive effects on quality with an associated increase of +0.33% median impact factor points for an additional young researcher. A possible explanation for such result is that research assistants work with senior researchers and benefit from their capital of reputation.

Institutional effects on quality from scientists affiliated to a public research organization like CNRS or INSERM were found, with an associated semi-elasticity of +1.62% median impact factor points for each institutional researcher in the laboratory. On the other hand, the effects of institutional non-research personnel displays a negative effect with an associated semi-elasticity of -0.35% impact factor points for each of these staff. These findings confirm the notion according to which research laboratories benefit from full time researchers rather than part time researchers who have to trade between teaching and research activities.

In addition, previous research production and quality measured in lagged contributions to publications and their respective median impact factors also present significant and positive effects on the current quality of laboratories. Their associated semi-elasticities show there is +0.06% higher median impact factor points for higher past production and +22.27% for better past quality.

Regarding our third model, according to which laboratory characteristics and past performances has an impact on the scientific production and quality during the four years around the ministry survey; we found that all three categories of scientific researchers present significant effects on research quality. We may notice here how both senior and junior researchers present negative impacts with an associated decrease of -1.15% and -1.16% median impact factor point for each additional senior and junior researcher respectively, while PhD's and postdoctors present positive effects with an associated increase of +0.3% median impact factor points for each additional young researcher. I draw your attention on the fact that both senior and junior researcher turned out to have negative effects on research quality, which seems to be counter intuitive; one may point out that accessing a higher status within the academic system does not necessarily mean performing better. Not all senior researchers produce outstanding science, and this proposition calls for further investigation on the scientific quality of these individuals.

These are contrasting results that let us imagine that as senior researchers become distracted with management activities targeted towards raising competitive funding their younger collaborators, meaning PhD students and post doctors, follow their guidelines and eventually become responsible for the research carried within the team at a technical level. This proposition could be tested if only we had nominative information on the PhD’s and post doctors and the publications they sign.

The researcher institutional affiliation also presents significant positive effects on research quality with an increase of +1.62% median impact factor points for each affiliation to the CNRS, the INSERM or other public research institution different than the university, which confirms that research laboratories benefit more from researchers who are committed to perform research on a full time basis rather that from researchers who have to share their time between teaching and researching activities; however, as noticed in the previous models, non-researcher personnel affiliated to a research institution present significant negative effects on research quality.

Table 6 Regression results for research quality explained by laboratory composition

Explanatory Variables	Median Impact Factor over 2 years preceding the survey	Median Impact Factor over the 2 years following the survey	Median Impact Factor from 2 years before to 2 years after survey	Average Impact Factor over 2 years preceding the survey	Average Impact Factor over the 2 years following the survey	Average Impact Factor from 2 years before to 2 years after survey
Senior	-0.0180**	-0.0061	-0.0116*	-0.0145**	-0.0037	-0.009
	(0.0079)	(0.0081)	(0.0068)	(0.0068)	(0.0074)	(0.0065)
Junior	-0.0029	-0.0160***	-0.0117**	-0.0078*	-0.0168***	-0.0130***
	(0.0055)	(0.0058)	(0.0048)	(0.0044)	(0.0045)	(0.0038)
PhD and Postdoc	0.0027	0.0033**	0.0030**	0.0011	0.0005	0.0011
	(0.0018)	(0.0016)	(0.0015)	(0.0018)	(0.0018)	(0.0016)
Staff	0.0045	0.0234	-0.0153	-0.0227	0.0315	-0.0181
	(0.0285)	(0.0198)	(0.0243)	(0.0280)	(0.0206)	(0.0236)
Inst. Researchers	0.0166***	0.0161**	0.0161***	0.0157***	0.0152***	0.0150***
	(0.0058)	(0.0064)	(0.0057)	(0.0048)	(0.0057)	(0.0049)
Inst. Personnel	-0.0048***	-0.0036***	-0.0040***	-0.0044***	-0.0028***	-0.0036***
	(0.0011)	(0.0012)	(0.0011)	(0.0011)	(0.0010)	(0.0010)
PhD Theses	-0.0026	-0.0061	-0.0057	-0.0009	-0.0012	-0.0019
	(0.0042)	(0.0038)	(0.0038)	(0.0040)	(0.0032)	(0.0032)
Age	0.0003**	0.0001	0.0001	0.0002**	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Career in Lab	0	0	0	0	0	0.0001
	(0.0002)	(0.0003)	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Lagged Publications	0.0005*	0.0006**	0.0007***	0.0006**	0.0005**	0.0006***
	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0003)	(0.0002)
Lagged Median Impact	0.2550***	0.2011***	0.2262***			
	(0.0288)	(0.0249)	(0.0255)			
Lagged Average Impact				0.2132***	0.1720***	0.1880***
				(0.0170)	(0.0217)	(0.0167)
Data Method	0.1380*	0.2729***	0.2460***	0.2221***	0.2242***	0.2233***
	(0.0818)	(0.0762)	(0.0700)	(0.0640)	(0.0691)	(0.0601)
Matter Science	-0.1663**	-0.1031*	-0.1332**	-0.1352**	-0.0572	-0.1174**
	(0.0654)	(0.0604)	(0.0530)	(0.0594)	(0.0648)	(0.0542)
Humanities	-0.5082**	-0.3817**	-0.2051	-0.2851*	-0.2006	-0.1315
	(0.2155)	(0.1763)	(0.1608)	(0.1497)	(0.1589)	(0.1362)
Life Science	-0.3511	0.2612	0.1108	-0.0397	0.3811	0.3042
(constant)	(0.3046)	(0.2794)	(0.2416)	(0.2482)	(0.2774)	(0.2358)
N	167	171	171	167	171	171
r2	0.7045	0.688	0.7473	0.7748	0.6906	0.7768
F	21.8971	28.7789	29.1558	30.39	20.2403	28.6391
rmse	0.3478	0.3343	0.2994	0.2944	0.3197	0.2734
chi2						
ll	-52.749	-47.4345	-28.5643	-24.8936	-39.7813	-13.0614
aic	135.498	124.869	87.1287	79.7871	109.5626	56.1227
bic	182.2679	171.994	134.2536	126.557	156.6876	103.2477
Standard errors in parentheses		* p < 0.1	** p < 0.05	*** p < 0.001		

One may also notice that past research and quality present positive significant effects on the current quality with an associated semi-elasticity of +0.07% median impact factor points for higher contribution counts produced in the past and +25.38% if such research reached higher median impact factor points as well. These results confirm that the quality of scientific output in research laboratories is path dependant and that learning effects are present in the process.

Finally, attention is drawn on the fact that both senior and junior researchers present negative effects on research quality, implying that the status within the academic system does not necessarily affect the quality of scientific research. Not all senior researchers produce outstanding science and compared with results from previous sections we may infer that quantity of publications primes over quality of publications; this calls for further investigation on the behavior of trade off between producing more and producing better for some young and senior researchers.

2.3.2 Effects of laboratory funding on research quality

We now focus our attention on the effects of different types of research funding on the median impact factor of the university laboratories. Following an analogous analysis as in the previous section, the results from the first model, reported on Table 7, shows that regional and private funding present positive significant effects on research quality with an associated semi-elasticity of +0.16% and +0.03% median impact factor points for each thousand additional Euros of allocated under each of these forms respectively.

However, public funding seems to present negative significant effects implying that higher public fund allocations are associated with lower quality of scientific production. It is interesting to notice that in the previous section on the influence of funding sources on scientific production we had found that public funds presented positive effects, implying that as public funds available to the laboratory increase, the volume of scientific publications increases as well; those results contrast with results in the present section on the influence of funding sources on the quality of scientific publications mentioned above, which tells us that even though this type of public funding help laboratories produce more, it does not necessarily help them perform better. These results let us believe that national public funding is associated with more output but of rather poor quality. The contrast between the effects of private and public funding indicate that while national public funding boosts the research output, the research quality is boosted by private competitive sources of funding.

Table 7 Regression results for research quality explained by laboratory funding and other characteristics

Explanatory Variables	Median Impact Factor over 2 years preceding the survey	Median Impact Factor over the 2 years following the survey	Median Impact Factor from 2 years before to 2 years after survey	Average Impact Factor over 2 years preceding the survey	Average Impact Factor over the 2 years following the survey	Average Impact Factor from 2 years before to 2 years after survey
Regional Funds	0.0016**	0.0012	0.0011	0.0014**	0.0007	0.0008
	(0.0007)	(0.0008)	(0.0008)	(0.0006)	(0.0007)	(0.0008)
European Funds	-0.0001	-0.0002**	-0.0001*	-0.0001	-0.0001*	-0.0001*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Private Funds	0.0003**	0.0003***	0.0003***	0.0003**	0.0002*	0.0002*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Public Funds	-0.0002*	-0.0002**	-0.0002**	-0.0003**	-0.0001	-0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Recurrent Public Funds	0	0.0001	0.0001	0.0001	0.0001*	0.0001*
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Size	-0.0015	0.0019	0.0004	-0.0002	-0.0017	-0.0008
	(0.0021)	(0.0030)	(0.0026)	(0.0028)	(0.0024)	(0.0025)
Inst. Researchers	0.0129	-0.001	0.0014	-0.0022	0.0096	-0.0013
	(0.0124)	(0.0164)	(0.0153)	(0.0147)	(0.0147)	(0.0164)
Inst. Personnel	-0.0004	-0.0031	-0.0021	-0.002	-0.0017	-0.0022
	(0.0028)	(0.0041)	(0.0035)	(0.0032)	(0.0034)	(0.0034)
PhD Theses	0.0013	-0.0132*	-0.0079	-0.0027	-0.0056	-0.0056
	(0.0052)	(0.0075)	(0.0067)	(0.0067)	(0.0056)	(0.0061)
Age	0.0003**	0.0001	0.0002	0.0004*	0.0002	0.0002
	(0.0001)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Career in Lab	-0.0002	0.0002	0	0.0007	0	0.0004
	(0.0008)	(0.0010)	(0.0009)	(0.0009)	(0.0009)	(0.0010)
Lagged Publications	-0.0006	0.0002	0.0001	0	0.0001	0.0004
	(0.0006)	(0.0007)	(0.0007)	(0.0006)	(0.0007)	(0.0008)
Lagged Median Impact	0.3312***	0.2437***	0.2911***			
	(0.0537)	(0.0829)	(0.0757)			
Lagged Average Impact				0.2686***	0.2153***	0.2328***
				(0.0527)	(0.0549)	(0.0599)
Data Method	-0.0996	-0.0186	-0.0144	-0.0766	0.1464	0.077
	(0.1581)	(0.2055)	(0.1842)	(0.1572)	(0.1827)	(0.1908)
Matter Science	-0.0046	-0.0714	0.0533	-0.0268	-0.0595	-0.0339
	(0.1412)	(0.1767)	(0.1703)	(0.1665)	(0.1746)	(0.1983)
Humanities	-0.2271	-0.5112	-0.1709	0.0044	-0.0105	0.0049
	(0.2160)	(0.4001)	(0.3569)	(0.2542)	(0.3220)	(0.3411)
Life Science (constant)	-0.6474*	0.1237	-0.3473	-0.6044	0.2197	-0.1115
	(0.3564)	(0.5346)	(0.4835)	(0.5203)	(0.6013)	(0.6370)
N	39	40	40	39	40	40
r2	0.8946	0.8474	0.8714	0.8834	0.8398	0.8381
F	89.0342	13.9831	17.9855	92.8351	35.5003	36.7271
rmse	0.2623	0.3118	0.2859	0.2952	0.3012	0.3212
chi2						
ll	8.0144	0.9218	4.393	3.4131	2.3036	-0.2596
aic	17.9713	32.1565	25.2139	27.1738	29.3927	34.5193
bic	46.2518	60.8674	53.9249	55.4543	58.1037	63.2302
Standard errors in parentheses		* p < 0.1	** p < 0.05	*** p < 0.001		

Furthermore, we also found positive significant age and past research quality effects that reflect an increase of +0.03% expected median impact factor points in laboratories with higher average age and +39.26% for better research performed in the past.

Regarding our second model, results show that only private funds have a positive significant effect on research quality, with an associated increase of +0.03% in the expected median impact factor points for each additional thousand Euros allocated under this type of funding.

On the other hand, we learn that both public and European funding present negative significant effects on the median impact factor, both with associated semi-elasticities of -0.02% points for each additional thousand Euros allocated under each of these sources respectively. As expected, the quality although not the quantity of past research performed by members of the laboratory presents positive significant effects with associated higher median impact factors (+27.59%) for higher quality research performed during the period of *t-5* to *t-2* years prior to the ministry surveys.

Finally, our third model, corresponding to the assumption that laboratory characteristics have an influence on the scientific production during four years around the declaration in the ministry survey, shows similar results reflecting positive significant effects from private funding and negative significant effects from European and public funding. This time we obtain results indicating that as public and European public funds increase, the associated scientific output presents lower quality, while an increased availability of private funds is associated with higher quality of scientific research, in essence, if public and European funding allow larger volumes of scientific output, they do not stimulate the quality of this one. The associated increase related to any additional thousand Euros of private allocations amounts up to +0.03% median impact factor points, while the associated decrease related to additional an thousand Euros of European or public funding corresponds to -0.01% and -0.02% points respectively. The associated impact of past quality on the other hand is positive, with an increase of +33.78% median impact factor points in the case a laboratory had had a one point higher median impact factor in its past research quality.

2.4 Conclusion

In conclusion, the present section on the collective scientific performance of research laboratories has provided insights on which determinants play an important role on the production of scientific publications and its quality at the laboratory level.

Two different types of statistical analyses were carried out in the study of scientific output:

- A Negative Binomial regression performed on contributions to scientific publications as count data.
- A Log Linear regression performed on the fractions of publications as continuous data.

In addition, a Log Linear analysis was carried out on the average and the median impact factors of journal in which the scientific output of laboratories has been published. This latter analysis has as objective the study of the determinants playing a role on the scientific quality of research laboratories.

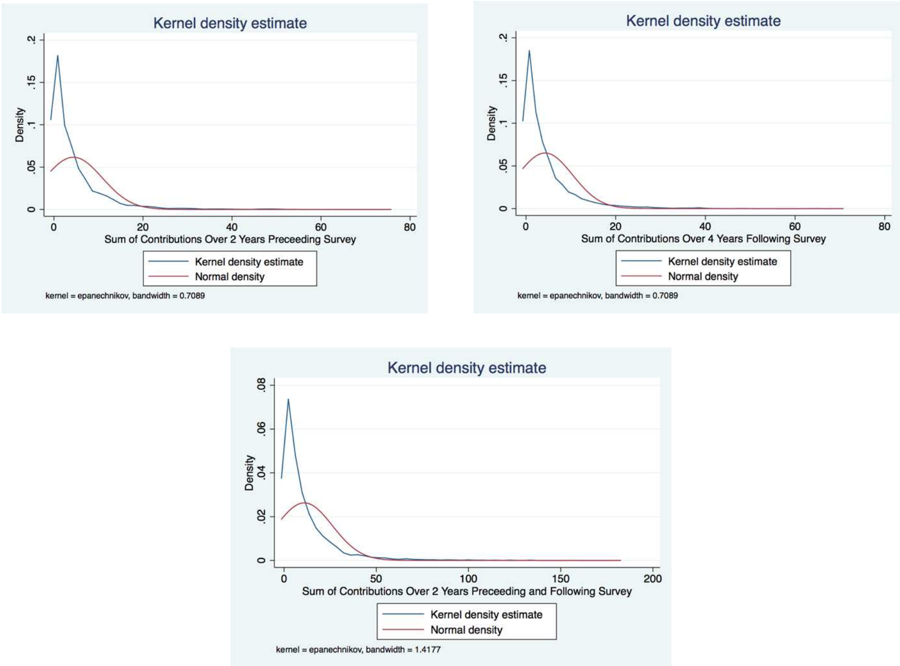
The different analyses were carried out across three main models defined by the way in which the tangible output, in this case scientific publications were aggregated across time. Please let me recall that three assumptions were formulated on this aggregation, which resulted in the following cases:

- Researchers affiliated to the laboratory during the period 1996 – 2000 may be accountable for the publications performed during a period of two years preceding the ministry survey of 2000 and 2 (model 1).
- They continued working for the laboratory during the period 2000 – 2004 and therefore may also be accountable for the laboratory publications signed during a period of two years following the ministry survey of 2000 (model 2).
- If the two previous assumptions hold, then we may also assume that researchers declared during this ministry survey may accountable for the laboratory publications signed over a period of four years around that survey (model 3).

As a general overview, I observed that results in section 1 indicate that collective scientific output is mainly driven by: senior researchers, post doctors and PhD candidates, it is enhanced by researchers who are affiliated to institutional research organization like CNRS or INSERM and is mainly influenced by public funding. In terms of collective scientific quality, these results indicate that senior and junior researchers present a negative influence on quality, which in contrast to results from the study of production would only mean that the more experienced researchers there is a trade off between quantity and quality. In addition the presence in the laboratory of researchers with an institutional affiliation different than the university plays an important positive role on collective scientific quality, while in terms of the funding structure, this quality driven by private funding which in contrast to the study on the scientific output indicates that while public funds boost quantity, private funds boost quality.

3 Scientific performance of individual researchers

In the second part of the present chapter I focus my attention on the total contributions and the fractional publication counts of individual researchers affiliated to the University Louis Pasteur during the periods covered by the ministry surveys from 1996 to 2008. I now attempt to explain the individual scientific production by the characteristics of these individual researchers, such as their right to direct scientific research, their researcher status, their age and their experience in the laboratory and also look for the effects of coauthorship with internal colleagues on individual scientific production given the decomposition of human resources into categories of rank/status; for this purpose I use the information on the types or researchers collaborating on a same publication.

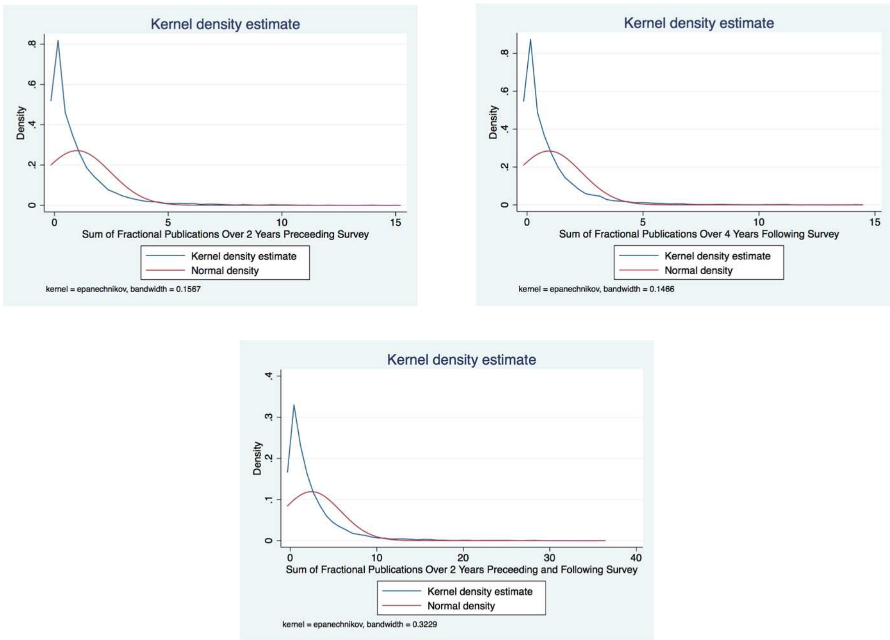


Set of Figures 14 Density of total individual contributions against normal density

In order to attain this objective I gathered the information concerning each publication in the database and then merged it with the personnel data. This procedure enabled me to identify all internal coauthors affiliated to a laboratory of the university working on a single publication. Then I separated these coauthors according to their different rank/status and defined the count variables that represent their total number.

I must clarify here that only three out of the four categories of researchers used in the previous sections can be identified using this method (senior, junior, and assistant researchers) due to the fact that the number of post doctors and PhD students appears directly aggregated on the laboratory dataset while the personnel dataset lacks information on these categories, which is the reason why it is impossible to identify them at the individual level. The origin of this problem is the fact that post doctors and PhD students were considered as temporary personnel during the ministry surveys of 1996 200 and 2004, and therefore were recorded in the individual personnel datasets of the ministry surveys covering these periods.

Similarly to the study of laboratory production in section 1, our dependent variables for the present analysis are the total contributions to publications and the total fractional publications of individual researchers during the 2 years preceding the survey (model 1), during 2 years following the survey (model 2) and during 2 years around the ministry survey (model 3); these variables, as one may observe from the Set of Figures 15, are highly non-normal and positively skewed.



Set of Figures 15 Density of total individual fractional publications against normal density

Following an analogous procedure to the one used in the previous studies on laboratory production, I performed an over-dispersion test in the case of the count variable (sum of contributions to publications). This test suggested that a negative binomial regression is better suited for the analysis of this variable; in the case of the continuous variable (sum fractional publications) a log linear analysis was chosen. Since the explanatory variables are defined by the characteristics of individual researchers and controlled by the disciplinary field, the method of publication data collection, and the size of the laboratory, our regressions may be expressed by the following expression:

$$\textit{Contributions and/or Fractional Publications}_{i,t} =$$
$$\exp(\beta_1 + \beta_2 \textit{ Habilitation to direct research}_{i,t} + \beta_3 \textit{ Junior Researcher}_{i,t}$$
$$+ \beta_4 \textit{ Assistant Researcher}_{i,t} + \beta_5 \textit{ Institution CNRS}_{i,t}$$
$$+ \beta_6 \textit{ Institution INSERM}_i + \beta_7 \textit{ Other Institution}_i$$
$$+ \beta_8 \textit{ Age}^2_{i,t} + \beta_9 \textit{ Career}^2_{i,t} + \beta_{10} \textit{ Matter Science}_{i,t} + \beta_{11} \textit{ Humanities}_{i,t}$$
$$+ \beta_{12} \textit{ OST data collection method}_{i,t})\, \varepsilon_{i,t}$$

Table 8 Correlations among individual characteristics

Dependent Variable Corr.	Habilit.	Indiv. Senior	Indiv. Junior	Indiv. Assist.	Univ.	CNRS	INSERM	Other Institution	Age	Career
Habilitation	1									
Individual Senior	0.6014	1								
Individual Junior	0.5797	-0.9842	1							
Individual Assistant	0.1136	-0.0748	-0.1028	1						
University	0.0763	-0.0008	0.0121	-0.0639	1					
CNRS	0.1229	0.0208	-0.0082	-0.0709	-0.7807	1				
INSERM	0.0275	-0.0001	0.0043	-0.0231	-0.2548	-0.2099	1			
Other Institution	0.1224	-0.0416	-0.0133	0.3091	-0.2413	-0.1988	-0.0649	1		
Age	0.562	0.4969	-0.4906	-0.0286	-0.0486	0.0724	0.0032	-0.0507	1	
Career	0.2507	0.2287	-0.2206	-0.0422	-0.0461	0.0991	-0.0394	-0.0671	0.4455	1

The correlations among these individual researcher characteristics, reported in Table 8, show that the status of senior researcher presents a positive and strong relationship with the habilitation to direct research and with an institutional affiliation to the CNRS, the status of junior researcher is rather associated with the habilitation to direct research (with lower strength than seniors as expected), while assistant researchers are only weakly related to the habilitation to direct research and present a negative relation with the institutional affiliation to either CNRS or INSERM. As one may expect, average age and career in the laboratory showed positive relationships with the status of senior researchers, as well as the affiliation to the CNRS.

