Individual Productivity Differences in Public Research : How important are non-individual determinants? An Econometric Study of French Physicists' publications and citations (1986-1997)*

Laure Turner[a] and Jacques Mairesse[b]

February 2005

We express grateful thanks to Michèle Crance and Serge Bauin of the Unité des Indicateurs de la Politique Scientifique (UNIPS, CNRS), for allowing us access to the data, for their help in compiling the data base and their advices. We thank as well Bronwyn Hall, Richard Blundell, Paul David, Emmanuel Duguet and Frank Windmeijer for their profitable comments and suggestions. We thank Emmanuel Duguet, Bronwyn Hall and Frank Windmeijer for allowing us to use their programs, respectively for the Zero-Inflated Poisson, the Poisson and log-linear estimation on panel data and the non-linear GMM estimation of exponential models. This paper also benefited from discussions at the 10th International Conference on Panel Data in Berlin 2002, at the SPRU Conference "Rethinking Science Policy" and at the Paris Conference in honor of Zvi Griliches 2003. Any remaining error is ours.

[a] ENSAE et CREST-LEI, 27 rue des Saint-Pères, 75007 Paris, e-mail: laure.turner@ensae.fr
[b] CREST-INSEE, 15 Bd Gabriel Péri, 92242 Malakoff, e-mail: mairesse@ensae.fr.

Introduction

In academic science, an empirical regularity has been evidenced by numerous studies : the strong inequality among researchers in terms of productivity and the persistence of the productivity hierarchies over the life cycle for a given cohort of scientists. Since Lotka's seminal article (Lotka, 1926) it is observed that in every scientific fields studied, a prolific minority of scientists contribute to the majority of the publications of their field, so that the distribution of publication counts is very left-skewed. In the data collected on French physicists and presented bellow, 10% of the of the scientists contributes to 30% of the total number of publications whatever the period - 1980-1985, 1986-1991, and 1992-1997 (figure 1). Moreover, 66% of the most productive researchers as well as 67% of the less productive researchers remain such over 1986-1997, which underlies a stability of the relative positions of the researchers in the distribution of publication counts over time (figure 2).

Inequality in outcomes in science is an issue for science policy and for the allocation of public resources into research. The fact that up to 50% of the contributions to a research field is made by a minority of scientists has served to question the efficiency of knowledge production in the public sector. For this matter, understanding the processes underlying the productivity distribution in science is of primary interest. Does the distribution of productivities mimic an unequal distribution of talents among researchers? Are there some cumulative phenomena at play, such that initial successes (failure) translate into permanent high (low) productivity? Is it to do with research incentives or to the scientists unobservable and unmonitorable ability for creative work?

In this paper, we explore the issue of the determinants of individual productivity differences in science, by looking at the relative role of three series of factors – individual characteristics, environment and incentives- in explaining individual productivity inequalities. We have built a longitudinal database for this study that concern French physicists and covers the recent period of 1980-1997.

We are interested in the hypothesis made in the economics of science literature of a link between incentives in research and individual scientific productivity. Incentives in research are specific in that they are to a large extent non monetary and reputation based. The researchers are rewarded for their productivity by the reputation gained among their peers, by prizes, nominations in prestigious institutions, the possibility to work in stimulating environments with greater resources, and so on. Zuckerman (1992) noticed the existence of about 3000 scientific prizes in North America in the beginning of the 1990s. The underlying attribution mechanism is the "priority rule", which selects the researchers who are first to discover and to publish (Merton 1957). Moreover, a process of "cumulative advantage" could be at work amplifying the impact of the non-monetary incentives on publication. In this process, an initial success in publication entails increasing productivity and reputation. The interpretation is that successful researchers have access to grants, time, stimulating laboratories and teams and so on, which help them to increase or at least maintain their publishing activity and their reputation. On the contrary, a scientist who has experimented a bad start in his research activity might be obliged at one point to quit research because of the accumulated obstacles (David 1994). In this first approach, the distribution of scientific outcomes stems from those two selective and cumulative mechanisms associated with the incentives in research. Empirically assessing the presence of cumulative advantages in research is beyond the scope of this study. Rather we look at promotions and laboratory affiliations as sorts of "rewards" of past productivity likely to have a lasting impact on future publication. Promotion and membership of a dynamic laboratory that is central in research collaboration are expected to be part of the process of cumulative advantage and at least, within the scope of our study, to stimulate researchers' individual productivity.