3.1 Individual researcher statuses and their influence on individual contributions

As discussed in previous sections, the first step of the count data analysis of individual contributions is to perform a Poisson regression and assess whether or not the specification is appropriate by performing a test of over-dispersion of the dependent variable. However, in order to assess the validity of the Poisson specification, we perform a test of over-dispersion of the dependent variable to corroborate whether the assumption of equi-dispersion of the Poisson model holds. The results from this test reveal the existence of over-dispersion at the 1% threshold and indicate that an alternative regression should be performed.

Overdispersion test	Model1	Model 2	Model 3
	Sum of contributions over 2 years preceding the survey	Sum of contributions over the 2 years following the survey	Sum of contributions from 2 years before to 2 years after survey
Overdispersion Statistic	1.1621***	1.1513***	1.1221***
	(0.0887')	(0.0802')	(0.0762')
a is significantly positive at the 1% threshold therefore implying overdisperion			

I use the category of senior university researchers in life sciences as the baseline and find that in terms of status within the French academic system, being a junior researcher is significantly associated with lower contribution counts. In addition, having the habilitation to direct research also presents positive significant effects on the individual expected contributions as well as the institutional affiliation of the researcher; these characteristics present positive effects from the affiliation to the CNRS or to the INSERM. Finally, I found significantly negative age effects and significantly positive career effects as may be observed in the results reported in Table 9. Such results are in accord with the notion that individual age effects are decreasing (Diamond, 1986) given that the expected returns of human capital are decreasing with the remaining activity period within the life cycle.

In the case of the first model, which is associated with the assumption that the scientific work carried by an individual researcher, corresponding to a ministry survey can be aggregated by the sum of his contributions during two years preceding the survey declaration, the results show that junior researchers are associated with a -50.7% lower count of contributions with respect to senior researchers. Having passed and obtained the habilitation to direct research presents positive and significant effects with a +55.7%

higher expected output with respect to non-habilitated researchers. Being associated with a research institution, CNRS or INSERM, presents very similar effects with an associated increased count of +56% to +60% of expected contributions. As researchers grow older, they produce slightly fewer contributions with an associated decrease of -0.019% expected contributions; however but this effect is evicted by a positive and significant career/experience effect in the laboratory associated with an increase of +0.02% expected contributions.

Regarding the results for the second model, where I suppose that the scientific work of a declared researcher is reflected in the sum of his contributions during two years following the ministry survey, I found that junior researchers have significant lower expected contributions counts of -52.7% with respect to senior researchers, as well as the habilitation, which indicates a +51.4% higher expected outcome. The institutional affiliation of the researcher also has a positive significant effect with +53.6% higher expected output in the case of CNRS researchers, and +45.5% in the case of INSERM researchers. Age effects are negative and associated with -0.029% lower expected output as the researcher grows a year older, and career/experience effects are positive with an associated increase of +0.02% higher expected output as the gains a year of experience.

Concerning the third model associated with the assumption that the scientific work of individual researchers declared in the different ministry surveys is reflected by their contributions during a period of four years around the official declaration, I found very similar results to those found in the case of the other two models, with positive effects of the habilitation to direct research associated with +54% higher contributions for those researchers who have obtained it, -52.4% lower expected contributions of junior researchers with respect to senior ones, and +50.4% to +57.2% higher expected outcome if the researchers is affiliated to the INSERM or the CNRS respectively.

Finally, let us notice that researchers working in the social and human sciences present much lower expected contributions with respect to those in the life sciences – -88.6% for model 1, -83.3% for model 2 and -86.4% for model 3–, while those researchers working in the matter sciences also present significant and negative effects in the case of two models, with lower expected contributions of -8.18% for model 1 and -7.68% for model 3.

3.2 Individual researcher statuses and their influence on individual fractional publications

In addition to the analysis on the sum of contribution to scientific publications, I performed an analogue analysis on the sum of actual fractional publications in order to model the effects of individual researcher characteristics on the sum of fractional publications. I studied the three different assumptions on the aggregation of publications subject to the declarations in the ministry surveys and found that in the case of our first model, the habilitation to direct research has positive effects on the expected output with +39.5% higher contributions counts with respect to a non-habilitated researcher. Being a junior researcher has a negative impact with -48.5% associated lower expected output with respect to a senior researcher. In this model I also found that being affiliated to the CNRS implies +11.5% more expected article shares with respect to other university researchers, while being affiliated to the INSERM implies +28.5% more.

In the case of the second assumption, we found similar effects from junior researchers, that is a lower expected outcome of -49.7% with respect to that of senior researchers, while the habilitation presents an effect of +34.9% higher expected article shares. The institutional affiliation of the individual researcher also presents significant and positive effects, with an impact of +14.13% higher expected outcome from researchers affiliated to the CNRS with respect to those affiliated to the university and an impact of +24.24% higher expected outcome for those affiliated to the INSERM; it is worth noticing that for the first time I find significant effects from researchers affiliated to a public research institution different than CNRS and INSERM, with a negative impact of -16% lower expected article shares, meaning that within the landscape of French research institutions, only CNRS and INSERM have the power to affect the output and therefore career of individual researchers.

Concerning the third assumption, the results show a similar pattern, with positive effects from the habilitation to direct research (+49.5% expected outcome), negative effects for being a junior researcher (-53.7% expected outcome with respect to senior researchers); we may also notice negative and significant effects from assistant researchers were found in this model with an impact of -33.7% lower expected articles shares for assistant researchers with respect to senior researchers. In terms of institutional affiliation, I found that being affiliated to the CNRS is associated with an increase of +31.36% expected shares, while being affiliated to the INSERM is associated with an increase of +50.9% expected shares. In addition, there are significant and negative age and career effects that with associated lower expected outcome (across all three models) as individuals grow older and gain experience.

Table 9 Regression results of individual output explained y individual characteristics

	Model 1 NB	Model 2 NB	Model 3 NB	Model 1 LogLinear	Model 2 LogLinear	Model 3 LogLinear
Explanatory Variables	Sum of contributions over 2 years preceding the survey	Sum of contributions over the 2 years following the survey	Sum of contributions from 2 years before to 2 years after survey	Sum of fractional publications over 2 years preceding the survey	Sum of fractional publications over the 2 years following the survey	Sum of fractional publications from 2 years before to 2 years after survey
Habilitation	0.4433***	0.4147***	0.4319***	0.3330***	0.2997***	0.4025***
	(0.0575)	(0.0512)	(0.0510)	(0.0426)	(0.0454)	(0.0410)
Junior	-0.7300***	-0.7469***	-0.7423***	-0.6646***	-0.6878***	-0.7704***
	(0.0484)	(0.0467)	(0.0454)	(0.0402)	(0.0419)	(0.0385)
Assistant	-0.4352	-0.0936	-0.2577	-0.0862	-0.2449	-0.4120*
	(0.3003)	(0.2365)	(0.2446)	(0.1840)	(0.2261)	(0.2211)
CNRS	0.4476***	0.4295***	0.4528***	0.1094***	0.1322***	0.2728***
	(0.0419)	(0.0389)	(0.0385)	(0.0336)	(0.0346)	(0.0326)
INSERM	0.4757***	0.3748***	0.4085***	0.2510***	0.2171***	0.4117***
	(0.0595)	(0.0640)	(0.0556)	(0.0572)	(0.0639)	(0.0551)
Other Institution	0	0.0301	-0.0268	-0.0782	-0.1770*	-0.1471
	(0.1168)	(0.1295)	(0.1184)	(0.0939)	(0.1019)	(0.0913)
Age	-0.0002***	-0.0003***	-0.0002***	-0.0001***	-0.0002***	-0.0002***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Career	0.0002***	0.0002**	0.0002***	-0.0001**	-0.0002***	-0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Data Method	-0.0534	-0.0722*	-0.0583	-0.0603	-0.0533	0.0071
	(0.0457)	(0.0430)	(0.0416)	(0.0387)	(0.0372)	(0.0359)
Matter Science	-0.0855**	-0.0547	-0.0800**	0.0986***	0.1654***	0.0792**
	(0.0393)	(0.0389)	(0.0369)	(0.0333)	(0.0345)	(0.0325)
Humanities	-2.1740***	-1.7934***	-1.9977***	-0.1815*	-0.1581*	-0.6066***
	(0.1671)	(0.1379)	(0.1234)	(0.1017)	(0.0892)	(0.0834)
Life Science (constant)	1.7211***	2.2042***	2.9126***	0.1729**	0.4477***	1.1534***
	(0.0843)	(0.0817)	(0.0783)	(0.0673)	(0.0705)	(0.0655)
N	5163	5163	5163	3756	3850	4385
r2				0.1398	0.1221	0.1829
F				59.9723	50.3704	94.7918
rmse				0.9484	1.0002	0.9911
chi2	1209.2766	1014.5178	1376.7017			
ll	-12450	-12880	-16970	-5124.709	-5457.8021	-6176.6644
aic	24924.9345	25793.0517	33967.5855	10273.418	10939.6043	12377.3288
bic	25003.5258	25871.643	34046.1768	10348.1913	11014.6742	12453.9602
Standard errors in parentheses		* p < 0.1	** p < 0.05	*** p < 0.001		

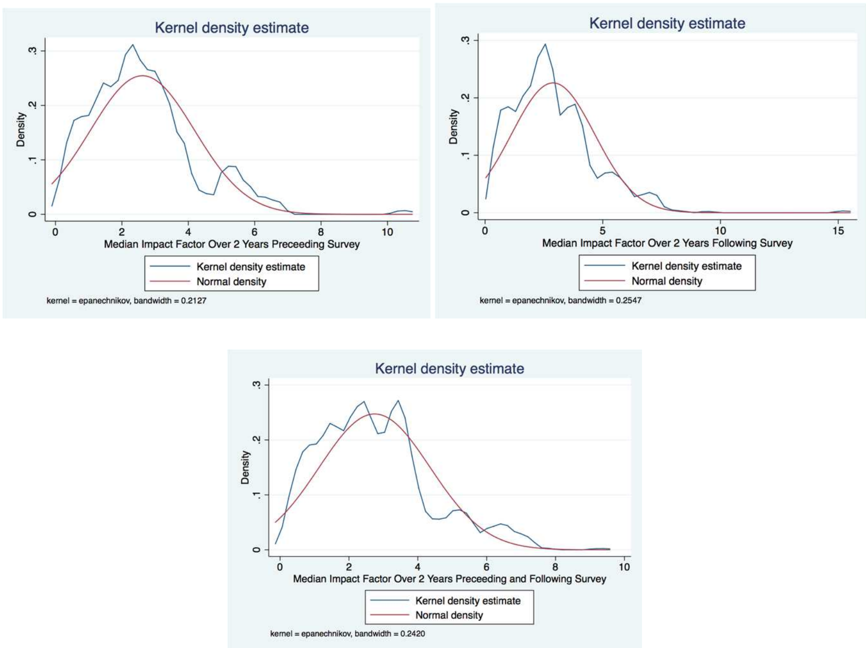
Finally, in terms of the disciplinary field in which researchers evolve, I must point out that for all three models, working in the human sciences presents significant negative impacts with lower expected article shares (-16.5% for model 1, -14.6% for model 2 and -45.4% for model 3) with respect to researchers in the life sciences, while working in the matter sciences seems to present positive and significant effects with higher expected outcome (+10.36% for model 1, +17.9% for model 2 and +8.2% for model 3).

3.3 Influence of individual researcher characteristics on research quality

The analysis of individual researcher characteristics on research quality is carried on the median and on the average impact factor of journals in which the individual researchers published their work. The dependent variables, which we may observe in the Set of Figures 16, are continuous and present positive skewedness and similarly to the case of collective research in the laboratory, they prove to be non-normal through a rejection of the null hypothesis of a normality test. In this subsection I study the median and the average impact factor and explain them by the characteristics of the individual researcher, augmented by the lagged research performance of the individual researcher in order to take into account the effects of the knowledge base he/she previously developed. The model I use may be detailed in the following expression.

$$\begin{aligned}
&\textit{Median/Average Impact Factor}_{i,t} = \\
&\exp\big(\text{ß}_1 + \text{ß}_2\,\textit{Habilitation to direct research}_{i,t} + \text{ß}_3\,\textit{Junior Researcher}_{i,t} \\
&+ \text{ß}_4\,\textit{Assistant Researcher}_{i,t} + \text{ß}_5\,\textit{Institutional CNRS}_{i,t} \\
&+ \text{ß}_6\,\textit{Institutional INSERM}_{i,t} + \text{ß}_7\,\textit{Other Institution}_{i,t} \\
&+ \text{ß}_8\,\textit{Age}^2_{i,t} + \text{ß}_9\,\textit{Career}^2_{i,t} + \text{ß}_{10}\,\textit{Lagged Contributions}_{i,t} + \text{ß}_{11}\,\textit{Lagged Quality}_{i,t} \\
&+ \text{ß}_{12}\,\textit{Matter Science}_{i,t} + \text{ß}_{13}\,\textit{Humanities}_{i,t} \\
&+ \text{ß}_{14}\,\textit{OST data collection method}_{i,t}\big)\,\varepsilon_{i,t}
\end{aligned}$$

The results from this analysis reported in tTable 10, show that in the first model the status of junior researcher has a positive significant impact on the median impact factor of contributions signed. This status is associated with an increase of +5.14% points with respect to senior researchers implying that individual researchers produce better quality output during their junior years. This is a counterintuitive result since, based on the notion that there are learning effects and path dependencies in research activities, one may tend to think that as researchers grow older and gain experience they become better at what they do. However in this case, junior researchers are associated with higher expected median impact factors showing that research quality does not stand solely on the shoulders of senior researchers; perhaps one may turn to the debate according to which senior researchers are becoming managers distracted from research activities; this would explain why juniors, who still carry on with technical activities but also have to look for competitive funding present better output.



Set of Figures 16 Density of median impact factors of individual publications against normal density

Concerning the institutional affiliation of researchers, we found that those who are affiliated to the CNRS are more likely to perform better research with an increased of +10.68% in the expected median impact factor with respect to university researchers, while those affiliated to the INSERM present negative significant effects with respect to university researchers with a decrease of -3.94% in the expected median factor of the publications they contribute to.

Individual age and career effects were also found in this model, implying that, as a researcher grows older, the quality of the publications he contributes to slightly increases since the semi-elasticity approaches zero, while as he gains experience in the laboratory, the quality of his research is less important.

Table 10 Regression results of research quality explained by researcher characteristics

Explanatory Variables	Median Impact Factor over 2 years preceding the survey	Median Impact Factor over the 2 years following the survey	Median Impact Factor from 2 years before to 2 years after survey	Average Impact Factor over 2 years preceding the survey	Average Impact Factor over the 2 years following the survey	Average Impact Factor from 2 years before to 2 years after survey
Habilitation	0.012	0.0124	0.0095	0.0066	0.0082	0.0024
	(0.0177)	(0.0174)	(0.0160)	(0.0164)	(0.0172)	(0.0156)
Junior	0.0502***	0.0433***	0.0553***	0.0472***	0.0365**	0.0422***
	(0.0160)	(0.0162)	(0.0150)	(0.0150)	(0.0156)	(0.0144)
Assistant	0.1243	0.2515***	0.1495***	0.0759	0.2166***	0.1104**
	(0.0760)	(0.0622)	(0.0577)	(0.0619)	(0.0427)	(0.0495)
CNRS	0.1015***	0.0957***	0.0923***	0.0876***	0.0928***	0.0887***
	(0.0134)	(0.0132)	(0.0124)	(0.0127)	(0.0132)	(0.0122)
INSERM	-0.0402**	-0.0718***	-0.0512***	-0.0074	-0.0255	-0.0305*
	(0.0191)	(0.0201)	(0.0168)	(0.0176)	(0.0182)	(0.0164)
Other Institution	0.0216	0.0109	0.0185	0.0387	0.0613**	0.0348
	(0.0383)	(0.0288)	(0.0295)	(0.0342)	(0.0301)	(0.0283)
Age	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Career	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***	-0.0001***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Lagged Publications	0.0023***	0.0012**	0.0021***	0.0024***	0.0018***	0.0022***
	(0.0006)	(0.0006)	(0.0005)	(0.0005)	(0.0005)	(0.0005)
Lagged Median Impact	0.2638***	0.2309***	0.2551***			
	(0.0050)	(0.0050)	(0.0047)			
Lagged Average Impact				0.2112***	0.1875***	0.1999***
				(0.0041)	(0.0041)	(0.0039)
Data Method	0.0735***	0.1274***	0.0988***	0.1038***	0.0852***	0.0894***
	(0.0126)	(0.0133)	(0.0123)	(0.0120)	(0.0125)	(0.0118)
Matter Science	-0.1723***	-0.1686***	-0.1588***	-0.1776***	-0.1066***	-0.1447***
	(0.0142)	(0.0137)	(0.0126)	(0.0131)	(0.0136)	(0.0122)
Humanities	-0.3500***	-0.4472***	-0.2698***	-0.2864***	-0.4305***	-0.2897***
	(0.0658)	(0.0791)	(0.0558)	(0.0713)	(0.0833)	(0.0566)
Life Science	0.0673*	0.2875***	0.1461***	0.2686***	0.4607***	0.3800***
(constant)	(0.0364)	(0.0358)	(0.0337)	(0.0346)	(0.0355)	(0.0333)
N	2738	2762	2762	2738	2762	2762
r2	0.7125	0.6703	0.7311	0.7414	0.6648	0.7309
F	399.4395	340.7096	423.7747	426.7405	302.1188	403.8882
rmse	0.3224	0.3238	0.2982	0.302	0.3186	0.2914
chi2						
ll	-778.9821	-797.3457	-569.9794	-600.1114	-753.1149	-506.0389
aic	1585.9642	1622.6913	1167.9588	1228.2227	1534.2298	1040.0779
bic	1668.774	1705.6233	1250.8908	1311.0325	1617.1617	1123.0098
Standard errors in parentheses		* p < 0.1	** p < 0.05	*** p < 0.001		

Finally, both the past scientific production of a researcher and its quality present positive significant effects on the median impact factor of current publications with an associated increase of +0.23% for higher contributions signed during the period of *t-5* to *t-2* years prior to the ministry surveys, and an increase of +30.18% for higher median impact factors associated to those publications.

In the case of the second model, we found that both junior and assistant researchers present positive significant effects on the median impact factor of publications signed during the two years following the ministry surveys. With respect to senior researchers, juniors are more likely to present an increase of +4.42% in their median impact factors, while assistant are more likely to produce research of +28.5% higher quality. Moreover, regarding the institutional affiliation of researchers, being a CNRS researcher also presents positive effects on individual research quality, with an associated increase of +10.04% expected median impact factor with respect to university researchers, while on the contrary, being an INSERM researchers has negative effects on research quality (-6.9% expected median impact factor) with respect to being a university researcher.

In this case we also found positive and significant effects from the research produced in the past and its quality. In fact those researchers who produced an additional contribution to publication during the *t-5* to *t-2* years preceding the ministry surveys are more likely to perform better science with an increase of +0.12% expected median impact factor. Furthermore, the quality of the research produced during that period also presents positive effects with an associated increase of +25.9% median impact factor points for each higher point of past quality.

Finally, the results associated to the third model show very similar effects to those obtained from our second model. I found a positive and significant impact for having the status of a junior or an assistant researcher; junior researchers are thus more likely to produce research of +5.68% higher expected median impact factor with respect to senior researchers and of +16.12% higher quality in the case of assistants.

In addition I also found that being a CNRS researcher has positive significant effects on research quality with an associated increase of +9.66% expected median impact factor points with respect to researchers solely affiliated to the university, while on the other hand, those researchers affiliated to the INSERM present negative effects with an associated -4.9% lower expected research quality. It's important to point out that past scientific production and quality also present positive and significant effects on the current quality of individual research with a respective impact of +0.23% and +29.05% for each additional contribution made in the past and each additional median impact factor point reached for that contribution.

Let us highlight the fact that researchers carrying scientific research in any other field different than the life sciences are producing lower quality output with associated semi-elasticites of -29.53% (model 1), -36.05% (model 2) and -23.64% (model 3) in the case of social and human sciences, and -15.82% (model 1), -15.51% (model2) and -14.68% (model3) in the case of matter sciences.

3.4 Conclusion

Throughout the present section I have carried out studies on the characteristics of individual researchers to provide a larger picture of the determinants that play a role in the process of scientific production in research laboratories.

In a general overview, results from this section indicate, as expected, that the status of the researcher plays a role on his / her individual scientific output with the production of senior researcher being more important that of junior researchers. However, these findings are in contrast with the results regarding the influence of these characteristics on the quality of individual research output, which indicates that junior researchers are associated with higher quality research.

A possible explanation for these effects would be that young researchers build their career during their early years in the job by making the effort to produce scientific publications of high quality. Once there are set in a trajectory of increasing quality publications they establish a reputation that has eventually the effect of allowing them to produce higher quantities of scientific publications of relatively stable and even lower quality.

A more important result regards the institutional affiliation of individual researchers indicating that researchers with an institutional affiliation different than the university produce larger quantities of scientific output and are associated with high quality research.

4 Categories of coauthors and their influence on individual scientific research

In the present section I develop further my analysis on the scientific production by studying the effects of collaborations in scientific publications at the individual level. For this purpose, and since I deal with individual data, it becomes straightforward to study the effects of the categories of coauthors on the scientific output of individual researchers and decided to extend the original model of individual output on the category of researchers and include the categories of coauthors associated to a publication.

The explanatory variables in this new setting are not only defined by the type of individual researcher producing the contribution or the fractional publication, but also the total number of coauthors belonging to each one of the three categories (senior, junior and assistant coauthors), the individual age and career of the individual researcher and the average age and career of his/her coauthors, controlling for the total number of administrative personnel available in his unit, and as usual, the data collection method and the disciplinary field of the laboratory of affiliation. This model is detailed in the following expression.

$$\begin{aligned}
&\textit{Contributions and/or Fractional Publications}_{i,t} = \\
&\exp(\, \beta_1 + \beta_2\, \textit{Habilitation to direct research}_{i,t} + \beta_3\, \textit{Junior Researcher}_{i,t} \\
&+ \beta_4\, \textit{Assistant Researcher}_{i,t} + \beta_5\, \textit{Institutional CNRS}_{i,t} \\
&+ \beta_6\, \textit{Institutional INSERM}_{i,t} + \beta_7\, \textit{Other Institution}_{i,t} \\
&+ \beta_8\, \textit{Age}^2_{i,t} + \beta_9\, \textit{Career}^2_{i,t} + \beta_{10}\, \textit{Senior Coauthors}_{i,t} + \beta_{11}\, \textit{Junior Coauthors}_{i,t} \\
&+ \beta_{12}\, \textit{Assistant Coauthors}_{i,t} + \beta_{13}\, \textit{Coauthor Age}^2_{i,t} + \beta_{14}\, \textit{Coauthor Career}^2_{i,t} \\
&+ \beta_{15}\, \textit{Matter Science}_{i,t} + \beta_{16}\, \textit{Humanities}_{i,t} + \beta_{17}\, \textit{OST data collection method}_{i,t})\, \varepsilon_{i,t}
\end{aligned}$$

Table 11 Correlations among individual characteristics and types of coauthors

Explanatory Variable Correlations	Habilit.	Senior	Junior	Assist.	Univ.	CNRS	INSERM	Other Inst.	Age	Career
Habilitation	1									
Senior	0.5794	1								
Junior	-0.5659	-0.9915	1							
Assistant	-0.1019	-0.0621	-0.0683	1						
University	-0.0378	0.0366	-0.0314	-0.0399	1					
CNRS	0.0702	-0.0205	0.0283	-0.06	-0.7918	1				
INSERM	0.0092	-0.0145	0.017	-0.0187	-0.2474	-0.2628	1			
Other Institution	-0.0952	-0.0207	-0.0155	0.278	-0.1767	-0.1877	-0.0586	1		
Age	0.5933	0.5298	-0.5273	-0.0177	-0.0433	0.0635	-0.0031	-0.0477	1	
Career	0.29	0.2504	-0.246	-0.0334	-0.0785	0.1302	-0.049	-0.0665	0.4406	1
Senior Coauthors	0.1935	0.3273	-0.3267	-0.0037	-0.1187	0.0807	0.0642	0.0077	0.1651	0.1514
Junior Coauthors	-0.0319	-0.0788	0.0791	-0.0031	-0.0821	0.0837	0.0015	-0.0075	-0.0266	0.0822
Assistant Coauthors	-0.0007	0.0212	-0.0536	0.2482	0.0419	-0.0869	0.0072	0.1051	0.0364	0.0162
Avg. Age Coatuthors	0.442	0.3713	-0.3718	0.0054	-0.0613	0.0751	0.0034	-0.0404	0.8259	0.3809
Avg. Career Coauthors	0.1917	0.1417	-0.1397	-0.0148	-0.0849	0.1191	-0.0295	-0.0485	0.3163	0.8951

Explanatory Variable Correlations	Senior Coauth.	Junior Coauth.	Assistant Coauth.	Age Coauth.	Career Coauth.
Senior Coauthors	1				
Junior Coauthors	0.6328	1			
Assistant Coauthors	0.0993	0.0581	1		
Avg. Age Coatuthors	0.1274	-0.0523	0.0339	1	
Avg. Career Coauthors	0.1265	0.089	0.0151	0.3895	1

As one may observe from the correlation among individual researcher characteristics and their different types of coauthors reported in Table 11, collaboration through coauthorship is positively related in most cases with researchers affiliated to a research institution (CNRS and INSERM) suggesting that full time researchers collaborate more than part time researchers; in addition, one may also notice that senior researchers are positively correlated with senior and assistant coauthors, junior researchers are positively correlated with only with junior coauthors and assistant researchers with assistant coauthors.

This observation suggests that scientists tend to collaborate with researchers of their own kind, which is quite surprising given that since scientific production is a distributed process among different actors, we should observe interactions among different types of researchers. Finally, one may pleasantly observe that as individual researchers obtain experience in the laboratory they tend to increase their collaborations with all kinds of peers, although age on the other hand, only displays this effect on the collaboration with senior and assistant coauthors.

4.1 Categories of coauthors and their influence on Individual production measured by scientific contributions

The present subsection studies the individual contributions to scientific publications; it develops a regression analysis performed on the count variable through a negative binomial process selected after carrying the over-dispersion test. The results from this analysis reported in Table 12 show that in the case the first assumption there are significant and positive effects from the habilitation to direct research, which is associated with +26.7% higher contributions counts with respect to researchers who don’t have it. Junior researchers, as expected, present significant and negative effects on the expected outcome with -11.5% lower expected contributions with respect to senior researchers. Institutional affiliation does also have significant impacts, with +9.49% higher contributions for CNRS researchers, +15.8% higher contributions for INSERM and -10.18% lower contributions for researchers associated to other institutions with respect to those researchers who are solely associated to the university.

Table 12 Regression results of output explained by individual characteristics and types of coauthors

	Model 1 NB	Model 2 NB	Model 3 NB	Model 1 LogLinear	Model 2 LogLinear	Model 3 LogLinear
Explanatory Variables	Sum of contributions over 2 years preceding the survey	Sum of contributions over the 2 years following the survey	Sum of contributions from 2 years before to 2 years after survey	Sum of fractional publications over 2 years preceding the survey	Sum of fractional publications over the 2 years following the survey	Sum of fractional publications from 2 years before to 2 years after survey
Habilitation	0.2370***	0.1877***	0.2608***	0.3400***	0.2930***	0.4010***
	(0.0308)	(0.027)	(0.0271)	(0.0424)	(0.0449	(0.0406
Junior	-0.1226**	-0.2218***	-0.2546***	-0.1606***	-0.2064***	-0.2651***
	(0.05410	(0.0376)	(0.046)	(0.045)	(0.0505	(0.0434
Assistant	0.1221	0.3366*	-0.0107	0.2418	0.1082	-0.2382
	(0.1351)	(0.1764)	(0.1534)	(0.1968)	(0.2541)	(0.2085)
CNRS	0.0907***	0.1286***	0.1887***	0.0383	0.1038***	0.2068***
	(0.0204)	(0.0191)	(0.0191	(0.0304)	(0.0314)	(0.0296)
INSERM	0.1467***	0.1287***	0.2190***	0.1082**	0.0817	0.2481***
	(0.0324)	(0.0344)	(0.0305)	(0.0496)	(0.0572)	(0.049)
Other Institution	-0.1074**	-0.1764***	-0.1476***	-0.1761**	-0.2671***	-0.2429***
	(0.0544)	(0.0457)	(0.0453)	(0.0751)	(0.0819)	(0.0751)
Age	-0.0038	-0.0097***	-0.0125***	-0.0070**	-0.0114***	-0.0155***
	(0.0026)	(0.0023)	(0.0025)	(0.0035)	(0.0037)	(0.0039)
Career	0.0017	0.0004	0.003	0.0120**	0.0006	0.0098*
	(0.003)	(0.0028)	(0.003)	(0.0054)	(0.0054)	(0.0058)
Senior Coauthors	0.0445***	0.0400***	0.0188***	0.0452***	0.0426***	0.0195***
	(0.0021)	(0.0018)	(0.0008)	(0.002)	(0.0024)	(0.0009)
Junior Coauthors	0.0174**	0.0293***	0.0139***	-0.0212***	-0.0162***	-0.0077***
	(0.0076)	(0.0043)	(0.0028)	(0.0021)	(0.0037)	(0.0011)
Assistant Coauthors	0.0297**	0.0077	0.0171**	0.0147	-0.001	0.0193***
	(0.0138)	(0.0109)	(0.0068)	(0.0235)	(0.0165)	(0.0072)
Average Age Coauthors	-0.0048	-0.0036	0.0013	-0.0085*	-0.0142***	-0.0088*
	(0.0036)	(0.0032)	(0.0035)	(0.0045)	(0.0048)	(0.0051)
Avg. Career Coauthors	-0.0004	-0.0014	-0.0044	-0.0204***	-0.0132**	-0.0206***
	(0.0034)	(0.0033)	(0.0036)	(0.006)	(0.006)	(0.0065)
Data Method	0.0869***	0.0762***	0.1090***	-0.0292	-0.0016	0.0471
	(0.0234)	(0.0213)	(0.0214)	(0.0353	(0.0339)	(0.0328)
Matter Science	0.0433**	0.0444**	0.0078	0.1764***	0.2145***	0.1498***
	(0.0196)	(0.0183)	(0.018)	(0.0297)	(0.0307)	(0.029)
Humanities	-0.4410***	-0.4055***	-0.7048***	0.0466	-0.0014	-0.4042***
	(0.1075)	(0.0695)	(0.0774)	(0.0955)	(0.0838)	(0.0793)
Life Science/Senior (constant)	1.2720***	1.5691***	2.1508***	0.1497	0.6617***	1.2869***
	(0.0952)	(0.0932)	(0.0943)	(0.1222)	(0.1277)	(0.1251)
N	3756	3850	4385	3756	3850	4385
r2				0.2998	0.2779	0.3305
F				89.3634	65.8816	112.879
rmse				0.8563	0.9077	0.8976
chi2	1870.9717	1748.5983	2648.7117			
ll	-9988.1336	-10340	-14680	-4738.1533	-5081.5323	-5740.0497
aic	20010.2672	20721.0184	29395.9796	9510.3066	10197.0646	11514.0993
bic	20116.1961	20827.3675	29504.5407	9616.2355	10303.4137	11622.6604
Standard errors in parentheses		* p < 0.1	** p < 0.05	*** p < 0.001		

Furthermore, let us recall that the interest of the present analysis is to observe the effects of different types of coauthors and their average age and career on the sums of individual contributions, which for this first model present positive and significant effects with higher expected contribution counts of +4.55% for each additional senior coauthor a researcher collaborates with during the period, +1.75% higher expected contributions for each additional junior coauthor and +3% more for an additional assistant coauthor.

Regarding the second model, I found significant effects from junior and from assistant researchers with respect to seniors, with an associated decrease of -19.9% expected contributions in the case of juniors and an associated increase of +40% in the case of assistant researchers. The habilitation to direct research also presents in this case positive and significant effects with an associated increase of +20.6% in the expected outcome for those researchers who have it. The institutional affiliation also plays an important role, with associated effects of +13.72% expected outcome for those individuals affiliated to the CNRS with respect to those affiliated solely to the university, also +13.72% for those affiliated to the INSERM and -16.17% for those affiliated to other institutions.

In terms of the effects from different types of coauthors, I found that only senior and junior coauthors present significant impacts on the total contributions of individual researchers during the two years following the ministry surveys. In fact, there is an increase of +4.08% expected contributions for each additional senior coauthor, while any additional junior coauthor is associated with an increase of +2.9% expected contributions.

In the case of the third model I found that having the habilitation direct research presents significant and positive effects with an associated increase of +29.7% expected outcome with respect to researchers who do not have it. In this case, and as in the first model, we only found effects from junior researchers, with an associated decrease of -22.47% expected contributions with respect to senior researchers.

In terms of institutional affiliation, I found, as usual, significant and positive effects from an affiliation to the CNRS and the INSERM, with respectively +20.76% and +24.48% expected contributions, and negative effects from an affiliation to other institutions (-13.72% expected contributions) with respect to researchers with no institutional affiliation.

Regarding the influence of coauthor types on the individual scientific production for this model, we found positive and significant effects from all three types of coauthors, with an associated increase of +1.89% expected contributions for an additional senior coauthor, +1.24% for an additional junior coauthor and +1.72% for an additional assistant coauthor.

We may finally notice that for the present analysis no effects were found from the average age of coauthors, while negative individual age effects were found in models 2 and 3. We also found significant effects from the disciplinary field of work, with researchers in the human sciences being associated with much lower contributions than researchers in the life sciences and researchers in the matter sciences being associated with slightly higher contributions.

4.2 Categories of coauthors and their influence on Individual output measured by fractional publications

The analogue analysis on the influence of individual characteristics and types of coauthors on the actual sum of fractional publications also reported on Table 12 reveals that in the first model, the habilitation to direct research plays a significant and positive role with an associated increase of +40.5% expected shares for researchers who have passed it with respect to those who have not. As usual, junior researchers present lower expected publications shares with an associated decrease of -14.8% with respect to senior researchers. In the present model we only found significant effects for the affiliation to the INSERM and other research institutions with an associated increase of +11.4% and a decrease of -16.14% of the expected outcome with respect to university researchers. We also found individual age and career effects, with a -0.69% diminishing effect on the expected outcome as individuals grow older, counterbalanced by a +1.2% increasing expected outcome as they gain experience in the laboratory.

Regarding my interest in the influence of coauthors, I realize that senior and junior coauthors present significant effects, with an additional senior coauthor representing an increase of +4.6% in the expected publications shares, while an additional junior coauthor is associated with a decrease of -2.09%. In addition, we also found significant and negative coauthor age and career effects that imply a fall in the expected publications shares of researchers as their coauthors grow older.

Concerning the second model, I found there are significant effects from the habilitation to direct research (+34% expected outcome) and from the junior researcher status (-

18.64% expected outcome). In addition, institutional affiliation to the CNRS is associated with higher publications shares (+10.93%), while other institutions are associated with a fall of –23.44% expected shares. We may notice here that no significant effects were found for the institutional affiliation to the INSERM and only individual age effects were found with an associated fall of -1.13% expected shares as the researcher grows older.

In terms of the impacts of different coauthors, I found significant and positive effects from senior coauthors, with an additional collaboration of this type being associated with an increase of +4.35% expected publications shares, while significant negative effects were found from junior coauthors, with each additional of such collaborations representing a fall of -1.6% expected output. Moreover, significant and negative coauthor age and career effects were found implying that as their average age and experience grows, the expected actual production of a researcher falls with respect to his/her previous work.

Furthermore, the results from our third model show there are positive effects from the habilitation to direct research, with an associated increase of +49.33% expected shares with respect to researchers who don't have the habilitation. Junior researchers also present lower expected shares with an impact of -23.27% with respect to senior researchers, while the institutional affiliation presents positive effects for CNRS researchers (+22.9%) and INSERM researchers (+28.15%) and negative effects for researchers affiliated to other institutions (-21.56%) with respect to university researchers. In this model I also found individual age and career effects with decreasing expected publications shares as the researcher grows older and increasing ones as he gains experience.

Finally, regarding the influence of coauthors on the individual scientific production, I found that all three categories of coauthors (seniors, juniors and assistants) present significant effects on the sum of publications shares during the period of four year around the ministry surveys. In fact, senior and assistant coauthors present positive effects with an associated increase of +1.96% and +1.94% respectively in the expected outcome for each addition collaboration of such type, while junior coauthors present negative effects with an associated fall in expected shares of -0.76% for each additional junior collaboration.

In the case we also found significant and negative coauthor age and career effects (-0.87% and -2% respectively) implying that as coauthors grow older, there is a fall in the individual expected publications shares, or in other words, it's not a good idea for young researchers to collaborate with older coauthors id the objective is to produce more and build reputation.

It is interesting to highlight that the different analysis on the effects of individual researcher characteristics and the effects of different types of coauthors behaves similarly across our three models corresponding to three different propositions on the aggregation of scientific production around the ministry surveys. These results shed some light on how the status of a researcher, his institutional affiliation, and the types of coauthors he collaborates with may influence his own scientific output. Focusing my attention on the coauthoring influence I realize it is always better to collaborate with senior researchers and to some extent with assistant researchers, even if coauthor's age and career effects are negative, meaning that as a researcher collaborates with more experienced pairs, positive status effects are more important than negative age effects.

4.3 Categories of coauthors and their influence on Individual research quality

I arrive at the final set of analyses on the determinants of scientific research with an interest on the influence of different types of coauthors on the quality of individual scientific research; this analysis is similar to the analysis of the effects of coauthors on the research output. The effects of individual researcher characteristics, the effects of the coauthor categories and finally the effects of the lagged performance of researchers define the regression models as follows.

$$
\begin{aligned}
&\textit{Contributions and/or Fractional Publications}_{i,t} = \\
&\exp\,(\,ß_1 + ß_2\,\textit{Habilitation to direct research}_{i,t} + ß_3\,\textit{Junior Researcher}_{i,t} \\
&+ ß_4\,\textit{Assistant Researcher}_{i,t} + ß_5\,\textit{Institutional CNRS}_{i,t} \\
&+ ß_6\,\textit{Institutional INSERM}_{i,t} + ß_7\,\textit{Other Institution}_{i,t} \\
&+ ß_8\,\textit{Age}^2_{i,t} + ß_9\,\textit{Career}^2_{i,t} + ß_{10}\,\textit{Senior Coauthors}_{i,t} + ß_{11}\,\textit{Junioer Coauthors}_{i,t} \\
&+ ß_{12}\,\textit{Assistant Coauthors}_{i,t} + ß_{13}\,\textit{Coauthor Age}^2_{i,t} + ß_{14}\,\textit{Coauthor Career}^2_{i,t} \\
&+ ß_{15}\,\textit{Lagged Research Output}_{i,t} + ß_{16}\,\textit{Lagged Research Quality}_{i,t} \\
&+ ß_{15}\,\textit{Matter Science}_{i,t} + ß_{16}\,\textit{Humanities}_{i,t} + ß_{17}\,\textit{OST data collection method}_{i,t})\ \varepsilon_{i,t}
\end{aligned}
$$

In the case of the first model, the results displayed in Table 13 show that both junior and assistant researchers are associated with positive effects of +4.77% and +22.14%

expected median impact factor points with respect to the research quality senior researchers. Furthermore, researchers affiliated to the CNRS are more likely to perform higher quality research with an effect of +9.58% expected median impact factor with respect to a university researcher, while those affiliated to the INSERM are associated with lower quality (-4.59%).

Now, focusing on the main interest of the present analysis (the influence of coauthors on individual output quality) we learn that as a researcher collaborates with an additional senior coauthor, the quality of his scientific output seems to be more important with an increase of +0.28% expected median impact factor, while adding collaborations with junior researchers also presents positive significant effects on research quality with an associated increase of +0.1% expected median impact factor for each additional junior coauthor.

I also found positive coauthor average age effects implying that as a researcher collaborates with older coauthors his research is expected to be of better quality, however, coauthor average career effects in a given laboratory present negative significant effects.

In terms of past scientific production I only found that the quality of past research presents positive significant effects on current research quality with an associated impact of +29.04% expected median impact factor for each median point reached with the publications to which the researcher contributed to during the *t-5* to *t-2* years preceding the ministry surveys.

Regarding the second model, results obtained throughout the analysis also show a positive influence form the status of individual researchers on research quality, with +7.16% expected median impact factor for juniors researchers and +22.9% in the case of assistant researchers, both with respect to senior researchers. Concerning the institutional affiliation, we also found significant effects from being a CNRS or an INSERM researcher with an associated impact of +8.84% expected median impact factor in the case of a researcher affiliated to the CNRS and an impacts of -6.83% in the case of a researcher affiliated to the INSERM, both with respect to all other university researchers.

In this model I only found significant effects from the collaboration with senior coauthors; in fact, the expected median impact factor increases by +0.36% for each collaboration with a senior coauthor; which in contrast to the results in the previous subsection indicates that young researchers benefit from collaborating with senior coauthors if their objective is to publish their research in high quality journals, or in other words, benefit from the reputation capital of senior researchers.

In addition, we found positive significant coauthor average age effects indicating that collaborating with older coauthors is associated with an increase in the expected research quality. Finally, the quality of past research also presents positive significant effects on the current research quality with an associated increase of +25.14% expected median impact factor for each point obtained by the articles produced in the past.

Moreover, being a junior is associated with an increase of +6.76% and being an assistant with an increase of +13.84% expected median impact factor with respect to seniors. The institutional affiliation of researchers also plays a determinant role with an associated increase of +8.82% in the expected median impact factor for CNRS researchers and an associated decrease of -5.51% for INSERM researchers.

Concerning the influence of different types of coauthors, this third model revealed positive significant effects from collaborations with senior and junior researchers. In fact an additional senior coauthor is associated with higher research quality (+0.15% expected median impact factor), while an additional junior coauthor is associated with an increase of +0.05%.

Table 13 Regression results for research quality explained by individual characteristics and types of coauthors

Explanatory Variables	Median Impact Factor over 2 years preceding the survey	Median Impact Factor over the 2 years following the survey	Median Impact Factor from 2 years before to 2 years after survey	Average Impact Factor over 2 years preceding the survey	Average Impact Factor over the 2 years following the survey	Average Impact Factor from 2 years before to 2 years after survey
Habilitation	0.0063	0.0186	0.0099	0.003	0.0094	0.0021
	(0.0183)	(0.0182)	(0.0165)	(0.0169)	(0.0181)	(0.0161)
Junior	0.0466***	0.0692***	0.0655***	0.0494***	0.0532***	0.0551***
	(0.0178)	(0.0189)	(0.0168)	(0.0167)	(0.0183)	(0.0162)
Assistant	0.2000**	0.2069***	0.1297*	0.0897	0.1712***	0.083
	(0.0918)	(0.0763)	(0.0780)	(0.0927)	(0.0561)	(0.0695)
CNRS	0.0915***	0.0848***	0.0846***	0.0764***	0.0764***	0.0801***
	(0.0138)	(0.0142)	(0.0125)	(0.0129)	(0.0141)	(0.0123)
INSERM	-0.0470**	-0.0708***	-0.0567***	-0.0154	-0.0359*	-0.0392**
	(0.0193)	(0.0208)	(0.0168)	(0.0172)	(0.0185)	(0.0163)
Other Institution	-0.0154	0.0167	0.0248	0.018	0.0680**	0.0415
	(0.0437)	(0.0310)	(0.0313)	(0.0397)	(0.0302)	(0.0286)
Age	-0.0011	0.0005	-0.0016	-0.0013	0.0009	-0.001
	(0.0013)	(0.0015)	(0.0013)	(0.0012)	(0.0014)	(0.0013)
Career	0.0008	-0.002	0.0011	0.0016	-0.0007	0.0013
	(0.0018)	(0.0019)	(0.0018)	(0.0016)	(0.0019)	(0.0017)
Senior Coauthors	0.0028***	0.0036***	0.0015***	0.0030***	0.0030***	0.0014***
	(0.0007)	(0.0006)	(0.0003)	(0.0006)	(0.0006)	(0.0002)
Junior Coauthors	0.0010***	0.0001	0.0005***	0.0004	0.0012**	0.0003**
	(0.0003)	(0.0007)	(0.0002)	(0.0003)	(0.0005)	(0.0002)
Assistant Coauthors	-0.0052	0.0031	-0.0007	-0.0034	0.0024	0
	(0.0068)	(0.0046)	(0.0020)	(0.0059)	(0.0034)	(0.0020)
Average Age Coatuthors	0.0084***	0.0054***	0.0104***	0.0090***	0.0042**	0.0087***
	(0.0018)	(0.0020)	(0.0019)	(0.0017)	(0.0019)	(0.0018)
Average Career Coauthors	-0.0067***	-0.0031	-0.0071***	-0.0079***	-0.0038	-0.0069***
	(0.0021)	(0.0024)	(0.0023)	(0.0020)	(0.0024)	(0.0022)
Lagged Publications	-0.0008	-0.0011	-0.0013*	-0.0006	-0.0009	-0.0009
	(0.0007)	(0.0007)	(0.0007)	(0.0007)	(0.0006)	(0.0006)
Lagged Median Impact	0.2550***	0.2243***	0.2521***			
	(0.0052)	(0.0054)	(0.0049)			
Lagged Average Impact				0.2045***	0.1817***	0.1981***
				(0.0041)	(0.0043)	(0.0039)
Data Method	0.0722***	0.1278***	0.0958***	0.0991***	0.0899***	0.0861***
	(0.0129)	(0.0143)	(0.0128)	(0.0123)	(0.0134)	(0.0123)
Matter Science	-0.1582***	-0.1533***	-0.1521***	-0.1573***	-0.0938***	-0.1335***
	(0.0146)	(0.0149)	(0.0129)	(0.0135)	(0.0146)	(0.0124)
Humanities	-0.2732***	-0.3616***	-0.2820***	-0.1430*	-0.3780***	-0.2454***
	(0.0709)	(0.1133)	(0.0682)	(0.0845)	(0.1266)	(0.0676)
Life Science – Senior (constant)	-0.1036	0.1420**	-0.1189*	0.0804	0.3555***	0.1523**
	(0.0675)	(0.0684)	(0.0626)	(0.0634)	(0.0682)	(0.0624)
N	2395	2331	2607	2395	2331	2607
r2	0.7155	0.6664	0.7345	0.7467	0.6622	0.7342
F	266.5498	223.8732	303.6959	298.266	204.771	290.5473
rmse	0.3039	0.3069	0.2885	0.2837	0.301	0.2811
chi2						
ll	-536.3478	-544.7511	-448.8242	-371.6359	-499.5525	-380.8595
aic	1110.6956	1127.5022	935.6484	781.2718	1037.1049	799.719
bic	1220.5372	1236.8292	1047.1016	891.1134	1146.4319	911.1721
Standard errors in parentheses		* p < 0.1	** p < 0.05	*** p < 0.001		

We also found, as expected, positive significant effects from the quality of past research with an associated higher current quality (+28.68% median impact factor) for each median point reached with the publications performed during the *t-5* to *t-2* years preceding the ministry surveys.