Laboratory variables are also used to capture environment effects on individual productivity. Long (1978) and Allison and Long (1990) underline the role of prestigious academic affiliation in encouraging individual scientific productivity. Carayol and Matt (2004) find a correlation between individual productivity and the labor force organization of the labs (shares of permanent, teaching and doc post-doc researchers) using data on about 80 laboratories belonging to a large French University. Mairesse and Turner (2002) show that the geographic proximity, size and productivity of the laboratories positively impact co-publications in networks at a laboratory scale. In this paper we look at the effect of the productivity of the colleagues on individual performance, as well as the effect of the size and of the share international collaborations of the labs.

We should also find a strong influence of individual observable variables on research productivity. The relation between age and publication has been often analysed. In the framework of the life cycle models, Diamond (1984) and Levin and Stephan (1991) modelled the quadratic relation between age and productivity assessed by publication counts according to which productivity was decreasing with age toward the end of the career. One issue at stake at the time was to assess whether the aging of the scientific community, as it was occurring in the United States, was going to impact negatively on national scientific output. Many other studies have also focused on gender discrimination in science, which has always been an important topic to the scientific community (Stephan, 1998). It appears also that education in a prestigious university or through a selective doctoral program has proved to have a positive impact on productivity (Crane, 1965, Long, Alison and McGinnis, 1979). Finally, we must take into account unobservable individual specific effects, interpretable as the «sacred spark» (Cole et Cole, 1973), the intrinsic motivation for research that determines scientists productivity independently of any incentives. Levin and Stephan (1991) speak of a taste for "puzzle-solving". It could also be seen as personal motivation for research or ability for creative work.

Our approach is ambitious in the sense that it aims at comparing simultaneously the relative impact of all those series of factors. It should therefore be considered as a first analysis that we need to develop further by focusing on several points. In particular, the way publication influences promotion and access to "good" laboratories should be taken into account, which is another work on its own. In the framework of this paper we have estimated an extended version of our model considering promotion and laboratory variables as endogenous, but our results were dissatisfactory and are not reproduced here. This limit will be discussed in section I.

Nevertheless our work sheds light on some interesting aspects of scientific productivity. We find a quadratic relation between age and individual productivity according to which productivity in terms of articles per researcher per year diminishes with age after 52 years old. Tenure significantly contributes to explain the productivity decline with age after this threshold, which is related to an important negative effect of long tenure in the "director of research" status and to a possible discouragement effect of non promoted researchers. Gender is also a major individual determinant of productivity as well as, to a lesser extent, high standard pre-doc formation ("Grande Ecole") which effect also seems to rely on the personal networks it provides. Among the laboratory characteristics, the international openness of the laboratory through collaboration and the laboratory's productivity influence individual performance with the same order of magnitude than the previous variables. Moreover, three elements are noticeable: 1) there is evidence of a positive peer effect on productivity; 2) the size of the laboratory has a small effect on individual productivity even though "talented" researchers seem more likely to be affiliated to larger labs; and 3) the accessibility of the technologies for experiments - captured by the Grenoble region dummy - has a positive impact on productivity.

The paper is organised as follow. The next section describes the data, the model specification and the estimation methodology (I). Section II describes the results when the productivity is assessed by three measures: the mean number of articles par researcher and per year, the average impact factor and the mean number of citations to the articles. Section III concludes.

I. Specification and Estimation Methodology

We have build a panel data base to study scientific productivity that document almost 20 years of publication for approximately 500 scientists. Productivity is measured both by publication counts and by two measures aimed at assessing the quality of the publications: citation counts and the impact factors of the journals. Another originality of our data set is that it includes information on the characteristics of the scientists (age, gender, pre-doc formation) as well as information on the timing of their promotions. We also know their laboratory affiliations over the period, which allows to study laboratory effects. These combined features makes the data set relatively rich and original for the study scientific productivity. To our knowledge, few sets of this kind exist. A data base built by C. Gonzalez-Brambila concerns Mexican researchers who have been selected to be part of the Mexican National System of Researchers (SNI) from 1991 to 2002, with information about the age, gender, year of PhD and field of the scientists. Levin and Stephan (1991) use a panel data base on publication and citation counts for American scientists over 1973-1979 to revisit some results on the age/productivity relation obtained on cross-sectional data. The next paragraph describes our data, and paragraph B presents the different variables that are used in the productivity model. Finally paragraph C develops our estimation methodology.