5 Conclusions

Throughout the present chapter I have exposed the empirical analyses performed on the scientific performance of both, research laboratories and individual researchers, of the former University Louis Pasteur. These different analyses were performed with the objective of understanding what are the determinants of the output and its quality associated to the university.

The first set of analyses focused on the research laboratory as the unit of observation first and then broadened the picture by studying the individual researcher; both the aggregated output and research quality indicators of these structures were defined according to three different assumptions on the scientists' turnover across different periods of time and the influence this turnover may have on the unit's scientific output. This set of analyses models the effects of explanatory variables categorized in two different dimensions, human and financial resources, with a decomposition into researcher ranks within the French academic system, and into different types of funding, in addition I used a set of laboratory characteristics such as the institutional affiliation of its researchers and the disciplinary field of the unit.

Regarding collective scientific output evidence supports the main assumption of this thesis according to which the composition of research laboratories affects its scientific production. As a general overview, these results show that post doctors and PhD candidates stimulate by senior researchers and scientific production; it is also stimulated by the presence of institutional researchers, affiliated to CNRS or INSERM. In addition, results also provide evidence of the role the funding structure on the scientific output, which is stimulated by public funding.

Moreover, the general picture is completed by the analysis of determinants playing a role on the individual scientific output for which results provide evidence that the status of individual researchers within the higher education system play a major role on their production. As discussed earlier, the status of senior researcher is indeed associated with higher numbers of publications, as it is the institutional affiliation to CNRS or INSERM.

Regarding the collective scientific quality, there is evidence that quality is stimulated by the presence of institutional researchers affiliated to CNRS or INSERM and it is path dependent with the quality of lagged publications stimulating the quality of current publications. In addition, evidence on the role of the funding structure of research laboratories indicates that private funding stimulates collective scientific quality, which contrasts with the case of collective scientific production, which is instead stimulated by public funding.

At the individual level, it is interesting to learn from these results that it is junior researchers who are associated with higher scientific quality implying that junior researchers make important efforts to publish their research in quality peer-reviewed journals towards their early career and then through reputational effects they tend to shift towards an intensive scientific production as they grow older.

Chapter 4

Complementarities in Scientific Research

Chapter 4: Complementarities in Scientific Research

1 Introduction

The present chapter studies complementary relationships between determinants of scientific research. The previous chapter was closed with a with an analysis of the characteristics of coauthors and the role they play on individual scientific production and quality; this analysis offered insight on how coauthoring with senior researchers has the highest impact on the output of individual researchers as expected, and that coauthoring with assistant researchers has a higher impact on the individual production than coauthoring with young researchers.

As mentioned in the previous chapter, the analysis of the status of coauthors is a particular case of complementarities between determinants playing a role on the scientific production and hence it is of particular interest to present a broader picture of complementary relationships at the laboratory level.

For this matter, I propose an application of the theory of supermodularity which allows conclusion on pairwise complementary relationships between arguments of a function. In this case it will allow assessing whether the scientific production and its quality present complementary relationships between different categories of researchers and between different categories of funding.

2 The notion of complementary determinants

Overtime, the economic literature has focused part of its interest on the notion of complementarities between two different arguments or variables explaining economic phenomena. These complementarities are defined by an increasing marginal value of the real valued function with respect to a positive joint variation of both arguments. As a consequence of this definition, traditional analysis of complementarities relies on the assumptions of continuity and concavity of the real valued function, and verifies the property of increasing differences of a pair of arguments or variables on which the complementarities are tested.

In the case of a continuous function defined by a couple of variables x_i and x_j with $i \neq j$[17], it is straightforward to express these complementary relations which are reflected by an increasing differences condition. This condition is none other than a positive double differentiation of the real valued function, first with respect to the argument x_i, then with respect to argument x_j. In this case there are complementarities between variables x_i and x_j if the second derivative of the continuous function $f(x_i, x_j)$ with respect to the variables is higher or equal to zero.

However, not all economic phenomena are described by continuous functions and in the case of a discrete function, the property of increasing differences we are interested in indicates that for any real valued function defined by a couple of variables x_i and x_j, the difference between the observed values of the function evaluated at a different levels of one of the arguments, holding the other constant, is positive. As an example, if we evaluate the function at two different levels of the variable $x'_j >= x_j$, with x_i constant, the property of increasing differences implies that $f(x_i, x'_j) - f(x_i, x_j) \geqslant 0$, indicating that the function $f(x_i, x_j)$ increases in x_j.

Under the case of a joint variation of a pair of variables x_i and x_j, one may find that complementarities raise when the values of the function are higher than the its values observed under the variation of a single one of either argument; hence one may represent the property of increasing differences in the case of discrete functions as the positive difference between the sums of the real values of the function evaluated at joint levels and at an alternate levels of the arguments.

$$f(x'_i, x'_j) + f(x_i, x_j) >= f(x'_i, x_j) + f(x_i, x'_j) \text{ , for } x' >= x, \text{ and } i \neq j.$$

Since the analysis of the scientific production and quality of laboratories and individual researchers is based on observations representing real valued functions that depend on a decomposition of human and financial resources, I may apply the notion of complementarities described above. In addition, I believe this analysis is appropriate given that knowledge on the functional for of the scientific output and quality is not

[17] In our case, the set X represents the set of scientific researchers, while the indexes *i* and *j* refer to the different types of researchers found in the scientific laboratories: senior and junior researchers, post doctors and PhD's, and support staff or assistant personnel.

perfect, and that the assumption of continuity and concavity of these functions may or may not hold. In the impossibility to confirm these assumptions, I may avoid them using the alternative approach of supermodularity of the objective function, to test the condition of increasing differences and confirm the existence of complementarities between different explanatory variables classes.

2.1 The property of supermodularity: A tool to assess complementarities

Supermodularity, is a second order property of any function in the space R^n; it represents a condition of non-negativity of the cross differences in any pair of variables defining the objective function. This property is analogue to the property of concavity of any continuous twice-differentiable function in R^n and formalizes the existence of complementarities between a pair of arguments.

The property of supermodularity is conditioned to real valued functions *f(x)* that operate on partially ordered sets of arguments. Let the set *X* be a sub-lattice such that for any element *x'* and *x''* belonging to *X*, with a partial ordering *x'* < *x''*, the set 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), implying that in an *n*-dimensional space, the elements ($x' \vee x''$) and ($x' \wedge x''$) may be defined by the following expressions.

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

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

Given pair of arguments defining a real valued function, the property of supermodularity is verified if the condition of increasing differences in both arguments holds, that is, for any function *f(x)*, with its arguments belonging to a partially ordered set, the function is supermodular if the sum of its real values evaluated at the levels ($x' \vee x''$) and ($x' \wedge x''$) is higher than the sum of its real values evaluated at the levels (*x''*) and (*x'*).

This definition is generalized in an *n*-dimensional space by the following construction:

$$F(x') - F(x' \wedge x'') =$$

$$\sum_{i=1}^{n}[F\begin{pmatrix} x'_1 \\ \cdots \\ x'_i \\ x'_{i+1} \wedge x''_{i+1} \\ \cdots \\ x'_n \wedge x''_n \end{pmatrix} - F\begin{pmatrix} x'_1 \\ \cdots \\ x'_{i-1} \\ x'_i \wedge x''_i \\ \cdots \\ x'_n \wedge x''_n \end{pmatrix}]$$

$$\leqslant \sum_{i=1}^{n}[F\begin{pmatrix} x'_1 \vee x''_1 \\ \cdots \\ x'_{i-1} \vee x''_{i-1} \\ x'_i \\ x''_{i+1} \\ \cdots \\ x''_n \end{pmatrix} - F\begin{pmatrix} x'_1 \vee x''_1 \\ \cdots \\ x'_{i-1} \vee x''_{i-1} \\ x'_i \wedge x''_i \\ x''_{i+1} \\ \cdots \\ x''_n \end{pmatrix}]$$

$$\leqslant \sum_{i=1}^{n}[F\begin{pmatrix} x'_1 \vee x''_1 \\ \cdots \\ x'_{i-1} \vee x''_{i-1} \\ x'_i \vee x''_i \\ x''_{i+1} \\ \cdots \\ x''_n \end{pmatrix} - F\begin{pmatrix} x'_1 \vee x''_1 \\ \cdots \\ x'_{i-1} \vee x''_{i-1} \\ x''_i \\ x''_{i+1} \\ \cdots \\ x''_n \end{pmatrix}]$$

$$= F(x' \vee x'') - F(x'')$$

Source: Topkis, 1998. pg, 45.

Which is none other than:

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

In the case of a two dimensional function, this condition indicates that the sum of the complementary states, or real values evaluated at the levels *(x''_i, x''_j)* and *(x'_i, x'_j)*, is higher than the sum of intermediate states, or real values evaluated at the levels *(x''_i, x'_j)* and *(x'_i, x''_j)*, as one may see in the following inequalities.

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

In other words, given different elements of the set *X*, the inequality implies that the value of the sum of the objective function for any combination of the type (highest, highest) and (lowest, lowest) elements of each of each variable is more important than the sum of any other combination of intermediate elements. This inequality not only expresses the property of increasing differences of the function *f(.)* in any pair of arguments (x_i, x_j), but also indicates the existence of complementarities between these variables that define it.

In addition, one may interpret the supermodular property of a function as a larger increase of its marginal value when both of the arguments of a pair wise study vary jointly rather than when they vary one at a time, as a result, we obtain an equivalent definition of the notion of complementarities to that reflected by the non negative cross differentiation in the case of a continuous function.

3 Application to the economics of science:

3.1 A parametric approach on the supermodularity of collective and individual scientific output.

Assuming that scientific production of research laboratories depends on the effort performed by different categories of researchers and other laboratory characteristics; and assuming that the underlying scientific production function is not truly known given that the scientific production is only observed as a set of ordered real values, it is appropriate to assume that one may not directly infer the functional form of the scientific production of research laboratories or individual researchers; neither may one able to state whether the functions defining the scientific quality and production are endowed with continuity and differentiability properties.

Therefore, the originality of the present analyses relies on the methods rather than the intuition of complementarities itself, which has been present in the economic literature for some period of time in the field of the economics of science. By implementing a supermodular analysis, I study the observed real values of the function denoting the total scientific output of the research group over a certain period of time adapting a parametric approach in which I estimate the scientific production and quality under the states of complementarily and substitutability of a given couple of arguments (Athey and Stern, 1998) to test whether the scientific output and quality in question present complementarities among different arguments of the function under a supermodular approach.

In the case of research laboratories, the scientific quality and production I am interested in is based on a set of variables which contain ordered elements and are defined by the decomposition of scientific personnel into several researcher classes. Given this setting, I test whether joint variations of the number of certain types of researchers, or whether joint variations of the amount of certain types of funding in a scientific laboratory with respect to a fixed value have any influence on the scientific quality and production of research laboratories.

Testing the outcome following these joint variation allows me to assess whether there are complementarities between couples of variables, which in this particular case refers to the complementarities between types of scientific personnel and between different types of funding resources available.

I focus my interest on the notion that the scientific production of research laboratories is a collective process in which a group of researchers work together in order to synthesize, assimilate, create and produce new bodies of knowledge. My first analysis is therefore based on a decomposition of the scientific personnel of research laboratories around the following five different classes. For the purpose of this study, the decomposition is 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 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 defined not only by the notion of juridical status represented by the different diplomas a researcher may have validated, but also by the 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 study whether a situation where complementary types of researchers affiliated to a same research laboratory work collectively in order to produce a scientific output materialized in the form of scientific publications.

In addition to my interest in studying the complementary relations between types of researchers, I also study whether there are complementary relations between different sources of research funds. For this purpose I use the information on research funding available for the last two surveys of the analysis and perform a decomposition of the research funding available at the laboratory according to the following categories (as described in detail in chapter 2).

- Regional collectivities: Covering funding from regional organisms.
- Public recurrent funding: From public organisms associated to the labs such as CNRS and INSERM.
- European Union: Covering funding from European structural programs regarding scientific research.
- Other public funding: Covering funding from other public institutions such as French ministries, Public investment banks and other CNRS project-based funding.
- National and international private funding: Covering funding received from contracts with different private organisms, licensing, and patent exploitation.

3.1.1 Human resource complementarities in the scientific quality and production of research laboratories

The objective of the present analysis is to look for the existence of complementarities between specific couples of researchers with respect to three different fixed values: the average and median headcounts of the researcher status within a given disciplinary field, the arrival of new researchers of a given status at the laboratory in between two periods, and the headcount of researchers of a given status in a combination of laboratories.

I take into account the fact that research laboratories choose the amount of scientific human resources distributed across different ranks or categories based on the experience of researchers. I assume that the turnover of the number of positions filled within a laboratory from period to period helps it develop and/or improve its scientific production and the quality of its output.

Following this assumption, I used the framework of supermodularity to test whether there are increasing differences in the scientific output and quality with respect to a fixed value given the complementary or intermediate states of particular pairs of researcher categories. These states, defined by the combinations[18] of the observation of elements belonging to the variables above or below the fixed value, are obtained by comparing the headcount of a given researcher category with the fixed value in question.

I justify using a linear transformation to create the complementary and intermediary states by the assumption that laboratories within the same disciplinary field belong to the same communities of practice, usually publishing in the same set of scientific journals, presenting their work in the same conferences, filing patent applications that are very close coded, applying for the same type of public and private financial sources and raising opportunities of close industrial applications and forming researchers with similar interests. From this point of view, I proceed with the assumption that a parameter that captures a behavioural aspect of a given laboratory must not be different from the same

[18] State 1 denotes observations of variables 1 and 2 both above a fixed value.
State 2 denotes observations of variables 1 and 2 both below the fixed value.
State 3 denotes observations of variable 1 above and variable 2 below the fixed value.
State 4 denotes observations of variable 1 below and variable 2 above the fixed value.

parameter estimated from a random sample drawn from the population in a given subset of laboratories in a same discipline and field.

The complementary and intermediate states of a pair of researcher categories with respect to the median practices in the disciplinary field are therefore obtained by comparing the headcounts of researchers in a specific rank with the median count of researchers in the same rank within laboratories of the same disciplinary field. Taking the couple of senior and junior researchers as an example, I first calculated the median count of seniors and juniors by disciplinary field (life science, matter science and social and human science), then for each unit of analysis I compared whether the senior and junior counts are higher or lower than their median headcount in the disciplinary field.

Following this comparison, I generated a set of four dummy variables, two of which will indicate which are the laboratories presenting a complementary state between the two variables (seniors and juniors). The first complementary state (S1) takes a value of 1 when both counts of senior and junior researchers are higher than their median count within the discipline and 0 otherwise, while the second complementary state (S2) takes a value of 1 when both counts are lower than their median within the field and 0 otherwise. The other two dummy variables and will indicate which are the laboratories presenting an intermediate state between the two variables (seniors and juniors), with the first intermediate state (S3) taking a value of 1 if the senior count is higher than the median senior count within the disciplinary field and the junior count is lower than the median count within the discipline, while the second intermediates state (S4) takes a value of 1 if the inverse is verified.

Moreover, I also established the complementary and intermediate states of pair wise couples of researchers categories with respect to their turnover, which is mentioned in previous chapters, detailing the evolution of seniors, juniors, PhD’s and post doctors and assistants. For this purpose, I also generated a set of dummy variables that take into account the joint turnover of a pair of categories. Taking once again the couple of senior and junior researchers as an example, this set of dummy variables indicate the following four cases: 1) both seniors and juniors present a negative turnover meaning that in between two given periods (ministry surveys) their headcount actually shrank, 2) while the seniors researchers present a positive turnover given a change of period, the junior researchers present a negative turnover, 3) while the senior researchers present a negative turnover, the junior researchers present a positive one and finally 4) both seniors and juniors present a positive turnover given a change of period.

It is important to notice that when using the turnover in time of researcher headcounts in a given rank or status, the analysis only takes into account the couples: *senior – junior, senior – assistant, and junior – assistant* researchers. The reason of this limitation is that we cannot observe the actual evolution of researchers straight from the laboratory dataset but rather from the personnel dataset, which does not provide any information on PhD's or post doctors since they are considered non-permanent personnel in the original data from the ministry and therefore lack nominative information.

Once the set of dummy variables in each of these two cases is defined I proceeded to test whether the scientific output and quality present increasing differences in the different pairs of researcher ranks or status. Following Athey and Stern, 1998, I performed the estimation of the scientific output (laboratory publications according to three different aggregation hypotheses) and its scientific quality (median and average impact factor of laboratory publications according to three different aggregation hypotheses) using the parametric regressions performed in chapter 3 on the determinants of research output and quality. In the present case, these estimations were augmented by the set of dummy variables taking into account the complementary and intermediate states given pair wise couples of explanatory variables.

As a result, in the case of the output estimation, the following model was used:

$$\begin{aligned}
&\textit{Contributions and/or Fractional publications}_{i,t} = \\
&\exp(ß_1\, Size_{i,t} + ß_2\, \textit{Institutional Researchers}_{i,t} + ß_3\, \textit{Institutional Personnel}_{i,t} \\
&+ ß_4\, \textit{Defended Theses}_{i,t} + ß_5\, Age^2_{i,t} + ß_6\, Career^2_{i,t} + ß_7\, \textit{OST data collection method}_{i,t} \\
&+ ß_8\, \textit{Matter Science}_{i,t} + ß_9\, Humanities_{i,t} + ß_{10}\, \textit{Life Science}_{i,t} \\
&+ ß_{11}\, Couple00_{i,t} + ß_{12}\, Couple01_{i,t} + ß_{13}\, Couple10_{i,t} + ß_{14}\, Couple11_{i,t})\, \varepsilon_{i,t}
\end{aligned}$$

While in the case of the quality estimation we used the model:

$$\begin{aligned}
&\textit{Average and/or Median Impact Factor}_{i,t} = \\
&\exp(ß_1\, Size_{i,t} + ß_2\, Institutional\ Researchers_{i,t} + ß_3\, Institutional\ Personnel_{i,t} \\
&+ ß_4\, Defended\ Theses_{i,t} + ß_5\, Age^2_{i,t} + ß_6\, Career^2_{i,t} \\
&+ ß_7\, Lagged\ Contributions_{i,t} + ß_8\, Lagged\ Quality_{i,t} + ß_9\, OST\ data\ collection\ method_{i,t} \\
&+ ß_{10}\, Matter\ Science_{i,t} + ß_{11}\, Humanities_{i,t} + ß_{12}\, Life\ Science_{i,t} \\
&+ ß_{13}\, Couple00_{i,t} + ß_{14}\, Couple01_{i,t} + ß_{15}\, Couple10_{i,t} + ß_{16}\, Couple11_{i,t})\, \varepsilon_{i,t}
\end{aligned}$$

In each of these models the variable Couple00 reflects the complementary state 1 where both arguments in the pair of interest are in their lower state, Couple01 reflects the intermediate state where the first argument in a pair is in the upper state and the second argument in the lower state, Couple10 represents the complementary state where the first argument is in the lower state and the second in the upper state and finally, Couple11 represents the complementary state where both arguments are in their upper state; these variables take the value 1 if the laboratory verifies that condition.

Once the models where estimated, I isolated the regression coefficients of the four dummy variables and performed a one sided Student test in which the null hypothesis indicates the absence of differences between the sum of complementary and the sum of intermediate states, against the alternative hypothesis, which indicates that the difference is positive. This rejection of the null hypothesis implies that there are increasing differences of the objective function in the couple of arguments under study and therefore that the property of supermodularity is verified. As a consequence, the dependent variable has complementarities in the pair of arguments in question.

$$H_0: ß[Couple00] + ß[Couple11] - ß[Couple01] - ß[Couple10] = 0$$

$$H_1: ß[Couple00] + ß[Couple11] - ß[Couple01] - ß[Couple10] > 0$$

Table 14 Human resource complementarities in output and quality with respect to median headcounts in the disciplinary field

Couple	Dependent Variable	Prob. Dependent 2 years before survey	Prob. Dependent 2 years after survey	Prob. Dependent 2 years before and after survey
Senior-Junior	Fractional Publications	0.9824095	0.9588623	0.9786398
Senior-PhD Postdoctor	Fractional Publications	0.6996165	0.606921	0.6690441
Senior-Assistant	Fractional Publications	0.1136377	0.0165555**	0.0235332**
Junior-PhD Postdoctor	Fractional Publications	0.5587825	0.5252246	0.5769975
Junior-Assistant	Fractional Publications	0.9549829	0.7145278	0.8602837
Phd Postdoctor-Assistant	Fractional Publications	0.997601	0.9477361	0.9895121
Senior-Junior	Contributions	0.9920784	0.9631689	0.9866799
Senior-PhD Postdoctor	Contributions	0.7762324	0.851388	0.8277264
Senior-Assistant	Contributions	0.1324674	0.0190826**	0.03342*
Junior-PhD Postdoctor	Contributions	0.4110067	0.5426458	0.473024
Junior-Assistant	Contributions	0.4546687	0.2859102	0.2949917
Phd Postdoctor-Assistant	Contributions	0.6283655	0.6347317	0.5836453
Couple	**Dependent Variable**	**Prob. Dependent 2 years before survey**	**Prob. Dependent 2 years after survey**	**Prob. Dependent 2 years before and after survey**
Senior-Junior	Average Impact Factor	0.596337	0.2954883	0.39302
Senior-PhD Postdoctor	Average Impact Factor	0.8645139	0.6485621	0.8257767
Senior-Assistant	Average Impact Factor	0.9010782	0.6753839	0.7501166
Junior-PhD Postdoctor	Average Impact Factor	0.2340702	0.0741724	0.0925124
Junior-Assistant	Average Impact Factor	0.9757115	0.8676404	0.9322123
Phd Postdoctor-Assistant	Average Impact Factor	0.944791	0.9115582	0.8875781
Senior-Junior	Median Impact Factor	0.7545736	0.8016902	0.5168068
Senior-PhD Postdoctor	Median Impact Factor	0.8409694	0.8414475	0.7192562
Senior-Assistant	Median Impact Factor	0.4645946	0.4208481	0.552868
Junior-PhD Postdoctor	Median Impact Factor	0.6086327	0.1422869	0.3832792
Junior-Assistant	Median Impact Factor	0.4185542	0.9098119	0.8543112
Phd Postdoctor-Assistant	Median Impact Factor	0.0588375	0.7361779	0.384979

When I performed this analysis in the case where the complementary and intermediates states are defined with respect to the median counts within the disciplinary field, I found that both, total contributions to publications and total fractional publications, display increasing differences in the couple *senior-assistant* for two aggregation cases[19] (models 2 and 3) with the null hypothesis being rejected at the 5% and 10% threshold.

These results, reported on Table 14 show that the supermodular condition of increasing differences is verified by the data in the couple of senior and assistant researcher categories. These findings imply that the scientific output of research laboratories is more

[19] The aggregation cases are the models 1, 2 and 3 of output aggregation defined in chapter 3:
Model 1 corresponds to the sum of contributions over 2 years before the ministry surveys.
Model 2 corresponds to the sum of contributions over 2 years following the ministry surveys.
Model 3 corresponds to the sum of contributions 4 years around the ministry surveys.

important in laboratories that choose to increase the headcounts of both senior and assistants researchers beyond their median levels within their disciplinary field rather than just one category at a time. On the other hand, these results from these analyses also show the absence of any increasing differences in any researcher couple when it comes to the study of the laboratory scientific quality measured either by median average publication impact factors.

When performed in the case where the complementary and intermediate states are defined with respect to the turnover in time of researcher categories, the analyses of supermodularity on the scientific output (total contributions to publications and total fractional publications) show an absence of increasing differences any couple of arguments, while the analysis on the scientific quality does reflect the existence of increasing differences in the couple *senior-assistant*.

In fact, the null hypothesis is rejected at the 5% threshold for the average impact factor of laboratory publications for all three aggregation models (2 years before, 2 years after and 4 years around the ministry surveys). These results, reported on Table 15, tell us that the average impact factor of publications is more important in laboratories where the headcount of senior researchers evolves in the same direction as the assistant researchers with respect to time.

Table 15 Human resource complementarities in output and quality with respect to researcher turnover

Couple	Dependent Variable	Prob. Dependent 2 years before survey	Prob. Dependent 2 years after survey	Prob. Dependent 2 years before and after survey
Senior-Junior	Fractional Publications	0.9768803	0.9185567	0.9667856
Senior-PhD Postdoctor	Fractional Publications	0.9228107	0.9296246	0.9557133
Senior-Assistant	Fractional Publications	0.9494203	0.9725817	0.972655
Senior-Junior	Contributions	0.9262099	0.9333935	0.923689
Senior-PhD Postdoctor	Contributions	0.7643848	0.7831147	0.8042425
Senior-Assistant	Contributions	0.9106068	0.9863397	0.9645615
Couple	**Dependent Variable**	**Prob. Dependent 2 years before survey**	**Prob. Dependent 2 years after survey**	**Prob. Dependent 2 years before and after survey**
Senior-Junior	Average Impact Factor	0.8666762	0.434693	0.5395915
Senior-Assistant	Average Impact Factor	0.0098488**	0.0157561**	0.0071473**
Junior-Assistant	Average Impact Factor	0.078205	0.6263347	0.2851948
Senior-Junior	Median Impact Factor	0.9912199	0.7446771	0.3998561
Senior-Assistant	Median Impact Factor	0.845216	0.2501828	0.2254952
Junior-Assistant	Median Impact Factor	0.8671675	0.2743752	0.6045821

3.1.2 Complementarities of funding sources in the scientific output and quality of research laboratories

Taking the analysis of complementarities in the scientific output and quality beyond the scope of human resources and focusing on the possible complementarities between different types of funding sources available at the laboratory. For this purpose I selected six different couples of funding types out of the 21 pair wise possibilities of combination; the first couple I defined is *{total private funding; total public funding}* which takes into account the aggregates of all sorts of private funding and the aggregates of all sorts of public funding. In order to define the other couples of the study, I used the decomposition of the total private and total public funding and sorted out the following couples *{regional funding; European funding}*, *{private funding; European funding}*, *{private funding; regional funding}*, *{private funding; recurrent public funding}* and finally *{private funding; other public funding}*.

Once the couples of funding sources were defined and following an analogous procedure as used in the previous section, I established the complementary and intermediate states of which each laboratory or research unit, first with respect to the median amount of funding available within the disciplinary field and then with respect to the evolution of these figures across time from the 2004 survey to the 2008 survey[20].

To proceed with the analysis I performed a standard estimation of the contributions to publications, the fractional publications and the average and the median impact factor of laboratory publications on the explanatory variables defined in the previous chapter, augmented by our recently defined complementary and intermediate states for each couple of funding source type; in the case of scientific output and quality, the models in question may be expressed as follows.

$$\textit{Contributions and/or Fractional publications}_{i,t} =$$
$$\exp(\beta_1\, Size_{i,t} + \beta_2\, Total\ Funding_{i,t} + \beta_3\, Institutional\ Researchers_{i,t}$$
$$+ \beta_4\, Institutional\ Personnel_{i,t} + \beta_5\, Defended\ Theses_{i,t} + \beta_6\, Age^2_{i,t} + \beta_7\, Career^2_{i,t}$$
$$+ \beta_8\, OST\ data\ collection\ method_{i,t} + \beta_9\, Matter\ Science_{i,t}$$
$$+ \beta_{10}\, Humanities_{i,t} + \beta_{11}\, Life\ Science_{i,t}$$
$$+ \beta_{12}\, Couple00_{i,t} + \beta_{13}\, Couple01_{i,t} + \beta_{14}\, Couple10_{i,t} + \beta_{15}\, Couple11_{i,t})\, \varepsilon_{i,t}$$

$$\textit{Average and/or Median Impact Factor}_{i,t} =$$
$$\exp(\beta_1\, Size_{i,t} + \beta_2\, Total\ Funding_{i,t} + \beta_3\, Institutional\ Researchers_{i,t}$$
$$+ \beta_4\, Institutional\ Personnel_{i,t} + \beta_5\, Defended\ Theses_{i,t} + \beta_6\, Age^2_{i,t} + \beta_7\, Career^2_{i,t}$$
$$+ \beta_8\, Lagged\ Contributions_{i,t} + \beta_9\, Lagged\ Quality_{i,t} + \beta_{10}\, OST\ data\ collection\ method_{i,t}$$
$$+ \beta_{11}\, Matter\ Science_{i,t} + \beta_{12}\, Humanities_{i,t} + \beta_{13}\, Life\ Science_{i,t}$$
$$+ \beta_{14}\, Couple00_{i,t} + \beta_{15}\, Couple01_{i,t} + \beta_{16}\, Couple10_{i,t} + \beta_{17}\, Couple11_{i,t})\, \varepsilon_{i,t}$$

Where the variables Couple00 and Couple11 denote the complementary state among the couple of funding types under analysis and the variables Couple01 and Couple10 denote their intermediate state.

Following the estimation of these models, I isolated the coefficients related to the complementary and intermediate dummy variables to perform a one sided test that would assess whether the effects of the complementary states are greater than the effects of the intermediate states. The null hypothesis of the test is defined by the equality between the effects of complementary and intermediate states, while the alternative hypothesis states that the effects of the complementary states are more important that the intermediate

[20] Let us recall that the data on funding is available for the period 2001 – 2008, which after aggregation corresponds to the funding related to the declarations in surveys 2004 and 2008.

states, which implies the existence of increasing differences and verifies the property of supermodularity of the objective function.

$$H_0: ß[Couple00] + ß[Couple11] - ß[Couple01] - ß[Couple10] = 0$$

$$H_1: ß[Couple00] + ß[Couple11] - ß[Couple01] - ß[Couple10] > 0$$

Table 16 Research funds complementarities in output and quality with respect to median headcounts in the disciplinary field

Couple	Dependent Variable	Prob. Dependent 2 years before survey	Prob. Dependent 2 years after survey	Prob. Dependent 2 years before and after survey
Total Private-Total Public	Fractional Publications	0.4849931	0.8193481	0.5969303
Private-Other Public	Fractional Publications	0.0207396**	0.0601348	0.0406906*
Regional-European	Fractional Publications	0.2073746	0.3330602	0.2136264
Private-Recurrent Public	Fractional Publications	0.8929004	0.9764475	0.9388305
Private-European	Fractional Publications	0.8821127	0.9757158	0.9330105
Private-Regional	Fractional Publications	0.0020415***	0.0360241*	0.0093504***
Total Private-Total Public	Contributions	0.123179	0.3351073	0.164146
Private-Other Public	Contributions	0.0059462**	0.0282073*	0.0128295**
Regional-European	Contributions	0.6178447	0.857596	0.6942675
Private-Recurrent Public	Contributions	0.2815856	0.6114983	0.3680109
Private-European	Contributions	0.9921809	0.9953012	0.9947348
Private-Regional	Contributions	0.010541**	0.0619932	0.0244895**
Couple	Dependent Variable	**Prob. Dependent 2 years before survey**	**Prob. Dependent 2 years after survey**	**Prob. Dependent 2 years before and after survey**
Total Private-Total Public	Average Impact Factor	0.0543508	0.8150028	0.2559608
Private-Other Public	Average Impact Factor	0.0705474	0.3456398	0.1237085
Regional-European	Average Impact Factor	0.8970969	0.6728547	0.7378079
Private-Recurrent Public	Average Impact Factor	0.4146907	0.8966217	0.61789
Private-European	Average Impact Factor	0.4015139	0.3982912	0.4359732
Private-Regional	Average Impact Factor	0.4984504	0.7034366	0.5146751
Total Private-Total Public	Median Impact Factor	0.0522981	0.3009634	0.3907497
Private-Other Public	Median Impact Factor	0.3751381	0.6742848	0.2736721
Regional-European	Median Impact Factor	0.9292322	0.8673056	0.8526491
Private-Recurrent Public	Median Impact Factor	0.8002831	0.8676892	0.8468049
Private-European	Median Impact Factor	0.4829669	0.8682185	0.873207
Private-Regional	Median Impact Factor	0.9935992	0.9826077	0.9726111

The results from the analysis, reported in Table 16, deal with the case where I tested the existence complementarities with respect to the median amount of funding available within the disciplinary field. These results show that the total fractional publications aggregated over 2 years preceding the ministry surveys and over 4 years around the surveys (models 2 and 3) present increasing differences in two different couples, *private-*

regional funding and *private-other public* funding rejecting the null hypothesis at the 5% threshold, while the total fractional publications aggregated over two years after the ministry surveys also presents increasing differences in the couple *private-regional* funding. In addition, a similar result is found in the case of the analysis of total contributions to publications, with increasing differences in the same couples *private-regional* funding and *private-other public* funding.

Table 17 Research funds complementarities in output and quality with respect the evolution of funding

Couple	Dependent Variable	Prob. Dependent 2 years before survey	Prob. Dependent 2 years after survey	Prob. Dependent 2 years before and after survey
Total Private-Total Public	Fractional Publications	0.7924325	0.7717285	0.8157827
Private-Other Public	Fractional Publications	0.4425136	0.6877862	0.6138496
Regional-European	Fractional Publications	0.975404	0.8295827	0.9513118
Private-Recurrent Public	Fractional Publications	0.698488	0.7664748	0.7546098
Private-European	Fractional Publications	0.8745474	0.5504888	0.7552507
Private-Regional	Fractional Publications	0.1256217	0.3902664	0.2126696
Total Private-Total Public	Contributions	0.5284546	0.6075554	0.5904669
Private-Other Public	Contributions	0.5859264	0.8962966	0.7476608
Regional-European	Contributions	0.8919354	0.6233038	0.8546662
Private-Recurrent Public	Contributions	0.3244113	0.4090492	0.4691262
Private-European	Contributions	0.6619228	0.4831989	0.5607742
Private-Regional	Contributions	0.3213538	0.5673594	0.420301
Couple	**Dependent Variable**	**Prob. Dependent 2 years before survey**	**Prob. Dependent 2 years after survey**	**Prob. Dependent 2 years before and after survey**
Total Private-Total Public	Average Impact Factor	0.4395955	0.3588798	0.4193687
Private-Other Public	Average Impact Factor	0.2751356	0.7044091	0.4803203
Regional-European	Average Impact Factor	0.362156	0.5106893	0.5183387
Private-Recurrent Public	Average Impact Factor	0.2258822	0.6159008	0.5363073
Private-European	Average Impact Factor	0.6361058	0.4252017	0.5247564
Private-Regional	Average Impact Factor	0.7536202	0.814167	0.8532561
Total Private-Total Public	Median Impact Factor	0.5327173	0.6800896	0.4529855
Private-Other Public	Median Impact Factor	0.27567	0.3302296	0.1523155
Regional-European	Median Impact Factor	0.3121834	0.5464404	0.3988759
Private-Recurrent Public	Median Impact Factor	0.1091365	0.402311	0.2511885
Private-European	Median Impact Factor	0.2819652	0.4524027	0.1924215
Private-Regional	Median Impact Factor	0.4853843	0.4810208	0.4823908

These results teach us that as laboratories receive higher amounts of private and regional funding, both over their median within the disciplinary field, their sum of fractional publications will by more important than if only one of these financial sources were to be higher than its median within the disciplinary field.

Unfortunately, the results from the analysis of complementarities of the average and median impact factors, with respect to the median amounts available within the discipline, reported on Table 16, show there is an absence of increasing differences (no rejection of H_0) in any couples of funding; while in the case of the analysis with respect to the evolution of funding in time, for which the results are reported on Table 17, none of the dependent variables indicating either output or quality present any increasing differences in any couple of financial sources.

3.1.3 Complementarities between coauthors in the individual scientific production

In addition to the analysis of complementarities in the scientific output and quality of research laboratories, I also performed an analogue analysis on the actual individual contributions to scientific publications. Given the limitations of the personnel dataset, I performed this analysis taking into account only the couples of coauthors: *senior – junior, senior – assistant, and junior – assistant* researchers.

Similarly to the analysis in the previous subsection where I studied the output and quality of laboratories, I established the complementary and intermediate states of coauthoring with respect to the median headcount of coauthors in a given rank within the discipline and with respect to the evolution in time of these figures. Once the set of dummy variables describing each of the states was defined, I performed a regression analysis of the dependent variables (individual output and quality) on the set of explanatory variables describing the environment in which researchers work; these models may be detailed in the following expressions.

$$\begin{aligned}&\textit{Contributions and/or Fractional Publications}_{i,t} = \\ &\exp(\beta_1\, \textit{Habilitation to direct research}_{i,t} + \beta_2\, \textit{Senior Researcher}_{i,t} \\ &+ \beta_3\, \textit{Junior Researcher}_{i,t} + \beta_4\, \textit{Assistant Researcher}_{i,t} + \beta_5\, \textit{Institutional CNRS}_{i,t} \\ &+ \beta_6\, \textit{Institutional INSERM}_{i,t} + \beta_7\, \textit{Other Institution}_{i,t} + \beta_8\, \textit{Age}^2_{i,t} \\ &+ \beta_9\, \textit{Career}^2_{i,t} + \beta_{10}\, \textit{OST data collection method}_{i,t} + \beta_{11}\, \textit{Matter Science}_{i,t} \\ &+ \beta_{12}\, \textit{Humanities}_{i,t} + \beta_{13}\, \textit{Life Science}_{i,t} + \beta_{14}\, \textit{Couple00}_{i,t} + \beta_{15}\, \textit{Couple01}_{i,t} \\ &+ \beta_{16}\, \textit{Couple10}_{i,t} + \beta_{17}\, \textit{Couple11}_{i,t})\, \varepsilon_{i,t}\end{aligned}$$

$$\begin{aligned}&\textit{Average and/or Median Impact Factor}_{i,t} = \\ &\exp(\beta_1\, \textit{Habilitation to direct research}_{i,t} + \beta_2\, \textit{Senior Researcher}_{i,t} \\ &+ \beta_3\, \textit{Junior Researcher}_{i,t} + \beta_4\, \textit{Assistant Researcher}_{i,t} + \beta_5\, \textit{Institutional CNRS}_{i,t} \\ &+ \beta_6\, \textit{Institutional INSERM}_{i,t} + \beta_7\, \textit{Other Institution}_{i,t} + \beta_8\, \textit{Age}^2_{i,t} \\ &+ \beta_9\, \textit{Career}^2_{i,t} + \beta_{10}\, \textit{Lagged Contributions}_{i,t} + \beta_{11}\, \textit{Lagged Quality}_{i,t} \\ &+ \beta_{12}\, \textit{OST data collection method}_{i,t} + \beta_{13}\, \textit{Matter Science}_{i,t} + \beta_{14}\, \textit{Humanities}_{i,t} \\ &+ \beta_{15}\, \textit{Life Science}_{i,t} + \beta_{16}\, \textit{Couple00}_{i,t} + \beta_{17}\, \textit{Couple01}_{i,t} + \beta_{18}\, \textit{Couple10}_{i,t} \\ &+ \beta_{19}\, \textit{Couple11}_{i,t})\, \varepsilon_{i,t}\end{aligned}$$

Once again, after the estimations were run, I isolated the coefficients related to the dummy variables and performed a one sided test on the difference between the sum of the effects of the complementary states and the sum of the effects of the intermediate states. The null hypothesis expresses the absence of differences between these sums, while the alternative hypothesis expresses the existence of increasing differences and therefore the property of supermodularity of the objective fonction.

When looking at the results from the analysis of complementarities in a given pair of coauthors with respect to the median counts of coauthors in the disciplinary field, reported on Table 18, one may learn that all dependent variables reflecting the scientific output display increasing differences in the couples *senior-assistant* and *junior-assistant* under all three aggregation models with the null hypothesis being rejected at the 5% threshold.

Moreover, when looking at the results of the same analysis on our quality indicators (average and median impact factors), we learn that only the average impact factor presents increasing differences in the couple *junior-assistant* for all three aggregation models around a survey and in the couple senior-assistant when the publications are added from 2 years before the survey (model 1) and when added to 2 years after the survey (model

2). What is important to notice here is that both output and quality of individual researchers present complementarities in the couples of senior-assistant and junior-assistant coauthors, meaning that the output and quality of an individual will be more important when both his/her number of collaborations with senior and/or junior coauthors and his/her number of collaborations with assistant coauthors are above the median practice within the discipline rather than one above and the other under this threshold.

Table 18 Coauthor complementarities in output and quality with respect to median headcounts in the disciplinary field

Couple	Dependent Variable	Prob. Dependent 2 years before survey	Prob. Dependent 2 years after survey	Prob. Dependent 2 years before and after survey
Senior-Junior	Fractional Publications	0.8341061	0.9982534	0.8372663
Senior-Assistant	Fractional Publications	0***	0***	0***
Junior-Assistant	Fractional Publications	0***	0***	0***
Senior-Junior	Contributions	0.2928228	0.9885841	0.5229291
Senior-Assistant	Contributions	0***	0***	0***
Junior-Assistant	Contributions	0***	0***	0***
Couple	**Dependent Variable**	**Prob. Dependent 2 years before survey**	**Prob. Dependent 2 years after survey**	**Prob. Dependent 2 years before and after survey**
Senior-Junior	Average Impact Factor	0.8401735	0.9966403	0.9403767
Senior-Assistant	Average Impact Factor	0.0001046***	0.0000394***	0.197248
Junior-Assistant	Average Impact Factor	0.00000137***	0.0002834***	0.000000119***
Senior-Junior	Median Impact Factor	0.9038072	0.7948627	0.8592337
Senior-Assistant	Median Impact Factor	0.0517593	0.1218691	0.3537748
Junior-Assistant	Median Impact Factor	0.9101458	0.1882279	0.3704177

Further into the analyses, the results concerning complementarities in a couple of variables with respect to their evolution in time, reported on Table 19, show that both output variables, contributions and fractional counts, present increasing differences in the couple *senior-assistant* when aggregated over 2 years following the ministry surveys (model 2) and over 4 years around the these surveys (model 3); the couple *junior-assistant* shows the same effects only when aggregated over 2 years following the surveys (model 2).

Regarding the quality of the individual research, the results from the analysis of complementarities with respect to the evolution of individual coauthor counts in time, also reported on Table 19, reveal that the average impact factor of individual publications presents increasing differences in the couple *senior-assistant* when aggregated over 2 years preceding the survey (model 1) and in the couple *junior-assistant* when aggregated over 2 years following the survey (model 2) and over 4 years around the survey (model

3); while the median impact factor only shows increasing differences in the couple *junior-assistant* when aggregated over 2 years following (model 2) and over 4 years around (model 3) the surveys.

What we must notice from these results is that both individual research output and quality present complementarities in the couples of senior-assistant and junior-assistant coauthors from a supermodular point of view, which means that their real values will be more important when the number of collaboration an individual has with senior and/or junior coauthors evolves in the same direction as his/her number of collaborations with assistant coauthors.

Table 19 Coauthor complementarities in output and quality with respect to their evolution in time

Couple	Dependent Variable	Prob. Dependent 2 years before survey	Prob. Dependent 2 years after survey	Prob. Dependent 2 years before and after survey
Senior-Junior	Fractional Publications	0.4790193	0.8687427	0.7228822
Senior-Assistant	Fractional Publications	0.0339043*	0.0033512*	0.0008579***
Junior-Assistant	Fractional Publications	0.9999506	0.000000298***	0.3897781
Senior-Junior	Contributions	0.8267853	0.5476565	0.0145394**
Senior-Assistant	Contributions	0.1084794	0.000496***	0.0019827***
Junior-Assistant	Contributions	0.9999149	0.0004228***	0.6730325
Couple	**Dependent Variable**	**Prob. Dependent 2 years before survey**	**Prob. Dependent 2 years after survey**	**Prob. Dependent 2 years before and after survey**
Senior-Junior	Average Impact Factor	1	1	0.9921218
Senior-Assistant	Average Impact Factor	0***	0.0256637*	0.2218094
Junior-Assistant	Average Impact Factor	1	0***	0***
Senior-Junior	Median Impact Factor	1	1	0.480719
Senior-Assistant	Median Impact Factor	0.1692595	0.3290173	0.9767523
Junior-Assistant	Median Impact Factor	0.999999	0***	0.000000656***

3.2 Complementarities in scientific output and quality: A supermodular non-parametric approach.

Up to the previous section I had based the analyses of complementarities solely on a non-parametric modification of the procedure developed by Athey and Stern; I used standard regression techniques to obtain an estimator of the scientific output and quality of research laboratories and individual scientists and studied the existent complementarities given a set of dummy variables capturing the complementary and intermediate states of couples of explanatory variables defined by the decomposition of human and financial resources into researcher categories, coauthor types and funding sources.

Following these estimations I retrieved the underlying complementarities with the help of a one sided test on the differences in the effects of the sum of the complementary state coefficients and the sums of the intermediate state coefficients, which expresses the condition of increasing differences, and verifies the property of supermodularity implying the existence of complementarities of a real valued function in a given pair of arguments.

However, the method developed in the previous section to asses the existence of complementarities between pairs of researcher categories may present a few shortfalls, for instance, when defining the complementary and intermediate states of the units of analysis given a pair of arguments we may find that in some cases, the sample of laboratories and individual researchers in the analysis may present only a few or even lack observations for one or more of these states. As an example, in the case of the states defined for the couple of arguments *senior-junior* affecting the output of research laboratories, if neither the count of senior researchers nor the count of junior researchers is over the fixed value (median counts within the disciplinary field or direction of the evolution in time) we will be in absence of observations describing the upper complementary state (couple11); in this case, the estimation of the effects of the complementary and intermediate states on the dependent variable will only include the effects captured by the remaining dummy variables (couple00, couple01 and couple10) creating a distortion in our hypothesis test that would assess the existence of increasing differences.