A. *The data*

The source of our publications and citations data is the *Science Citation Index (SCI)* which is produced by the *Institute for Scientific Information (ISI).* The *SCI* covers all the scientific domains and the articles of approximately 3200 most cited journals. The quality of the data is excellent, which make the *SCI* the international reference for bibliometric work.

We drew from the *SCI* the publications of 497 French physicists over the period 1986-1997. For 352 of them we have been able to collect as well all the citations received by their articles[1]. We considered the citations received within two years, in order to keep the maximum number of years of data in the citations set[2]. The period covered in the citation set is therefore 1986-1994. Citations are traditionally used in bibliometric studies to weight articles in order to account for their impact. In this

[1] For the citations only, the sample is not made of 497 scientists but is a sub-sample of 352 researchers who were born between 1936 and 1955 instead of 1936 and 1960. This is due to the timing of the data collection.
[2] On average, an article receives approximately 40% of its citations within two years according to our data.

paper, the average number of citations (received within two years) per article per scientist and per year is taken as a qualitative measure of productivity. The second qualitative measure of productivity considered in our study is the average impact factor per article published, per year and per researcher. The impact factor of a journal gives information on the journal's reputation and visibility. This measure has been calculated by *ISI* for the CNRS as the mean number of citations obtained by the articles of the journal within two years after publication[3]. An impact factor calculated over five years is sometimes used as well, but the productivity measure based on a two years impact factor has the advantage of being comparable to our productivity measure in terms of citations. Interestingly, the correlation between those two qualitative productivity measures is 0.37, which is less important that what could have been expected considering that they are often viewed as substitutes when assessing the quality of scientific publications. The correlation of the annual number of publications per scientist with the average impact factor of the journals of publication per scientist and per year is 0.36, and it is 0.26 with the average number of citations per scientist and per year.

The sample consists of the scientists working at the CNRS[4], the French public institution for research, in the field of condensed matter[5]. The field of condensed matter was chosen for two reasons. First, its characteristics are suited to our study: its research is classified as pure basic science; journals with a sound reputation are clearly identifiable; the size of the field covered is clearly defined; and there is very little mobility among researchers from public research to teaching or to private research. Second, condensed matter is a fast-growing field, honoured by the Nobel Prize for Physics awarded to Pierre-Gilles de Gennes in 1991, and which currently accounts for close to half of all French academic physics[6].

Since it is desirable to have a stable number of scientists over the period under study for the econometric estimations, we considered only the researchers whose first publication or entry at the

[3] More precisely, for a given journal, the impact factor is the ratio of the number of citations received during years T and T-1 by the articles published during years T-1 and T-2 over the number of considered articles. An article is qualified by the impact factor of the journal into which it was published.
[4] The Centre National de la Recherche Scientifique (CNRS) is a public organization for research, affiliated to the Ministry responsible for Research. With 25,000 employees (11,000 researchers and 14,000 engineers, technicians and administrative staff), a budget of 2.5 billion euros in 2001, and laboratories throughout the country, the CNRS covers all fields of knowledge. University researchers are often affiliated to CNRS labs therefore called “mixed units”.
[5] The group of 497 physicists studied here represents almost all CNRS researchers in this field (654 in 1996). They were selected according to their year of birth: 1936-1960.
[6] Condensed matter includes all states of matter, on various scales (atom, molecules, colloids, particles or cells), between liquids and solids, in which molecules are relatively close. Its study is based on a heritage of traditions, both experimental (crystallography, diffusion of neutrons and electrons, magnetic resonance imagery, microscopy, etc.) and theoretical (static physics). It is also prompted to develop more and more relations with industry around materials used in electronics, granulars, plastics, food or cosmetic gels, etc.

CNRS dates back before 1986[7]. Our final sample is balanced and concerns 465 scientists over 1986-1997[8]. The next section describes our explanatory variables.

B. Explaining Scientific Productivity

We study the determinants of researchers' productivity measured along three dimensions - in terms of the annual number of publications per scientist, in terms of the average impact factor of the journals of publication per scientist and per year, and in terms of the average number of citations per article per scientist and per year - which correspond to three sets of regressions.