The solution to this problem lays on the fact that the supermodularity framework allows the study of the real values of the objective function without imposing any functional form or any condition of continuity and concavity on it; therefore, I adapted and carried the analysis on the actual observed values of scientific output and quality following a non-parametric procedure based on Beresteanu, 2005. For this purpose I drew a non-parametric estimator of the real valued function under study so that it reflected the phenomena under the effects of one of our complementary or intermediate states. The non-parametric estimator was obtained through a sample of combination of real valued observed in the dataset.

To verify that our data fulfils the condition of increasing differences under the framework of supermodular functions, I built several samples of combinations of 4 different units of analysis (collective and individual scientific output and quality) randomly drawn from their original dataset throughout a process of sampling without repetition that would always ensure that one observation is defined for each one of four states. I built such samples with objective of obtaining a set of four real values of the scientific production and quality function of research laboratories and institutional researchers that may fit the condition of increasing differences given the partially ordered set of arguments (categories of researchers, categories of funding sources and categories of coauthors depending on the analysis of interest).

Each one of the observations in a given combination is ranked from 1 to 4 according to its respective laboratory share of researcher categories of interest for a given pair wise comparison so that each observation may represent one of the four possible states of the objective function within the condition of increasing differences being the complementary and the intermediate states[21]. Once each of the four observations is identified with its position within the increasing differences inequality we may evaluate it to assess whether or not the sum of complementary states reflect a higher outcome than the sum of intermediate states.

$$y_1 + y_2 \geq y_3 + y_4$$

By applying simple algebra one may reinterpret the condition as the ratio of the distance between complementary states to the sum of the distances between the initial and intermediate states and assess which of these distances is greater.

$$y_1 + y_2 \geq y_3 + y_4$$
$$\Longleftrightarrow y_1 + y_2 - 2y_1 \geq y_3 + y_4 - 2y_1$$
$$\Longleftrightarrow (y_2 - y_1) / [(y_3 - y_1) + (y_4 - y_1)] - 1 \geq 0$$

If the scientific production and quality shows complementarities within its pairs of researcher categories, then the variation of the distance of the output or quality, when the level of researchers in both categories moves jointly in the same direction towards a higher element in the order with respect to the initial state, should be greater than the

[21] The complementary states are defined by:
State 1: The initial state, or where both arguments of the function are at their lower levels within the sample. This state is equivalent to model 1 in the analysis of the estimated output.
State 2: The state where there is a joint evolution of the arguments with respect to the state1. This model is equivalent to model 2 in the analysis of the estimated output.
State 3 and 4: The state where there has only been an evolution of one or the other argument with respect to state1. These states are equivalent to models 3 and 4 in the analysis of the estimated output.

variation of the sum distances between the observed output or quality when the level of researchers in each category moves separately.

I 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 (x_k, x_l), or shares of researcher categories[22]. Then I 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 lowest shares of researchers in each category, rank 2 being the research group with the highest shares of researchers in each category, and ranks 3 and 4 being the other two research groups with intermediary shares of researchers.

As an example or illustration, imagine four research laboratories randomly drawn from the sample and call them alpha, beta, gamma, and delta; then rank them according to a pair of categories of researchers (seniors and juniors), the first research group, alpha, has a share of 10% senior and 10% junior researchers, therefore it holds rank 1, the second research group, beta, has a share of 40% senior and 40% junior researchers and it holds rank 2, through their observed production these two laboratories represent the complementary states in the combination; the other two laboratories, gamma and delta, have shares of 20% senior and 30% junior, and 30% senior and 20% junior researchers respectively, and thus they hold ranks 3 and 4, and their observed production represents the intermediate states in the combination.

In order to implement this method, I obtained an unbiased estimator of the output and quality using a non-parametric bootstrap estimation with one hundred repetitions calculating and storing the estimated values of the dependent variable. For this purpose I used the exact same regression models described in chapter 3 on the determinants of scientific production and quality.

The next step in the analysis was to generate several sets of observations containing an element for each of the four states according to the procedure described above to establish the following indicator:

$$(\hat{y}_2 - \hat{y}_1) / [(\hat{y}_3 - \hat{y}_1) + (\hat{y}_4 - \hat{y}_1)]$$

[22] We recall the categories (x_k, x_l) are defined by the following couples: (Seniors, Juniors), (Seniors, Post doctors and PhD’s), (Seniors, Support Staff), (Juniors, Post doctors and PhD’s), (Juniors, Support Staff), (Post doctors and PhD’s, Support Staff).

This indicator captures the distance between these estimated values when their arguments evolve jointly and separately[23], therefore preserving the existence of pair wise complementarities among the arguments defining the function.

The process was repeated several times to obtain a random sample without replacement of evaluations of the total contributions, the total fractional publications and the average and median impact factor of publications according to all three models of publication aggregation in the laboratories.

In fact, if the scientific production and quality presents pair wise complementarities among its arguments, then the distance between the estimated output or quality at any initial level of two arguments and their estimated value when both arguments have increased jointly should be larger than the sum of distances from their estimated value at the initial level to the estimated value when only one argument or the other has increased.

Hence for the sample of laboratories I generated the statistic based on the distance indicator, which in presence of complementarities between the researcher categories any two categories of arguments *k* and *l* should be greater than 1.

$$\Delta = d_{comp} / (d_{inter\ category\ k} + d_{inter\ category\ l})$$

To assess the existence of complementarities between the couples of arguments I performed a one-sided test of the type $H_0: \Delta\text{-}1 = 0, H_1: \Delta\text{-}1 \geq 0$ where the null hypothesis indicates that the distance between two estimated values of the scientific production or quality given a joint variation of a couple of arguments is equal to the sum of distances between the estimated values given a separate variation of the arguments. The alternative hypothesis on the other hand indicates that the distance between the estimated values in the case of a joint variation of arguments is larger than the sum of distances between the values in case of a separate variation of those arguments, revealing the existence of increasing differences and verifying the existence of complementary relationships among the couple of arguments in question.

This analysis on the estimated values of the scientific production and quality of research laboratories assesses the existence of complementarities within the process of scientific research production and quality based on a supermodular non-parametric framework and attempts to be more realistic than a simple parametric analysis of these phenomena.

[23] Where the distance between complementary states is defined by $d_{comp} = (\hat{y}_2 - \hat{y}_1)$, and the distance between intermediate states is defined by $d_{inter\ category\ k} = (\hat{y}_3 - \hat{y}_1)$ and $d_{inter\ category\ l} = (\hat{y}_4 - \hat{y}_1)$.

3.2.1 Human resource complementarities in collective scientific production and quality under the non-parametric approach on supermodularity.

The results from the analyses on the scientific output of research laboratories under the non-parametric approach reported on Set of **Tables *20*** show that the total contributions to publications present complementary relationships in all categories of researchers during the whole period of analysis and under all three assumptions on the output aggregation in the case of contributions to publications and fractional publications.

Regarding the relationship among seniors and technical staff (assistant researchers), the evidence of complementarities in all periods of time confirms some of the insights obtained in the previous section; the fact that these complementary links are constant in time across all periods reveals that the scientific output relies in on a strong link between couples of researchers.

Set of Tables 20 Increasing differences in types of researchers by period (estimated laboratory output)

CONTRIBUTIONS TO PUBLICATIONS					
Model	**Pair of researchers**	**1996**	**2000**	**2004**	**2008**
Model 1: Aggregation over the 2 years preceding the surveys	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior**	4.16e-08**	6.74e-09**	2.26e-08**	4.73e-06**
	Senior-PhD Postdoctor**	1.72e-09**	0.1081446	4.13e-08**	2.53e-14**
	Senior-Assistant**	.0000765**	0.1092477	2.54e-06**	1.59e-10**
	Junior-PhD Postdoctor**	1.52e-13**	3.05e-07**	8.91e-11**	1.67e-12**
	Junior-Assistant**	8.24e-12**	1.77e-08**	4.98e-10**	1.30e-07**
	PhD Postdoctor-Assistant**	4.23e-15**	1.43e-06**	.0091689**	2.50e-16**
Model 2: Aggregation over the 2 years following the surveys	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior**	4.35e-16**	7.49e-07**	5.37e-10**	.0002597**
	Senior-PhD Postdoctor**	3.68e-14**	2.75e-06**	1.11e-12**	1.09e-07**
	Senior-Assistant**	1.55e-13**	3.96e-13**	9.55e-06**	2.75e-10**
	Junior-PhD Postdoctor**	6.54e-15**	3.25e-10**	1.05e-13**	1.57e-09**
	Junior-Assistant**	3.88e-12**	5.14e-13**	1.09e-08**	7.75e-07**
	PhD Postdoctor-Assistant**	1.30e-14**	9.85e-11**	2.00e-16**	3.51e-14**
Model 3: Aggregation from 2 years before to 2 years after survey	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior**	.0000143**	3.75e-08**	8.37e-06**	.0000195**
	Senior-PhD Postdoctor**	1.60e-10**	4.35e-08**	2.23e-08**	2.19e-10**
	Senior-Assistant**	2.81e-11**	5.00e-12**	.0000141**	.0059764**
	Junior-PhD Postdoctor**	9.38e-07**	3.90e-09**	7.54e-08**	8.62e-12**
	Junior-Assistant**	1.15e-10**	2.39e-13**	2.31e-07**	.0131172**
	PhD Postdoctor-Assistant**	7.09e-13**	3.75e-14**	7.10e-08**	8.07e-16**

FRACTIONAL PUBLICATIONS					
Model	Pair of researchers	1996	2000	2004	2008
Model 1: Aggregation over the 2 years preceding the surveys	Couple	Prob.H0	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior**	2.30e-08**	.000043**	.0010154**	.0068741**
	Senior-PhD Postdoctor**	3.96e-08**	1.85e-06**	.0002075**	.0012039**
	Senior-Assistant**	9.53e-07**	1.15e-07**	.0000469**	.0046768**
	Junior-PhD Postdoctor**	1.71e-11**	9.08e-06**	.0002039**	.0022358**
	Junior-Assistant**	2.03e-09**	.0002572**	2.31e-06**	.0388305**
	PhD Postdoctor-Assistant**	5.60e-08**	4.07e-12**	8.91e-14**	.0002847**
Model 2: Aggregation over the 2 years following the surveys	Couple	Prob.H0	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior**	8.75e-09**	.0000562**	1.27e-06**	.0008458**
	Senior-PhD Postdoctor**	6.27e-07**	.0002957**	2.90e-09**	.0106984**
	Senior-Assistant**	1.89e-10**	7.50e-12**	9.63e-13**	.0000837**
	Junior-PhD Postdoctor**	.000042**	4.80e-09**	3.32e-12**	.0127635**
	Junior-Assistant**	2.76e-06**	2.38e-08**	3.20e-09**	.00624**
	PhD Postdoctor-Assistant**	.0000828**	7.60e-11**	5.73e-13**	.0049295**
Model 3: Aggregation from 2 years before to 2 years after survey	Couple	Prob.H0	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior**	4.09e-08**	.0000425**	.0010374**	.0149012**
	Senior-PhD Postdoctor**	2.37e-11**	5.97e-07**	.0000793**	.0116709**
	Senior-Assistant**	2.20e-06**	2.48e-07**	.0001897**	.0059488**
	Junior-PhD Postdoctor**	2.85e-12**	1.56e-06**	1.24e-07**	.018431**
	Junior-Assistant**	2.83e-06**	.0001677**	.000034**	0.0543545
	PhD Postdoctor-Assistant**	1.22e-09**	2.83e-11**	1.46e-13**	.0035609**

When looking for complementarities of the scientific quality of research laboratories in couples of researcher categories whose results are reported on Set of **Tables *21***, I find that increasing differences exist only in a few couples such as *junior-PhD post doctor* for all three models of output aggregation related to the survey of 2000, then in the couples *junior-assistant* and *assistant-PhD post doctors* for models 2 and 3 related to the same survey.

It is important to notice that these complementary relationships only hold one period, which implies the existence of rather weak or unstable complementary links between the younger researchers and the technical staff, which have an effect on the scientific quality of research laboratories.

This supermodular non-parametric analysis of complementarities between different researcher categories tells us that both, output and quality of research laboratories have complementary relationships between almost every single couple of researcher categories, suggesting that there is a cascade of complementarities along the hierarchic structure of researcher ranks.

Finally, since the data reflects the existence of complementarities between highly experienced researchers and less experienced ones using the non-parametric method method, I am able to conclude there is a certain “research delegation effect” towards and assistants or support staff, which is also supported to some extent by the results from the parametric approach. Research delegation in this case implies the existence of strong links between different sections of the scientific production distribution curve, which would partially explain its positive skewedness (Lotka, 1926).

Set of Tables 21 Increasing differences in types of researchers (estimated laboratory quality)

AVERAGE IMPACT FACTOR				
Model	**Pair of researchers**	**2000**	**2004**	**2008**
Model 1: Aggregation over the 2 years preceding the surveys	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior	0.300622	0.6790571	0.9999998
	Senior-PhD Postdoctor	0.3279824	0.463814	1
	Senior-Assistant	0.0980511	0.8273572	1
	Junior-PhD Postdoctor**	.0387401**	0.9808275	1
	Junior-Assistant	0.1130782	0.9990954	0.9999997
	PhD Postdoctor-Assistant**	.0238697**	0.9488193	1
Model 2: Aggregation over the 2 years following the surveys	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior	0.2120664	0.3977979	1
	Senior-PhD Postdoctor	0.1708586	0.7433521	1
	Senior-Assistant	0.0742965	0.8274387	1
	Junior-PhD Postdoctor**	.0051626**	0.888234	1
	Junior-Assistant**	.0354639**	0.8746132	1
	PhD Postdoctor-Assistant**	.0111147**	0.992299	1
Model 3: Aggregation from 2 years before to 2 years after survey	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior	0.3698976	0.655158	1
	Senior-PhD Postdoctor	0.3511089	0.7448702	1
	Senior-Assistant**	.0253335**	0.9230397	1
	Junior-PhD Postdoctor**	.0200323**	0.9000999	1
	Junior-Assistant**	.0265669**	0.9745845	1
	PhD Postdoctor-Assistant**	.0027016**	0.8555346	1

MEDIAN IMPACT FACTOR				
Model	Pair of researchers	2000	2004	2008
Model 1: Aggregation over the 2 years preceding the surveys	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior	0.200277	0.3772919	0.9999945
	Senior-PhD Postdoctor	0.4703613	0.3488405	1
	Senior-Assistant	0.7728555	0.3100204	1
	Junior-PhD Postdoctor**	.0318646**	0.2536734	1
	Junior-Assistant	0.5029187	0.6413443	0.9999877
	PhD Postdoctor-Assistant	0.1045952	0.8407483	1
Model 2: Aggregation over the 2 years following the surveys	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior	0.1524911	0.2085558	0.8272485
	Senior-PhD Postdoctor	0.1024997	0.8304955	1
	Senior-Assistant	0.2081617	0.8233722	0.9998818
	Junior-PhD Postdoctor**	.0055936**	0.1466051	1
	Junior-Assistant	0.1901129	0.3995839	0.9999153
	PhD Postdoctor-Assistant	0.0828264	0.9144211	1
Model 3: Aggregation from 2 years before to 2 years after survey	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior	0.3279318	0.1948134	0.9999887
	Senior-PhD Postdoctor	0.0654881	0.596324	1
	Senior-Assistant	0.4114061	0.5837656	0.9999621
	Junior-PhD Postdoctor**	.0149409**	0.2701349	1
	Junior-Assistant	0.1319939	0.2742544	0.9999835
	PhD Postdoctor-Assistant	0.2264678	0.8886082	1

3.2.2 Research funds complementarities in the collective scientific production and quality.

Moving on to the next analysis, I deal with the complementary relationships between different types of funding sources available under the non-parametric approach. I used the non-parametric procedure described at the beginning of the section to generate an indicator that would capture the increasing differences of the real valued function under study (scientific output and quality) on the different types of funding sources. As described earlier in this section, the first step was to generate an estimation of the output (total contributions and total fractional publications) and its quality (average and median impact factor of laboratory publications) using a bootstrapped regression of the models described in chapter 3.

After obtaining the unbiased coefficients, I generated an estimated output for each observation and then extracted several random sets of four observations from the dataset with the objective of building a new dataset of increasing differences, with each element of an observation fulfilling a certain position that indicates either a complementary or an intermediate state given a pair of arguments and once this dataset of elements within the increasing differences condition was obtained we defined the distance indicator to be tested.

In this case, my interest was to test whether the dependent variables (output and quality) display increasing differences in the couples referring to the different types of funding sources[24] available at the laboratory during the surveys of 2004 and 2008. The results reported on Set of **Tables *22*** show the existence of only a few complementary relations between some of the funding source couples.

Set of Tables 22 Increasing differences in types of research funds. Estimated laboratory production

CONTRIBUTIONS TO PUBLICATIONS			
Model	**Pair of funding**	**2004**	**2008**
Model 1: Aggregation over the 2 years preceding the surveys	**Couple**	**Prob.H0**	**Prob.H0**
	Total_Private-Total_Public	0.2695217	
	Private-Public		0.9759133
	Regional-European	0.9682976	
	Recurrent_Public-Private**	.0376619**	
	European-Private	0.2534162	0.9994082
	Regional-Private	0.1549786	0.6320802
Model 2: Aggregation over the 2 years following the surveys	**Couple**	**Prob.H0**	**Prob.H0**
	Total_Private-Total_Public	0.2674462	
	Private-Public		0.9358592
	Regional-European	0.9605178	
	Recurrent_Public-Private**	.0384975**	
	European-Private	0.2544527	0.9996243
	Regional-Private	0.1749068	0.6170775
Model 3: Aggregation from 2 years before to 2 years after survey	**Couple**	**Prob.H0**	**Prob.H0**
	Total_Private-Total_Public	0.2589121	
	Private-Public		0.9673207
	Regional-European	0.9716551	
	Recurrent_Public-Private**	.0434063**	
	European-Private	0.2410059	0.999585
	Regional-Private	0.1618521	0.6816107

[24] Let us recall that the couples of funding sources in our analysis are: *{regional funding ; European funding}*, *{private funding ; European funding}*, *{private funding ; regional funding}*, *{private funding ; recurrent public funding}* and finally *{private funding ; other public funding}*.

FRACTIONAL PUBLICATIONS			
Model	Pair of funding	2004	2008
Model 1: Aggregation over the 2 years preceding the surveys	Couple	Prob.H0	Prob.H0
	Total_Private-Total_Public	0.8300417	
	Private-Public		0.8718342
	Regional-European	0.9998452	
	Recurrent_Public-Private	0.0914455	
	European-Private	0.9722062	0.9991688
	Regional-Private	0.3130662	0.9956647
Model 2: Aggregation over the 2 years following the surveys	Couple	Prob.H0	Prob.H0
	Total_Private-Total_Public	0.7959827	
	Private-Public		0.8639067
	Regional-European	0.9974005	
	Recurrent_Public-Private**	.0222785**	
	European-Private	0.9823137	0.9999369
	Regional-Private	0.5482571	0.9971278
Model 3: Aggregation from 2 years before to 2 years after survey	Couple	Prob.H0	Prob.H0
	Total_Private-Total_Public	0.8142179	
	Private-Public		0.8812107
	Regional-European	0.9995877	
	Recurrent_Public-Private	0.1074772	
	European-Private	0.9731717	0.9997818
	Regional-Private	0.258199	0.9977587

In fact, the contribution to publications display increasing differences only in the couple of recurrent public funding and private funding, that is between the regular allowance research laboratories obtain from public organisms such as the CNRS, INSERM or the ministry of higher education and research on a periodic basis and private funding obtained from competitive allowances and research commercialisation under all three models of output aggregation and for the information related to the survey of 2004. In addition, the total fractional publications also display increasing difference in the same couple, *recurrent public-private funding*, under the assumption of publication aggregated over two years following the survey.

Finally, when studying complementarities between different types of funding sources on the scientific quality of research laboratories the results reported on Set of **Tables *23*** show that neither the average nor the median impact factor of publications performed by our set of laboratories present any increasing differences in any couple of research funding sources.

These analyses destined to assess the existence of complementarities between different sources of research funds show that the collective scientific production in research laboratories may be more important in the case where levels of recurrent public allocations are complementary with the levels of private research funds obtained; that is, the laboratory’s best option is to concentrate proportional efforts into defending the evolution of their public allowances and at the same time look for competitive and commercial private funds; however, even if this strategy can improve the amount of science produced, it cannot ensure an improvement of the research quality of the laboratory since there is no evidence that a complementary relation between these two variables has any effect on the output quality.

Set of Tables 23 Increasing differences in types of research funds (estimated laboratory quality)

AVERAGE IMPACT FACTOR			
Model	**Pair of funding**	**2004**	**2008**
Model 1: Aggregation over the 2 years preceding the surveys	**Couple**	**Prob.H0**	**Prob.H0**
	Total_Private-Total_Public	0.9797208	
	Private-Public		0.9457175
	Regional-European	0.9991354	
	Recurrent_Public-Private	0.8390927	
	European-Private	0.999979	0.8806873
	Regional-Private	0.9998866	0.9869396
Model 2: Aggregation over the 2 years following the surveys	**Couple**	**Prob.H0**	**Prob.H0**
	Total_Private-Total_Public	0.9974678	
	Private-Public		0.8386776
	Regional-European	0.9991339	
	Recurrent_Public-Private	0.844889	
	European-Private	0.9999995	0.9556921
	Regional-Private	0.9998856	0.8637914
Model 3:Aggregation from 2 years before to 2 years after survey	**Couple**	**Prob.H0**	**Prob.H0**
	Total_Private-Total_Public	0.9935848	
	Private-Public		0.8840326
	Regional-European	0.9981198	
	Recurrent_Public-Private	0.8404682	
	European-Private	0.9999983	0.8725193
	Regional-Private	0.9999518	0.9073203

MEDIAN IMPACT FACTOR			
Model	Pair of funding	2004	2008
Model 1: Aggregation over the 2 years preceding the surveys	Couple	Prob.H0	Prob.H0
	Total_Private-Total_Public	0.8920056	
	Private-Public		0.9585014
	Regional-European	0.9990097	
	Recurrent_Public-Private	0.8349995	
	European-Private	0.9991102	0.9603256
	Regional-Private	0.999944	0.9731002
Model 2: Aggregation over the 2 years following the surveys	Couple	Prob.H0	Prob.H0
	Total_Private-Total_Public	0.9763095	
	Private-Public		0.9782993
	Regional-European	0.9968733	
	Recurrent_Public-Private	0.7930451	
	European-Private	0.9999697	0.9812296
	Regional-Private	0.9995354	0.8739995
Model 3:Aggregation from 2 years before to 2 years after survey	Couple	Prob.H0	Prob.H0
	Total_Private-Total_Public	0.6636305	
	Private-Public		0.9658259
	Regional-European	0.9927326	
	Recurrent_Public-Private	0.8149142	
	European-Private	0.9719517	0.9355955
	Regional-Private	0.9999474	0.9529929

3.2.3 Complementarities between coauthors on individual scientific production and quality

In the present subsection I carry out a third analysis, which deals with the actual scientific output of individual researchers and I look for complementarities between the different types of coauthors an individual researcher may have. Using the same procedure developed in the previous sections, I establish several sets of four different individual researchers for whom the levels of their different types of coauthors reflect one of the four complementary or intermediate states, which will then allow me to build a set of increasing differences indicators.

Please note here that I look for complementarities between senior, junior, and assistant coauthors given that the information on individual coauthors only provides the levels for these three categories (see chapter 2 on datasets).