Table I.1 indicates the main statistics for these variables as well as for the explanatory variables used in the models. The 465 physicists in our sample published approximately 8000 articles over the period 1986-1997, which corresponds to a mean number of 2.7 papers per researcher and per year, with a standard error of 3. The annual number of articles publication varies greatly among the scientists, between 0 and 62, the maximum over the period. The mean proportion of researchers with no publication in a year is 27%. The mean number of authors per article is 3.2, and the mean number of pages is 5.5. The scientists get published in journals whose articles receive 2.7 citations on average over two years. The quality of the journals of publications are different across the researchers, ranging between almost 0 and 21.5 citations in two years. Approximately 32 000 citations (within two years) were received by the publications of the scientists studied, which amounts to 3.5 citations per researcher and per year on average over the period, with a standard error of 6.

We study the relative impact of three series of determinants on scientific productivity that are presented in paragraphs 1, 2 and 3 below: individual factors - age, gender, education -, factors related to the incentive scheme of academic research - the career trajectory or experience, and environment factors - the size and activity of the laboratories in which the scientists work and a mobility indicator. We also take into account the unobserved individual specific effects. A major problem arise when the aim is to measure the relative impact of variables such as age and experience, because those variables are strongly correlated. Moreover, in our specification, we need to take into account an observed fact,

[7] The researchers entered the CNRS at different dates, between 1960 and 1997. Few researchers enter after 1990 (17), so they have been eliminated from the data. Among the remaining 480 scientists, 433 entered the CNRS before 1986, so we expected them to publish during the whole period of observation 1986-1997. 47 scientists entered between 1986-1990. Either they started to publish after their entry, in which case they are not considered as being in position to publish during the whole period of 1986-1997. Or they have started to publish before 1986, so we kept them in the sample. Between the year of their first article and their entry at the CNRS, they were assigned the status that they reached at their entry.

[8] As mentioned, for the citations only, the sample is reduced to 352 researchers over 1986-1994.

that is the exogenous increase of publication with time, and add time dummies in our model. This identification issue is considered in paragraph 4.

1. Individual variables

Age

The age dispersion among the scientists under study is important. On average, they are 44.6 years old but the standard error amounts to 8.0. Economic studies of the age/publication relation use the framework of life cycle models (Diamond, 1984, Levin and Stephan, 1991). These models enlighten the consequences of the end of the career on individual productivity and on the allocation of research efforts over time. In Levin and Stephan (1991), the scientists allocate time in order to maximize their utility function over their career. At each date, the utility depends on the future financial rewards associated with teaching and consulting activities and on the current research output seen as a proxy for the "puzzle-solving reward". One implication of the model is that research activity declines over the life cycle. Interestingly for our study, this proposition was verified on publication panel data drawn from the *SCI* and the *Survey of Doctorate Recipients* that concern six sub-fields of physics and earth science including solid state and condensed matter physics over the period 1973-1979[9]. We are able to look at this effect as well and to compare our estimates with the ones of Levin and Stephan (1991). The age variable used in the model is described in paragraph 4.

Gender

Men represent 82% of our sample. A main concern in the issue of the gender influence on publication is whether the rewards in science are gender biased. Several studies have concluded that women scientists publish less than men and that they earn less as well (see Stephan 1998). Zuckerman, Cole and Bruer (1991) show that the process of cumulative advantages might be a reason of the persistent position of women in the "outer circle of science" because it amplifies an initial situation where women published less than men. But in the empirical studies, the relation between gender and outcome is often biased in the sense that the estimations rely on cross-sectional data that can not allow to account for unmeasurable individual effects reflecting personal motivation, talent or any omitted individual variable explaining productivity. Moreover, the samples used are non random but consists of the successful scientists, which introduce a selection bias in the results. Consequently, it is interesting to test the gender/publication relation on panel data, because we can control for the unobservable specific effects and take into account the evolution that led to the observed situation. In this framework, Stephan (1998) finds that gender is not a significant determinant of salary changes in

[9] This effect was not properly identified on cross-sectional studies, since they did not control for cohort effects.

US academe during the 1970's. In our model gender is introduced as a dummy variable that equals one if the scientist is a woman.

"Grande Ecole" dummy

We introduced as a dummy variable equal to one when the researchers studied in a "Grande Ecole" in addition to graduate from their PhD[10]. 16% of our sample did so. Among them, over 60% belonged to the Ecole Normale Supérieure, 6% to the Ecole Polytechnique, 10% to the Institut Supérieur d'Electronique du Nord, 6% to the Ecole Supérieure d'Electricité de Paris, etc. We expect this dummy variable to play a role on individual productivity since different studies have shown the importance of pre-doctoral formation in explaining productivity differences in research (see for instance Long, Allison and McGinnis, 1979). The intuition is that the knowledge, values, and scientific performance criteria learned during this period have a lasting positive impact on their work.

Individual specific effects

Apart from the explanatory variables, we also introduce in the model random effects specific to the individuals, in order to take into account the unobserved individual heterogeneity. The choice of our estimation method is determined by the existence of a correlation between the individual effects and the explanatory variables.

2. Incentives variables

Career paths

The researchers are distributed according to the evolution of their career, so that we can study the link between publication and promotion and more generally account for an incentive mechanism of the scientific institution. A researcher with a typical career profile enters the CNRS as "Chargé de Recherche" (CR), is then promoted research director of class 2, "Directeur de Recherche de 2ème classe" (DR2), and finally research director of class 1, "Directeur de Recherche de 1ère classe" (DR1). Yet, many researchers in our sample are never promoted and remain in the same status during the observed period (respectively for the status CR : 46.7% of the sample, DR2 : 10.4%, DR1 : 3%). Almost 30% of the sample get to be promoted DR2. The most difficult promotion to obtain is DR1: only 10% of the sample succeed to be promoted DR1.

---

[10] In the French educational system, after they graduate from high school, the students can either go to the University, which does not require any level nor grade achievement in high school, or they can apply to a preparatory class where they will be taught during two years the knowledge required to compete for the hard admission into a "Grande Ecole". Every student of the Grandes Ecoles therefore succeeded in two selection processes: selection on the basis of their grades in high school, and exams to enter the Ecoles.

The interactions between the career inside the scientific institution and publication behaviour are numerous. A descriptive study (Turner, 2003) on our sample established the existence of a positive link between status and publication at each date, whereas the relation between current position and past publication is more complex. We would expect it to be positive as well, since the CNRS incentive scheme is such that promotion rewards publication.

Conversely, does promotion give incentives to publish more, for instance by offering a better access to the resources needed for research? The descriptive study mentioned shows that on average the promoted researchers remain as productive as before their promotion. But the mean number of publications differs before and after the promotion according to the status reached, the age, and the period. For instance, the oldest scientists publish less after any promotion, especially after a DR1 promotion obtained during the last sub-period whereas younger researchers publish more after a DR2 promotion.

In our econometric model, we introduce a certain number of variables related to the career trajectory or promotion profile as explanatory variables. The variables are chosen to capture observed changes in the publication behaviour as the number of years since promotion increases. We observe that the productivity of the scientists who are never promoted and who remain CR first increases with time and then decreases after a certain number of years spent in the status, as a sign of discouragement. We also observe that the DR2 productivity seems to remain on a constant trend, whatever the promotion perspectives or the total experience at the CNRS. Finally, it appears that the DR1 productivity is decreasing with tenure in the status. Consequently, we retain as career variables the tenure in each status and dummies for the status DR2 and DR1[11]. The tenure variables are described more precisely in paragraph 3.

3. Environmental variables

A study of the collaboration among the scientists presented in Mairesse and Turner (2002) has underlined the impact of some laboratory characteristics – the size, the productivity of its members, the quality of its publications, its international collaborations, its thematic specialisation – on the intensity of co-publication at the laboratory level. We want to assess the impact of the same laboratory variables on the individuals productivity. This “laboratory effect” is an issue when evaluating the recent policies that stimulate the creation of large research structures like technopoles. We have in mind that the membership of a dynamic laboratory stimulates researchers' individual productivity and

[11] Several trials have later confirmed that any variable controlling for the total number of years the scientist has been working at the CNRS or the perspective of the career ending had less significant effect on productivity. The major effects are the effects of tenure per status.

may be part of a process of cumulative advantages in which well-known scientists enhance their productivity and recognition by working in this type of laboratory.