We proceed with a random selection of several thousands of combinations of four different individuals with the help of a random sampling without replacement; then for any given combination I verified that the individuals fulfilled the following conditions:

- One of the individuals presents the lowest level of coauthors in both categories of interest[25], which represents the initial state.
- One of the individuals presents the highest level of coauthors in both categories, representing the state of joint evolution of coauthors.
- Each of the two other individuals presents an intermediate level in one or the other category of coauthors. These two individuals represent the two intermediate states.

Once the final sample of combinations of researchers fulfilling the conditions was obtained, I proceeded with the definition and analysis of the distance between the complementary states, and the sum of distances between the initial state and each one of the intermediate states captured by our distance indicator[26] defined in the previous section.

Set of Tables 24 Increasing differences in types of coauthors (estimated individual output)

CONTRIBUTIONS TO PUBLICATIONS					
Model	**Pair of researchers**	**1996**	**2000**	**2004**	**2008**
Model 1: Aggregation over the 2 years preceding the surveys	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior	0.1543888	0.0789805	0.0560965	0.2436056
	Senior-Assistant	0.1589771	0.1269665	0.1710837	0.1449066
	Junior-Assistant	0.1719429	0.1004676	0.1854354	0.2174144
Model 2: Aggregation over the 2 years following the surveys	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior**	0.0879499	0.0239058**	0.1623289	0.103705
	Senior-	0.1561231	0.0021593**	.0466709**	0.1995428
	Junior-Assistant	0.1513886	0.0914238	0.1609381	0.1457872
Model 3: Aggregation from 2 years before to 2 years after survey	**Couple**	**Prob.H0**	**Prob.H0**	**Prob.H0**	**Prob.H0**
	Senior-Junior**	.0297513**	0.1580278	.033246**	0.0837586
	Senior-	0.146765	.0070258**	.0277839**	0.1745171
	Junior-	0.1575805	.0230891**	0.0812788	0.1013956

[25] With the categories of interest being the couples of coauthors: (*senior, junior*), (*senior, assistant*), *and* (*junior, assistant*).

[26] $\Delta = d_{comp} / (d_{inter\ category\ k} + d_{inter\ category\ l})$, where the numerator defines the distance between the complementary states and the denominator defined the sum of distances between the intermediate and the initial-complementary state.

FRACTIONAL PUBLICATIONS					
Model	Pair of researchers	1996	2000	2004	2008
Model 1: Aggregation over the 2 years preceding the surveys	Couple	Prob.H0	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior**	0.0670052	0.69629	.0401055**	0.9579129
	Senior-Assistant	0.8835221	0.1310252	0.4027028	0.4089992
	Junior-	0.8882222	0.6363296	0.3413433	.0105353**
Model 2: Aggregation over the 2 years following the surveys	Couple	Prob.H0	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior**	0.3866439	.027882**	.0419013**	0.2428835
	Senior-Assistant	0.1745176	0.0528788	0.3744994	0.8925235
	Junior-Assistant	0.9928056	0.5286864	0.1162486	0.3660666
Model 3: Aggregation from 2 years before to 2 years after survey	Couple	Prob.H0	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior**	0.2430106	.02419**	.0064984**	0.0892956
	Senior-Assistant	0.1642112	0.0889791	0.1067015	0.1724265
	Junior-Assistant	0.9271279	0.9110247	0.8330892	0.1101841

Once I had established the distance indicators, I performed a one sided test to assess whether the null hypothesis[27] is rejected, which in this case will help me conclude that the distance between the individual scientific outputs responding to joint variations in the levels of each category of coauthors is higher than the sum of distances between the individual outputs related to separate variations of coauthors in the levels of one or the other category, hence implying the existence of complementarities between the two categories of coauthors.

The results from this analysis of complementarities on the estimated individual contributions reported on Set of **Tables *24*** confirm the existence of increasing differences in the couple of *senior* and *junior* coauthors during the period related to the 2004 survey and between the couple of *senior* and *assistant* coauthors during both periods related to 2000 and 2004 under the assumption of publication aggregation over two years following the survey (model 2).

[27] H_0: $\Delta - 1 = 0$

Set of Tables 25 Increasing differences in types of coauthors (estimated individual quality)

AVERAGE IMPACT FACTOR				
Model	Pair of researchers	2000	2004	2008
Model 1: Aggregation over the 2 years preceding the surveys	Couple	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior	0.9353501	0.9959733	0.9418592
	Senior-Assistant	0.5928715	0.9085156	
	Junior-Assistant	0.9994203	0.8346764	
Model 2: Aggregation over the 2 years following the surveys	Couple	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior	0.5246545	0.407362	0.9778306
	Senior-Assistant	0.9440508	0.9668617	0.8789026
	Junior-Assistant	0.3043694	0.9850513	0.8321254
Model 3: Aggregation from 2 years before to 2 years after survey	Couple	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior	0.4484839	0.9929512	0.9985628
	Senior-Assistant	0.5565575	0.9999999	0.7656929
	Junior-Assistant	0.8671036	0.9999985	0.7868726

MEDIAN IMPACT FACTOR				
Model	Pair of researchers	2000	2004	2008
Model 1: Aggregation over the 2 years preceding the surveys	Couple	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior	0.9997478	0.9931945	0.9723084
	Senior-Assistant	0.9059151	0.9774473	
	Junior-Assistant	0.9995039	0.7618114	
Model 2: Aggregation over the 2 years following the surveys	Couple	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior	0.9909393	0.6738614	0.9975945
	Senior-Assistant	0.4503874	0.963338	0.7432896
	Junior-Assistant	0.5500327	0.9692501	0.8913579
Model 3: Aggregation from 2 years before to 2 years after survey	Couple	Prob.H0	Prob.H0	Prob.H0
	Senior-Junior	0.9571921	0.9817952	0.999241
	Senior-Assistant	0.8952246	0.9999999	0.8160908
	Junior-Assistant	0.2401886	0.9999967	0.8373017

In addition, in the case of model 3 (publications aggregated over a 4 years around the surveys) I find that all three couples of coauthors present increasing differences at some period in time; while the case of the fractional publications shows there are increasing differences in the couples of *senior* and *junior* coauthors at some periods for all three models and in the couple of *junior* and *assistant* coauthors in the case of model 3 for the survey of 2008.

On the other hand one may notice that the results concerning from analysis of complementarities among different types of coauthors on the individual scientific quality, reported on Set of **Tables** ***24***, do not provide any insights on the existence of any complementary relationships neither for the average nor for the median impact factor of individual publications.

A further look into the results obtained from the analysis of complementarities of coauthor categories on the individual scientific production and quality allows us to conclude that the individual scientific production benefits from collaboration with researchers located at every level of the scientific personnel hierarchy. That is, the individual research output will be more important when a researcher decides to increase his/her collaborations with more experienced researchers and technical researchers at the same time rather than just increasing the collaborations with only experienced or with only technical researchers.

These results comfort the main hypothesis of this research, which is that scientific production does not only rely on star scientists but also on less experienced researchers; they provide insights on the existence of a cascade of complementarities from highly experienced researchers to younger researchers and the technical staff that influences the individual scientific production measured either by contributions or by fractional publications; we may interpret these complementarities as a form of “research delegation” within the scientific production.

4 Conclusions

Analyses on the complementary nature of different kinds of researchers issued from the decomposition of the human researchers and the complementary nature of different kinds of project-funding issued from the decomposition of funding were performed using an original functional study approach based on the theory of supermodularity. The novelty of the present research lays in the use of this approach to study pair wise complementary relationships between couples of determinants of the scientific production process.

The supermodular study used to highlight complementarities between explanatory variables was undertaken under two different mechanisms. The first is a parametric approach in which a set of indicators point out the complementary and intermediary states of two determinants are run along a regression analysis with the objective of evaluating whether the effects of the sum of complementary states are higher than the effects of the intermediary states, thus concluding the existence of a complementary relationship between the two determinants.

The second mechanism is a non-parametric approach in which random combinations of observations are bundled together with each observation representing on of the possible states among the upper and lower complementary and the upper and lower intermediary state of the determinants. These bundles of observations are then used to create an index representing the ratio of the distance between the complementary states related the distance of between the initial and the intermediary states which in the case providing insights on the whether the complementary relationships exist between the determinants.

The results from these different analyses provide insights on the complementary relationships between different variables defining the decomposition of human resources and funding structure. There is evidence on the existence of important complementary link between senior researchers and technical staff and between junior researchers and technical staff enhancing the scientific production of research laboratories. In addition, there is also evidence on the existence of complementary links between private and regional funding and between private and public recurrent funding enhancing the scientific production of research laboratories whilst complementary relationships between senior and junior coauthors with influence on the individual scientific production were also found.

In terms of quality of research, influential complementary relationships were found between senior researchers and technical staff and between junior researchers and technical staff at both the laboratory level as head counts and at the individual level as types of coauthors collaborating, which reveals that delegating research tasks among different categories of researchers does have an impact on the scientific output and quality.

General Conclusion

General Conclusion

Throughout the present research work I have set in place a framework to study the output and quality of scientific laboratories and individual researchers; its first component is the literature base, which sets the grounds of the studies by connecting previous studies on different issues related to collective research. The literature reviewed in chapter one provides notions and findings on seven different axes: the collective nature of the scientific production, the concepts that create a support for such a collective process, the role that hierarchic structures have on within research organizations, notions related to the importance of reputation within the organizations, the possible funding structures and their influence on scientific research, the possible incentive structures linking quality to performance and finally the trade off between performing basic or applied science. These axes represent the support upon which the scientific production process can be studied stressing its character of collective (laboratories) and collaborative (individuals) process.

The second component of this framework is the data used to carry out the analysis of determinant playing a role on scientific production and quality and their complementary relationships. These datasets concern the public research laboratories of a well-known and recognised French university, the University Louis Pasteur; they containing information on the characteristics and composition of these laboratories, as well as the characteristics of individual members. In addition, the datasets present a high degree of reliability given that they come from official sources with juridical value, that is, the compulsory four-year ministry surveys that research laboratories affiliated to a university are to provide to the ministry of higher education and research. These surveys are then used by the ministry of higher education and research for evaluation purposes and decision-making concerning the operability of these research units.

The datasets containing information about the funding structure of these laboratories was gathered on a basis of project-funding amounts for research groups, which were matched to their respective laboratories. Finally, the output and quality of the scientific production of individual researchers was gathered from the Thomson Reuter’s Web of Science. Different matching processes allowed building consistent datasets providing a big picture of the scientific production of public research laboratories with detailed data on the output and quality of scientific research for both laboratories and individuals.

The third component of the framework are the different analyses were carried out with the objective of investigating the role determinants play on scientific production and quality, with particular focus on the effects of the composition of human resources and the nature of research project funding. These two bodies of explanatory factors were further decomposed into different categories enabling a deeper comprehension of the effects the structure of human resources and types of funding has on the scientific production.

A set of analyses focused on the research laboratory as the unit of observation and the aggregated output and research quality indicators of their production process were defined according to three different assumptions on the scientists' turnover across time[28]. The decomposition of the human resources was made according to the researcher ranks within the French academic system, while the decomposition of the funding structure was made according to the different types of funding received by projects. These analyses showed that senior researchers and public and private funding are determinant and stimulate the scientific performance of research laboratories. They also show that laboratories with higher counts of research-oriented researchers perform better than others where researchers with an institutional affiliation different than the university stimulate the scientific output of the laboratory. On the other hand, analyses regarding the quality of the output showed that institutional researchers and past output and quality play a major role on the current quality of output in research laboratories. These results indicate that not only quality is stimulated by institutional researchers (CNRS or INSERM) but also and most important, quality is path dependent, that is, previous

[28] *1)* Researchers affiliated to the laboratory during the period 1996 – 2000 may be accountable for the publications performed during a period of two years preceding the ministry survey of 2000 and 2.

2) They continued working for the laboratory during the period 2000 – 2004 and therefore may also be accountable for the laboratory publications signed during a period of two years following the ministry survey of 2000 (model 2).

3) If the two previous assumptions hold, then we may also assume that researchers declared during this ministry survey may accountable for the laboratory publications signed over a period of four years around that survey (model 3).

conditions of in the volume and quality of the scientific output play a major role on the present quality or scientific output.

Further analyses focused on the individual researcher as the unit of observation to provide a wider picture of scientific production. It showed that certain individual characteristics such as the seniority of the researcher and certain types of collaboration such as the senior and staff type stimulate individual scientific output. These results also provided insights on the fact that as researchers grow older, they produce more scientific output, though the role they play on the quality of publications is not as determinant as is the role junior researchers play as evidence shows that even if they produce less scientific output, they are determinant to the quality of the overall output. In addition, the analyses also provide insights on the fact that institutional researchers perform better than others and that past research quality is also an important factor determining future quality of research implying that trade offs between the time allocated to perform scientific research and the time allocated to perform teaching activities has a major impact on scientific production.

The fourth and final component of the framework is the study of the complementary nature of determinants of scientific production and quality. How different kinds of researchers issued from the decomposition of the human researchers and how different kinds of project funding issued from the decomposition of funding may present complementary relationships playing a role on scientific output and quality.

Several analyses were performed using an original functional approach based on the theory of supermodularity. The novelty of the present research lays in the use of this approach to study pairwise complementary relationships between couples of determinants of the scientific production process.

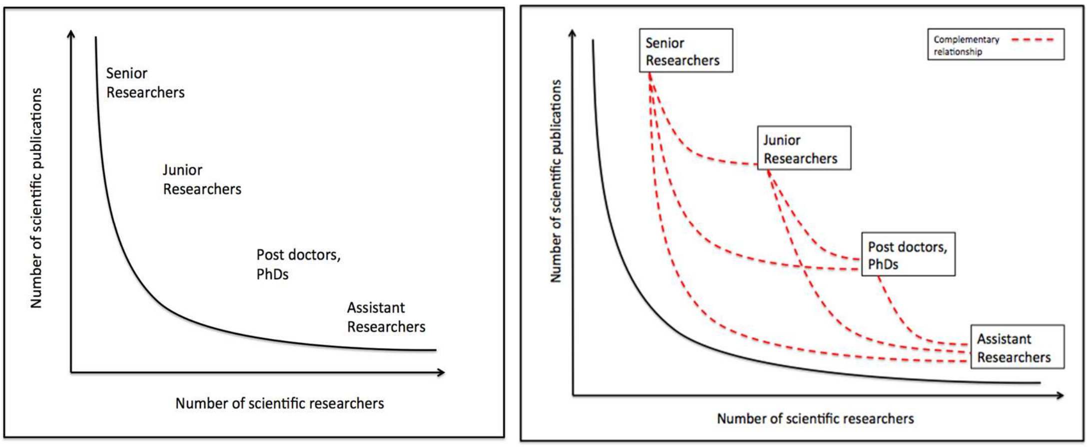
The supermodular study used to highlight complementarities between explanatory variables was undertaken according to two different mechanisms. The first is a parametric approach in which a set of indicators point out the complementary and intermediary states of two determinants are run along a regression analysis with the objective of evaluating whether the effects of the sum of complementary states are higher than the effects of the intermediary states, thus concluding the existence of a complementary relationship between the two determinants.

The second mechanism is a non-parametric approach in which random combinations of observations are bundled together with each observation representing on of the possible states among the upper and lower complementary and the upper and lower intermediary state of the determinants. These bundles of observations are then used to create an index representing the ratio of the distance between the complementary states related the distance of between the initial and the intermediary states which in the case providing insights on the whether the complementary relationships exist between the determinants.

The two approaches are valid for the application of supermodularity to study a function, although in the present case, the second non-parametric approach proved to be more suitable given the structure of the data. In fact, given that the study of the scientific production process is carried using a decomposition of the human resources and the funding structure that is defined by positive continuous and count data, it is necessary to impose a fixed threshold against which one is able to compare and tell whether each of complementary and the intermediary states are observed. This is translated into the creation of a set of dichotomous indicators that are plugged into the regression analysis to study whether the effects of the complementary states are higher than the effects of the complementary states; however some may interpret the creation of these indicators as too restrictive. The non-parametric approach is preferred given the fact that it can provide results straight from the data without the need of transforming the original variables.

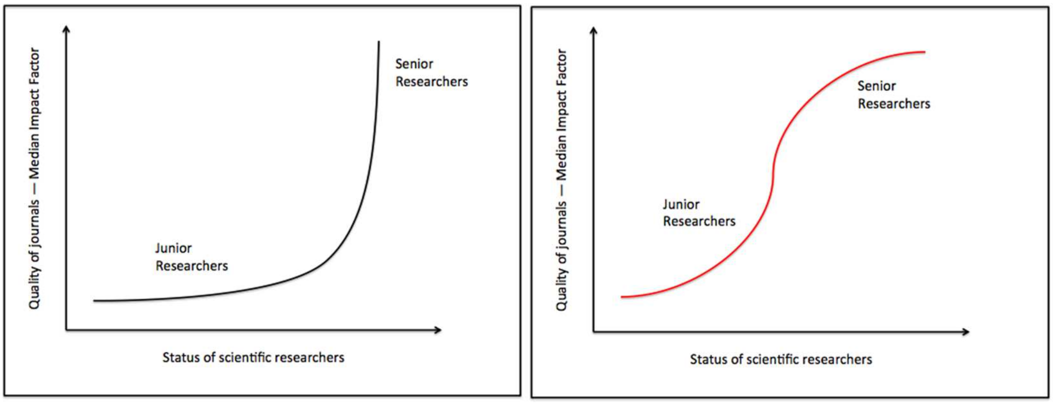
In a general overview, the picture I obtain from the present framework indicates that senior researchers, post doctors and PhD candidates and public funding are the main drivers of the scientific output volume in research laboratories. In addition, institutional researchers are also accounted for higher volumes of output. These key findings already provide an insight of relationship between highly experienced researchers and young post doctors and PhDs. Interestingly, this wider picture shows that post doctors and PhDs and institutional researchers play a major role on the quality of the scientific output of laboratories as if research oriented young individuals were already thinking ahead and setting the grounds of a reputational dynamic on their career trajectory. In addition, this wider picture shows the quality of the output is stimulated by private funding, and is path dependent, that is the volumes and quality of output produced during previous years plays a major role on current research production. Further analysis on complementary relationships between determinants provide this wider picture with evidence on the existence of pairwise complementarities between all types of researchers operating o the volume of scientific output, wile only pairwise complementarities between junior researchers and post doctors / PhDs and between assistant researchers and post doctors /

PhDs exist and operate on the quality of the output. Moreover, according to this evidence, the quality of the output presents pairwise complementarities between recurrent public funding and private funding. In essence, the analysis of the scientific production, particularly regarding the complementary relationships between determinants may be represented by the Set of Figures 17, which places the decomposition of scientific researchers on a generalised version of the scientific production distributions observed by Lotka (1926) and displays a cascade of pairwise complementary links between them, which is the main objective of the present work.



Set of Figures 17 Decomposition and complementarities of human resources on the distribution of scientific production

Furthermore, in addition to results regarding the decomposition of human resources, their impact on scientific production and the nature of the relationship between them, important insights on about the effects of the decomposition of human and funding resources on the quality of scientific output measured by the median and average impact factors of journals in which the output is published were found. Set of Figures 18 offers a graphical interpretation of how quality is influenced by the status of scientific researchers; the left hand figure proposes a widely adopted heuristic view according to which as researchers obtain experience and grow old they become wiser and hence the quality of their scientific output must be ever increasing. Opposed to this view I present in the right hand figure an interpretation of the results obtained throughout this work; it is during the early years of the scientific career that young researchers make the highest effort to publish their research in high impact factor journals, but after their career and reputation is consolidated the quality level of their research may only increase at a decreasing rate eventually stabilizing. What this interpretation offers graphically is an *S*-shaped curve that describes the evolution of scientific quality across different types of scientific researcher.



Set of Figures 18 Influence of researcher status on output quality

Finally, a graphical interpretation of the results on the decomposition of funding resources and its influence on the scientific production is offered in Figure 19; this figure displays in a Cartesian quadrant in I place the stimulus from private and public funding. The upper left hand side of the quadrant indicates that, according to results throughout this work, private funding stimulates quality of the scientific output, while the lower right hand side indicates that public funding stimulates the volume of the output. This figure also reflects the absence of insight on the nature of the relationship between different types of funding; as the analysis of complementarities between them did not provide any conclusive results regarding the existence of complementary links between types of funding on scientific quality, it is impossible to fill either the lower left hand or the upper right hand quadrants of the figure with any information.

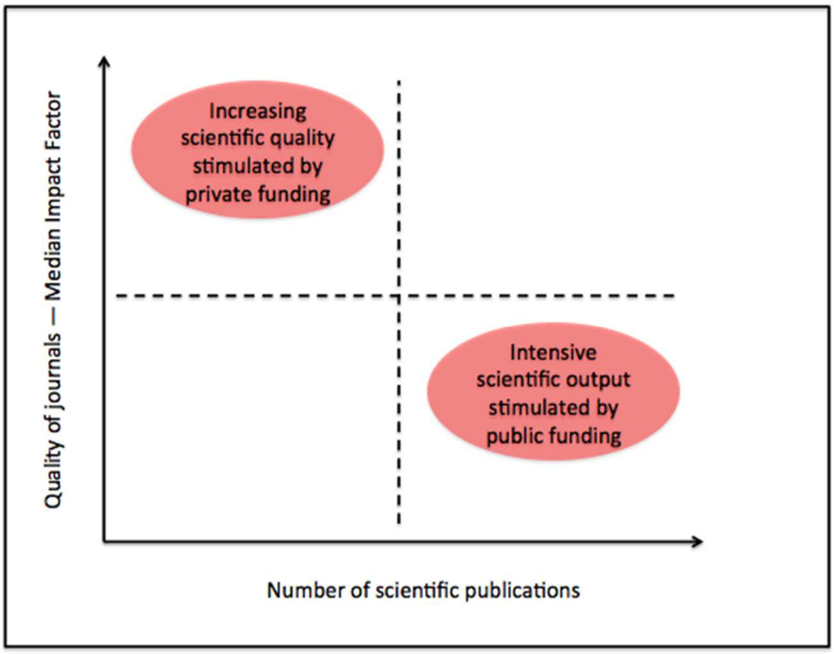


Figure 19 Stimulus of scientific research funding

I finalise the present work by recalling the current position of the European Union regarding scientific research. The European Commission[29] currently states its will for results maximisation by inviting Member States to:

- "Introduce or enhance competitive funding through calls for proposals and institutional assessments as the main modes of allocating public funds to research and innovation, introducing legislative reforms if necessary".
- "Ensure that all public bodies responsible for allocating research funds apply the core principles of international peer review".

In fact, the Commission believes that best practice performance of Member States in terms of scientific research solely relies on:

- Inciting researchers to reach international competitiveness by "allocating public funds though open calls for proposals evaluated by non-domestic experts".
- Conditioning institutional funding decision to the assessment of quality of research organisations.

The contribution of the present studies allows pointing out, in contrast to the Commission's position, that best practice performance in scientific research may instead

[29] See the European Research Area website: http://ec.europa.eu/research/era/more-effective-national-research-systems_en.htm

be increased by adopting a wider view on scientific production where a decomposition of human and funding resources are at the core of a collective process. Such a new point of view would imply:

- Utilising different types of funding, especially private funding in the collective scientific production process; this could be achieved by stimulating links and favoring joint programs between research teams and non-public stakeholders.
- Assessing the composition of research teams to stimulate the interaction between different types of researchers and take advantage of the complementary relationships between them.