The variables are calculated for the laboratories where the researchers were working in 1997. They are therefore time invariant and some are aggregates over the whole period of 12 years. More precisely, the laboratory variables are the following:

- Size: the size of the laboratory is the total number of researchers in the laboratory in 1997 including University scientists affiliated to the lab. The size variable is also centred and squared in order to measure quadratic effects.
- Productivity of the laboratory : it is calculated for every researcher individually over the whole period 1986-1997, by subtracting its personal contribution to the production of his colleagues who we have in our database and who are working in the same laboratory in 1997. Subsequently, it proxies the productivity of the researcher's colleagues or environment. We take the logarithm of this variable.
- Quality at the laboratory level: it is also calculated for every researcher individually over 1986-1997 by subtracting its personal contribution to the average impact factor of his colleagues who we have in our database and who are working in the same laboratory in 1997. It reflects the quality of the researcher's environment. Again we take the logarithm of this variable.
- International openness of the laboratory: it is the proportion of articles at the laboratory level and over the whole period co-published with at least one foreign co-author (see Mairesse and Turner, 2002).
- Region dummies: a dummy for the Grenoble region and the Paris region are introduced to qualify the laboratories in 1997 because those two regions account for a major part of the total number of physicists and publications over the period. The regions are defined as Grenoble (resp. Paris) plus the set of towns geographically close to Grenoble (resp. Paris) - less than 100 km - in which CNRS laboratories are located in 1997.

Of course, the mobility of the researchers is of central interest for the evaluation of a "laboratory effect", and taking the 1997 laboratory of the researcher seems to give an incomplete information. This is not so because of two main reasons. First, the actual mobility is very low: over the period of 18 years between 1980 and 1997, 55% of the researchers never changed laboratories, 33% changed only once, 11% changed twice, and 2% changed three times. Secondly, the results remain when we consider the sub-sample of the 55% of scientists who do not change lab. We add a dummy when the number of changes equals more than one. It is to capture an anticipated effect of mobility according to which the researchers who moved more than once are affiliated to more productive laboratories on average.

The fact that the productivity, quality and international openness of the laboratory are calculated over 12 years instead of being time-variant is also not a main issue because it is conceivable that a long period is needed to assess the genuine or "steady-state" productivity and quality of the researchers who contributed during a certain period to the reputation of the laboratory. We have done the same calculations over a six years period but it did not change the results.

4 Time, Age and Tenure effects: the well-known identification problem

First statistics on our data suggest that publication increases toward the end of the period independently of the age of the researchers. We introduce years dummies in the regressions in order to capture this effect and account for any changes in the work environment or in the state of the art in condensed matter physics. But the estimation method that we use (section C) require to write the variables in deviation from their mean, which rise a major identification problem since the transformed age and time variables are collinear. Therefore, we decided to assess the age effect by four age groups instead of continuous age and age squared variables (see table I.1). It breaks the link between age and time so that it is possible to estimate both effects simultaneously.

This issue relates to the well-known identification problem when simultaneously estimating period-age and cohorts effects. Hall, Mairesse and Turner (2005) specifically concentrate on this issue with similar data in order to explore a satisfying solution.

A last question remains. Including both the tenure variables and the time periods in the model rises another identification problem. Because a significant number of researchers are never promoted, time and tenure in status also show some collinearity, especially when considering the variables in deviation from the means. The solution proposed is to form three groups of tenures for each status. Again, it has the advantage to break the link between tenure and time and to allow the estimation of time-varying tenure effects. For instance for the DR1 status, we look at the following three groups of tenure: 0-1 years as DR1, 2-5 years as DR1 and more than 5 years as DR1 (see table I.1).

The next section presents the methodology used to estimate the model.

C. *Methodology*

To estimate the first equation in which the dependant variable is the number of articles, a count variable, we need to estimate a Poisson model of productivity. We run the estimations under the main hypothesis that the explanatory variables are strictly exogenous as respect to the errors.

The model is the following, with i={1,…, N}, N=465, and t={1,…,T}, T=12:

$$E(y_{it} | X_{it}, Z_i) = exp(m + Z_i g + X_{it} b + a_i) \quad \text{with } y_{it} \rightarrow Poisson \qquad (1)$$

The variables in Z are stable across time but not across individuals and the variables in X vary in both dimensions. The random individual effects are α_i. We assume that the errors are not serially correlated. But according to the Hausman tests, we assume that the individual effects are correlated with the explanatory variables. In order to be able to estimate the coefficients of the time-invariant variables of the model, we assume that all the correlation with the individual effects is due to the time-varying variables in X, and that the time-invariant variables in Z are not correlated with the individual effects[12]. Naming u_{it} the error term, we have:

$EX_{it}'\alpha_i \neq 0$, $EZ_i'\alpha_i = 0$,

$EX_{it}'u_{it} = EZ_i'u_{it} = 0$,

$E\alpha_i = Eu_{it} = 0$,

$E\alpha_i u_{it} = 0$, (H1)

$E\alpha_i \alpha_j = \sigma_\alpha$ if i = j, and $E\alpha_i \alpha_j = 0$ otherwise,

$Eu_{it}u_{is} = \sigma_u$ if i = j and t=s, and $Eu_{it}u_{is} = 0$ otherwise.

To solve the problem of the correlated unobserved individual effects to the explanatory variables, we treat them as fixed effects in our estimation .

A two step estimation is used (TS in what follows). We estimate β in (1) by the Conditional Maximum Likelihood Estimation (CMLE) used by Hausman, Hall and Griliches (1984). In a second step, to estimate the coefficients of Z, we replace β by its CMLE estimate and estimate equation (2) using the non linear least squares method :

[12] By doing so, we do not estimate the raw effect of the time invariant variables. For instance, the estimation does not separate the effect of gender from some unmeasured effects that might exist and be correlated to

$$y_{it} / \exp(X_{ti}\hat{b}) = \exp(m + g\ Z_i) + e_{it} \tag{2}$$

We obtain consistent estimates of γ and μ. We run the Two Step estimation (TS) as well as the level estimation - the basic Poisson model (named TOTAL in what follows) – in order to assess the size of the unobservables effect. Only the TS regressions take the individual effects into account. The results are in *table II.1*.

When the dependent variable is the average quality of the papers per researchers and per year – measured by the impact factor or by the number of citations, the Poisson model is replaced by the log linear model, since the dependant variables are continuous:

$$\log\ (\mathrm{y_{it}}) = dum(y_{it}=0) + m + Z_i g + X_{it} b + a_i + u_{it} \tag{3}$$

The same Two Step method is used to estimate the time-invariant variables. We estimate β by the WITHIN estimator. In a second step, to estimate the coefficients of Z, we estimate equation (4) using the linear least squares method :

$$\log\ (\mathrm{y_{it}}) - X_{it}\hat{b} = dum(y_{it}=0) + m + Z_i g + a_i + u_{it} \tag{3}$$

The results are in *table II.2* for the impact factor and in *table II.3* for the citations.

The next section presents the results of the three estimations successively, isolating the determinants of publication, impact factor and citation respectively.

II. Results

A. The determinants of publication

This section describes successively the impact of individual, promotion and laboratory variables on individual productivity assessed by the mean number of articles per year and per researcher.

gender: the number of children or maternity leaves, marital status, etc. The woman variable embodies the fact of being a woman plus all the unmeasured correlated facts absent in the regression.

When looking at the **age** groups, the estimation suggests a quadratic relation between the age of the scientists and the average number of their publications per year. According to the estimation, the researchers productivity increases between the first and the third age group, that is before 50, and then declines after 51 years. More precisely, the researchers aged 39 to 45 years publish on average 0.26 paper more per year than the youngest researchers aged 26 to 38, and the scientists aged 46 to 50 publish 0.36 paper more than the youngest group. The oldest researchers, aged 51 to 61, publish only 0.13 paper more than the youngest researchers.

To refine our idea on the effect of age on productivity, we run an annex regression in which the continuous variables age and age squared replace the age cohorts. As explained previously, the time dummies can no more be correctly estimated, but we are simply interested in the age estimates. We check that the age estimates do not vary too much when we exclude the time dummies from the equation. The TS estimates of age and age squared are respectively 0.022 and –0.0016 when the time dummies are included in the regression, and 0.025 and –0.0014 when the time dummies are excluded. Consequently, the quadratic relation between age and productivity is confirmed. According to the annex TS regression in which time dummies are included, the conditional effect of age is such that the researchers are more productive every year until they turn 52 but at a diminishing rate: they publish 0.9 paper per year on average at 30 years, 2.3 papers at 40, 2.9 papers at 52 and 2.5 papers at 60. The curves illustrating the age/publication relation according to the TS estimation and conditionally on the other variables is represented on *graph II.1*. It includes the mean point of 2.7 papers at 44.6 years old.