Bibliographic References

P. Aghion and J. Tirole (1994). The management of innovation. *Quarterly Journal of Economics*. Vol 109 (4), pp 1185 – 1209.

A. Agrawal and R. Henderson (2002). Putting patents in context: Exploring knowledge transfer from MIT. *Management Science*. Vol 48 (1), pp 44 – 60.

A. Alchian and H. Demsetz (1972). Production, information costs, and economic organization. *American Economic Review*. Vol 62 (Dec), pp 777 – 795.

R. C. Allen (2003). Progress and poverty in early modern europe. *The Economic History Review*. Vol 56 (3), pp 403 – 443.

P. D. Allison and J. A. Stewart (1974). Productivity differences among scientists: Evidence for accumulative advantage. *American Sociological Review*. Vol 39 (4), pp 596 – 606.

R. Amir (2005). Supermodularity and Complementarity in Economics: An Elementary Survey. *Southern Economic Journal*. Vol 71 (3), pp 636 – 661.

N. Arnold (2007). The application of the concept of governance to the structures of german extra-university research organizations from a legal perspective. *New Forms of Governance in Research Organizations*, pp 175 – 186.

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

S. Athey and S. Stern (1998). An empirical framework for testing theories about complementarity ini organizational design. *NBER working paper*.

L. Bach and P. Llerena (2007). Indicators of higher-education institutes and public-research organizations technology transfer activities: insights from France. Science and Public Policy. Vol 34 (13), pp 709 – 721.

J. Barney (1991). Firm resources and sustained competitive advantages. *Journal of management.* Vol 17 (1), pp 99 –120.

S. Belenzon and M. Schankerman (2007). The impact of private ownership, incentives and local development objectives on university technology transfer performance. *Center for economic performance, LSE.*

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

A. Bonaccorsi and C. Daraio (2003). Age effects in scientific productivity. *Scientometrics.* Vol 58 (1), pp 49 – 90.

R. Boumahdi, N. Carayol, and P. Llerena (2008). The private contractual funding of academic laboratories: a panel data analysis. *Applied Economics Letters*. Vol 15 (6), pp 465 – 468.

S. Breschi, F. Lissoni, and F. Montobbio (2007). The scientific productivity of academic inventors: new evidence from italian data. *Economics of Innovation & New Technology*. Vol 16 (2), pp 101 – 118.

G. Buenstorf (2009). Is commercialization good or bad for science? individual-level evidence from the max planck society. *Research Policy*. Vol 38 (2), pp 281 – 292.

D. Bureau and M. Mougeot (2007). Performance, incitations et gestion publique. *Rapport du CAE*. Vol 66.

Q. Burrell and R. Rousseau (1999). Fractional counts for authorship attribution: A numerical study. *Journal of the American Society for Information Science*. Vol 46 (2), pp 97 – 102.

D. Cable and M. Graham (2000). The determinants of job seekers' reputation perceptions. *Journal of organizational behavior.* Vol 21 (8), pp 929 – 947.

M. Calderini, C. Franzoni (2007). If star scientists do not patent: The effect of productivity, basicness and impact on the decision to patent in the academic world. *Research Policy*. Vol 36 (3), pp 303 – 319.

Y. Caloghirou, G. Hondroyiannis, and N. S. Vonortas (2003). The performance of research partnerships. *Managerial and Decision Economics*. Vol 24 (2), pp 85 – 99.

N. Carayol and M. Matt (2004). Does research organization influence academic production?: Laboratory level evidence from a large European university. *Research Policy*. Vol 33 (8), pp 1081 – 1102.

N. Carayol and M. Matt (2006). Individual and collective determinants of academic scientists' productivity. *Information Economics and Policy*. Vol 18 (1), pp 55 – 72.

N. Carayol (2007). Academic incentives, research organization and patenting at a large French university. *Economics of Innovation and New Technology*. Vol 16 (2), pp 119 – 138.

C. Chambers and F. Echenique (2009). Supermodularity and preferences. *Journal of Economic Theory*. Vol 144 (3), pp1004 – 1014.

R. Coombs, M. Harvey, and B. Tether (2003). Analysing distributed innovation processes of provision qnd innovation. *Industrial and corporate change.* Vol 12 (6), pp 1125 –1155.

K. Cravens, E. Oliver and S. Ramamoorti (2003). The reputation index: Measuring and managing corporate reputation. *European management journal.* Vol 21 (2), pp 201 212.

M. Crow and B. Bozeman (1987). R&D laboratory classification and public policy: The effects of environmental context on laboratory behavior. *Research policy*. Vol 16 (5), pp 229 – 258.

D. Czarnitzki and A. Fier (2003). Publicly funded R&D collaborations and Patent Outcome in Germany. *ZEW - Center for European Economic Research, discussion papers.*

P. Dasgupta and P. David (1985). Information disclosure and the economics of science and technology. *CEPR Discussion Papers.*

P. Dasgupta and P. David (1994). Toward a New Economic of Science. *Research policy*. Vol 23 (5), pp 487 – 521.

A. Diamond (1986). What is a citation worth? *The Journal of Human Resources*. Vol 21 (2), pp 200 – 215.

A. Dixit (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*. Vol 37 (4), pp 696 – 727.

A. Dixit (2003). Clubs with entrapment. *The American Economic Review*. Vol 93 (5), pp 1824–1829.

G. Dosi (1982). Technological paradigms and technological trajectories : A suggested inter- pretation of the determinants and directions of technical change. *Research Policy.* Vol. 11 (3), pp 147 – 162.

D. Dougherty (1996). Organizing for innovation. *In S. Clegg, C. Hardy and W. Nord (eds). Managing organizations: current issues*. pp 174 – 189.

L. Egghe (2007). Distributions of the h-index and the g-index. *In D. Torres-Salinas and H. Moed (eds). Proceedings of the 11th international society for scientometrics and informatics. CSIC.* pp 245 – 253

L. Egghe (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*. Vol 60 (10), pp 2142 – 2144.

K. R. Fabrizio and A. D. Minin (2008). Commercializing the laboratory: Faculty patenting and the open science environment. *Research Policy*. Vol 37 (5), pp 914 – 931.

G. Filliatreau (2008). Indicateurs de sciences et de technologies. *Rapport de l'Observatoire des sciences et des techniques, édition 2008*.

F. Galia and D. Legros (2004). Testing for complementarities between team, incentives, training and knowledge management: Evidence from france. *Technical report, ERMES working paper*. Sept. 2004.

W. Glänzel (1996). The need for standards in bibliometric research and technology. *Scientometrics.* Vol 35 (2) pp, 167 – 176.

C. Gouriéroux, A. Holly, and A. Monfort (1982). Likelihood ratio test, wald test, and kuhn-tucker test in linear models with inequality constraints on the regression parameters. *Econometrica*. Vol 50 (1), pp 63 – 80.

M. Granovetter (1973). The strength of weak ties: a network theory revisited. *Sociological Theory, 1*. Vol 1, pp 201 – 233.

M. Granovetter (1985). Economic Action and Social Structure: the Problem of Embeddedness. *American Journal of Sociology*. Vol 91, pp 481 – 510.

S. J. Grossman and O. D. Hart (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *The Journal of Political Economy*, Vol 94 (4), pp 691 – 719.

H. Grupp and M. E. Mogee (2004). Indicators for national science and technology policy: how robust are composite indicators? *Research Policy*. Vol 33 (9) 1373 – 1384.

G. Guidetti, S. Mancinelli, and M. Mazzanti (2009). Complementarity in training practices methodological notes and empirical evidence for a local economic system. *The IUP Journal of Applied Economics*, Vol 8 (1), pp 39 – 56.

J. Hage and R. Clignet (1982). Coordination styles and economic growth. *Annals of the American Academy of Political and Social Science*. Vol 459, pp 77 – 92.

M. Harfi and C. Mathieu (2006). Mobilité internationale et attractivité des étudiants et des chercheurs. *Horizons Stratégiques*. Vol (1), pp 28 – 42.

R. Harper (1992). Looking at ourselves: an examination of the social organisation of two research laboratories. *In Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. pp 330 – 337.

T. Heinze and S. Kuhlmann (2007). Analysis of heterogeneous collaboration in the German research system with a focus on nanotechnology. *New Forms of Governance in Research Organizations*, pages 187–209.

T. Heinze and N. Arnold (2008). Governanceregimes im Wandel. Eine analyse des außeruniversitären, staatlich finanzierten Forschungssektors in Deutschland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Vol 60 (4), pp 686 – 722.

T. Heinze and S. Kuhlmann (2008). Across institutional boundaries?: Research collaboration in German public sector nanoscience. *Research Policy*. Vol 37 (5), pp 888 – 899.

T. Heinze, P. Shapira, J. D. Rogers, and J. M. Senker (2009). Organizational and institutional influences on creativity in scientific research. *Research Policy*. Vol 38 (4), pp 610 – 623.

M. Heller, R. Eisenberg, M. McKelvey and L. Orsenigo (2006). Can patents deter innovation? The anticommons in biomedical research. *In The economics of biotechnology, vol 2. pp 391* – 394.

S. Hemlin, C. Allwood, and B. Martin (2008). Creative knowledge environments. *Creativity Research Journal*. Vol 20 (2), pp196 – 210.

B. R. Holmstrom and J. Tirole (1989). The Theory of the Firm. *Handbook of industrial organization.* Vol 1. pages 61 – 133.

P. Jensen, J. Rouquier, and Y. Croissant (2008). Scientists connected with society are more active academically. *Science and Public Policy*. Vol 35 (7).

P. Jensen, J. Rouquier, and Y. Croissant (2009). Testing bibliometric indicators by their prediction of scientists' promotions. *Scientometrics*. Vol 78 (3), pp 467 – 479.

J. S. Katz and B. R. Martin (1997). What is research collaboration?. *Research Policy*. Vol 26 (1), pp 1 – 18.

D. A. Kodde and F. C. Palm (1986). Wald criteria for jointly testing equality and inequality restrictions. *Econometrica*. Vol 54 (5), pp 1243 – 1248.

B. Kogut and U. Zander (2003). A memoir and reflection: Knowledge and an evolutionary theory of the multinational firm 10 years later. *Journal of International Business Studies*. Vol 34 (6), pp 505 – 515.

P. Laredo and P. Mustar (2001). French research and innovation policy: Two decades of transformation. *In P. Laredo and P. Mustar Research and innovation policies in the new global economy: An international comparative analysis.* pages 447 – 496.

C. Leilich (2005). Flat or Steep Hierarchies in Research Institutes? Empirical Findings for the Max-Planck-Institutes. *Trier: IAAEG, unpublished*.

L. Leydesdorff and H. Etzkowitz (2001). A Triple Helix of University-Industry-Government Relations: 'Mode 2' and the Globalisation of National Systems of Innovation. *In Science under Pressure. The Danish Institute for Studies in Research and Research Policy*. pp 7 – 34.

S. Liebowitz and S. Margolis (1995). Path dependence, lock-in, and history. *Journal of law, economics and organisation*. Vol 11 (1), pp 205 – 226.

P. Llerena and M. Benaim (2010). Les sources de financement de la recherche et de l'innovation en Alsace. *Note N°2 de l'ORRI.*

J. H. Love and S. Roper (2009). Organizing innovation: Complementarities between cross-functional teams. *Technovation*. Vol 29 (3), pp 192 – 203.

R. A. Lowe and C. Gonzalez-Brambila (2007). Faculty entrepreneurs and research productivity. *Journal of Technology Transfer*. Vol 32 (3), pp 173 – 194.

B. Lundvall (1992). National systems of innovation: towards a theory of innovation and interactive learning. *Pinter Pub Ltd, London*.

M. Luwel, H. F. Moed (1998). Publication delays in the science field and their relationship to the ageing of scientific literature. *Scientometrics*. Vol 41 (1-2), pp 29 – 40.

M. H. MacRoberts and B. R. MacRoberts (1982). A re-evaluation of lotka's law of scientific productivity. *Social Studies of Science*. Vol 12 (3), pp 443 – 450.

J. Mairesse and M. Pezzoni (2010). Prima-facie evidence on the importance of collaboration in scientific research: Are italian physicists more productive than french physicists? *Forthcoming (work in progress)*.

R. Merton (1968). The Matthew effect in science: The reward and comunication systems of science are considered. *Science*. Vol 159 (3810), pp 56 – 63.

F. Meyer-Krahmer and H. Wessels (1989). Intersectoral technology flow: Network between producer and user - An empirical analysis of the Federal Republic of Germany. *Jahrbucher fur Nationalokonomie un Statistik.* Vol 206 (6), pp 563 – 582.

P. Mohnen and L.-H. Röller (2005). Complementarities in innovation policy. *European Economic Review*. Vol 49 (6), pp 1431 – 1450.

R. Nelson (2004). The market economy and the scientific commons. *Rsearch policy.* Vol 33 (3), pp 455 – 471.

D. Partha and P. A. David (1994). Toward a new economics of science. *Research Policy*. Vol 23 (5), pp 487 – 521.

R. Radner (1986). The internal economy of large firms. *The Economic Journal*. Vol 96, pp 1 – 22.

R. P. Rumelt (1982). Diversification strategy and profitability. *Strategic Management Journal*, Vol 3 (4), pp 359 – 369.

S. Rynes (1989). Recruitment, job choice and post hire consequences: A call for new research directions. *CAHRS working paper series.* pp 398.

P. Shapira and S. Kuhlmann (eds) (2003). Learning from science and technology policy evaluation: Experiences from the United States and Europe. *Edward Elgar Publishing.* 2003.

S. Slipersaeter (2007). Variables for analysing higher education institutions in Europe. *In A. Bonaccorsi and C. Daraio (eds). Universities and strategic knowledge creation.* pp 433 – 442.

P. E. Stephan, S. Gurmu, A. J. Sumell, and G. Black (2007). Who's patenting in the university? Evidence from the survey of doctorate recipients. *Economics of Innovation & New Technology*. Vol 16 (2), pp 71 – 99.

P. E. Stephan (1996). The economics of science. *Journal of economic literature.* Vol 34 (3), pp 1199.

P. E. Stephan (2008). Science and the university: Challenges for future research. *CESifo Economic Studies*. Vol 54 (2), pp 313 – 324.

S. Stern (2009). Do scientists pay to be scientists? *Management Science*. Vol 50 (6), pp 835 – 853.

E. Strube and M. Resende (2009). Complementarity of Innovation Policies in the Brazilian Industry: An Econometric Study. *CESifo working paper series*.

D. Topkis (1998). Supermodularity and complementarity. *Princeton Univ Press*.

D. B. Turban and D. W. Greening (1997). Corporate social performance and organizational attractiveness to prospective employees. *The Academy of Management Journal*. Vol 40 (3) pp 658 – 672.

D. B. Turban and D. M. Cable (2003). Firm reputation and applicant pool characteristics. *Journal of Organizational Behavior*. Vol 24 (6), pp 733 – 751.

K. Unsworth and S. Parker (2003). Proactivity and innovation: Promoting a new workforce for the new workplace. *In D. Holman, T. Wall, C. Clegg, P. Sparrow, and A. Howard (eds), The New Workplace: A Guide to the Human Impact of Modern Working Practices.*

X. Vives (2007). Supermodularity and Supermodular Games. *Occasional paper*.

A. Wald, K. Franke, and D. Jansen (2007). Governance reforms and scientific production evidence from German astrophysics. *In D. Jansen (ed). New Forms of Governance in Research Organizations*, pages 211 – 232.

F. A. Wolak (1989). Testing inequality constraints in linear econometric models. *Journal of Econometrics*. Vol 41 (2), pp 205 – 235.

Annex 1 Descriptive statistics

Table 26 Laboratory statistics, main variables

Variable	Year	Average	Standard Deviation	Variance	Minimum	Maximum
Total Researchers	1996	17.40964	17.55829	308.2936	2	79
Total Researchers	2000	17.79878	17.98248	323.3695	0	81
Total Researchers	2004	19.25	18.52012	342.9949	0	93
Total Researchers	2008	37.69111	33.38342	1114.453	0	146
CNRS Researchers	1996	20.21475	37.44766	1402.327	1	259.7
CNRS Researchers	2000	19.36403	32.37554	1048.175	1	205.8
CNRS Researchers	2004	19.537	32.35343	1046.744	1	202
CNRS Researchers	2008	46.41212	52.67352	2774.5	1	257
INSERM Researchers	1996	6.088889	9.286977	86.24794	1	46.1
INSERM Researchers	2000	6.968182	9.971552	99.43185	1	56
INSERM Researchers	2004	6.916667	11.05083	122.1208	1	54
INSERM Researchers	2008	17.33462	31.30517	980.0139	1	106
Academic Researchers	1996	11.87222	11.38588	129.6382	1	76
Academic Researchers	2000	12.00779	12.67901	160.7574	1	62
Academic Researchers	2004	15.60282	16.30194	265.7532	1	85
Academic Researchers	2008	31.05	25.1705	633.5543	1	98
Total PhD Students	1996	14.55422	15.9054	252.9818	2	82
Total PhD Students	2000	13.03846	13.8127	190.7907	1	89
Total PhD Students	2004	15.9589	17.91371	320.9011	0	120
Total PhD Students	2008	32.16667	32.86774	1080.289	0	164
Total Personnel	1996	13.37349	31.84602	1014.169	0	209.4
Total Personnel	2000	13.44207	31.51415	993.1415	0	233
Total Personnel	2004	15.45878	33.41349	1116.461	0	237
Total Personnel	2008	37.85667	55.87266	3121.754	0	305
Theses per PhD student	1996	.9327198	.54457	.2965565	.125	2.777778
Theses per PhD student	2000	1.183711	.7488755	.5608145	.1538462	5.5
Theses per PhD student	2004	1.010575	.8038317	.6461454	0	5
Theses per PhD student	2008	.8998018	.476742	.2272829	.25	2.666667
Total Theses	1996	12.85	14.7529	217.6481	1	60
Total Theses	2000	13.64935	13.88654	192.836	1	68
Total Theses	2004	13.13699	15.7174	247.0365	0	105
Total Theses	2008	26.09524	26.59251	707.1614	1	117

Table 27 Laboratory statistics, output, quality, composition and funding

Variable	Year	Average	Standard	Variance	Minimum	Maximum
Total Publications	1996	83.54762	121.8191	14839.89	0	743
Total Publications	2000	81.05952	110.1541	12133.94	0	662
Total Publications	2004	85.5	99.55691	9911.578	0	513
Total Publications	2008	118.44	163.1443	26616.05	0	843
Average Impact Factor	1996	2.921831	2.044256	4.178985	.21	13.66427
Average Impact Factor	2000	3.059156	1.743519	3.039857	.286	8.735464
Average Impact Factor	2004	3.571163	1.813172	3.287594	.7269262	12.01332
Average Impact Factor	2008	4.103522	2.07268	4.296001	.8844873	13.08545
Junior Researchers	1996	9.5	10.18208	103.6747	0	48
Junior Researchers	2000	9.630953	10.09089	101.826	0	52
Junior Researchers	2004	10.64103	10.61728	112.7266	0	48
Junior Researchers	2008	15.48	17.48613	305.7649	0	65
PhD and Postdoctors	1996	22.75	33.95811	1153.154	0	270
PhD and Postdoctors	2000	17.78572	24.61221	605.7607	0	196
PhD and Postdoctors	2004	20.33974	26.0943	680.9122	0	192
PhD and Postdoctors	2008	38.82	45.38214	2059.538	0	224
Senior Researchers	1996	7.285714	8.631089	74.4957	0	41
Senior Researchers	2000	6.976191	7.34679	53.97533	0	35
Senior Researchers	2004	7.25641	8.392074	70.4269	0	44
Senior Researchers	2008	10.52	13.00273	169.071	0	57
Assistant Researchers	1996	13.22619	31.68747	1004.096	0	209.4
Assistant Researchers	2000	13.19345	31.18977	972.8018	0	233
Assistant Researchers	2004	15.1532	32.57909	1061.397	0	237
Assistant Researchers	2008	34.151	54.14628	2931.82	0	305
Total Private Funding	1996					
Total Private Funding	2000					
Total Private Funding	2004	337.518	463.9558	215255	0	2085.157
Total Private Funding	2008	383.7383	501.1622	251163.6	0	1863.112
Total Public Funding	1996					
Total Public Funding	2000					
Total Public Funding	2004	1399.626	1802.671	3249622	7.829	10088.57
Total Public Funding	2008	2528.842	2577.941	6645778	22	9145.376

Christian MARTINEZ DIAZ

ANALYSE DE LA PRODUCTION SCIENTIFIQUE : Rôle de la complémentarité des facteurs

AN ANALYSIS OF SCIENTIFIC PRODUCTION: The complementary role of factors

La compréhension par les économistes des modes d'organisation et de production des connaissances scientifiques reste limitée malgré les développements récents de l'économie de la science. Or la recherche et l'innovation sont devenues des enjeux majeurs pour la compétitivité des économies européennes et leur croissance. Ces travaux de thèse portent une attention particulière non seulement au rôle de l'individu/chercheur au sein du processus de production scientifique mais aussi à celui des laboratoires de recherche en tant que organisations composés des groupes de chercheurs. Le cas du système de recherche académique Français, dans lequel nous trouvons une structure mixte au sein des laboratoires de recherche académique, nous permet également de distinguer le rôle des organismes de recherche publique comme le CNRS ou l'INSERM dans cette organisation et production collective de connaissances. Ces organismes étant consacrés à la recherche de base ou fondamentale, leur apport à la croissance économique est indiscutable. Plus particulièrement, l'objectif ces travaux est de mettre en lumière les caractéristiques propres aux phénomènes de production et de qualité scientifique. Cette thèse s'intéresse aux questions relatives aux caractéristiques de la production et à la qualité de la recherche scientifique, d'une part aux déterminants en matière des ressources humaines et financières de ces phénomènes, et d'autre part aux relations existantes entre différents éléments issus de la décomposition de ces deux types de ressources. Ainsi, ces travaux sont centrés sur des questions relatives à la composition des équipes et aux modes de financement des laboratoires de recherche en tant que déterminants de la production et de la qualité scientifique, à la collaboration entre différents types de chercheurs et aux relations de complémentarité entre différents types de financement de la recherche publique.

The understanding of organization and production modes of scientific knowledge by economists remains limited despite recent developments in the economics of science (Dasgupta and David, 1994, Nelson, 2004, Stephan, 2008). However, research and innovation have become major issues for the competitiveness of European economies and their growth (Meyer-Krahmer, 1989). This thesis draws the public's attention not only towards the role of the individual researcher in the scientific production process but also towards the role of research laboratories and their composition in terms of human resources and types of funding. The case of the French academic research system, in which we find a mixed structure between the academic and institutional laboratories, also allows us to distinguish the role of public research organizations such as CNRS or INSERM in the organization and production of collective knowledge. These organizations, which are devoted to basic or fundamental research, provide an undeniable contribution to economic growth (Stephan, 1996). More specifically, the objective of this work is to highlight the characteristics of the scientific production and quality processes. It focuses on issues related to the characteristics of scientific production and quality namely the determinants of human and financial resources of these phenomena on the one hand and the existing relationships between different elements resulting from the decomposition of these two types of resources on the other hand. In summary, this work focuses on issues related to team and fund composition within research laboratories as determinants of scientific production and quality, on the collaboration between different types of researchers and complementary relationships between different types of funding for public research.