We compare our TS results to the ones of Levin and Stephan (1991) obtained for 182 scientists in solid state and condensed matter over 1973-1979 using a Tobit model with correlated fixed effects and time dummies (model B in the paper). The mean number of papers is 3.8 over *two* years, which is slightly smaller than the average productivity in our sample but is still comparable. The quadratic relation between age and publication is confirmed with a coefficient on age of 2.41 and on age^2 of -0.027 (publications are counted over two years). Yet, the life cycle effect is stronger in their model, in the sense that their results imply a relation more quadratic then ours. According to their model, the solid state and condensed matter physicists are productive between 33 and 57 years old, publishing 0.71 paper *per year* at 35 years old, a peak of 2 papers per year at 45, and 0.72 paper at 55. The differences in the findings could in part be explained by the fact that the specification of our model takes into account the count nature of the data, the great proportion of zeros and the career of the scientists.

The age effect is complemented by a **time effect** according to which the productivity increases with time for all the researchers. The scientists publish 0.9 paper more on average in 1991 than in

1986, and 1.6 papers more in 1996 than in 1986. This tends to suggest that the scientists progressively experiment a wider and faster access to publication related to the increasing numbers of existing journals.

We find a strong influence of the other individual variables on productivity.

The **gender** effect is important. All other variables being equal, woman publishes almost 0.9 paper less than a man on average per year according to our estimation. This results needs to be developed in a specific study on gender. If the estimates suggest that men are more productive than women, they do not tell neither about the reasons of this phenomena nor about the true abilities of women scientists. Several sociological reasons could explain this feature. And our estimation method does not allow us to distinguish the pure effect of being a women from all the related unmeasured "sociological" facts (number of children, marital status, etc.) which impact as well on the value of the estimated coefficient (see note 12).

The result on **pre-doc formation** is puzzling. The scientists who have been educated in a Grande Ecole publish 0.7 paper more than the others per year on average, according to the TS regression. But this figure amounts to 0.3 in the TOTAL estimation. The gap between 0.7 and 0.3 means the following: all other observable effects being equal, and all unmeasured effects being equal as well, the scientists who have been to a "Grande Ecole" publish 0.7 articles more each year than the other scientists; but when we take into account the fact that unmeasured effects are different for the scientists who have been to a "Grande Ecole" and the others, the gap is reduced to 0.3 papers. The correlation between unmeasured effects and "Grande Ecole" is negative: taking into account this correlation implies that the scientists who have been to a "Grande Ecole" publish only 0.3 articles more each year than the other scientists. To interpret this result we need to interpret what the unmeasured effects could be: it could well be a "prestige" or a personal network effect. As a matter of fact, alumni's networks are very developed and active in France and they play a role in collaboration, promotion and mobility. The hypothesis could then be that relying on an active personal network or on "prestige" induce lesser "effort" to publish than without those supports. Of course, this hypothesis needs further investigation.

The TOTAL and TS estimations give a different picture of the **career** influence on productivity. In the TOTAL estimation, DR1 publish 1.5 papers more on average per year than CR researchers and a DR2 scientist 0.5 paper more. Whereas in the TS estimation, the DR2 coefficient is not statistically significant (-0.1), and the DR1 coefficient is negative, the DR1 publishing on average 0.8 paper less than a CR. Reaching the status DR1 has an negative impact on productivity according to the TS regression. If we believe in our model, this result suggests that all other effects being equal, if "talent" was equally distributed a DR would be less likely to publish than a CR (TS estimates); but if we take into account the positive correlation of “talent” with the status variable, a DR publishes more on average per year than a CR scientist (TOTAL estimates).

The **tenure** estimates give more details on career influence on productivity. According to the TS regression, a scientist who has been **DR1** for more than 5 years publishes 1.3 papers less on average than a newly promoted DR1 researcher. It means that the impact on productivity of being DR1 diminishes as the number of years spent in the status increases. Consequently, the incentives to publish appears to be lower with time for those researchers who have reached the higher status. The increase in their amount of administrative tasks must here be mentioned. Again, this result holds all other variables being equal in particular talent among the DR1. The higher TOTAL estimates suggest that “talent” is positively correlated with tenure when looking at the DR1, which is intuitive because the ones who have been DR1 for a longer time, all else being equal, have been promoted sooner. Taking this correlation into account reduces the effect of tenure almost twice: according to the TOTAL estimates, a scientist who has been DR1 for more than 5 years publishes 0.8 papers less on average than a newly promoted DR1 researcher.

Similar effects are noticed for the **DR2**. A scientist who has been in the status DR2 for 4 to 8 years publishes 0.3 paper less on average than a newly promoted DR2, and a DR2 of the third tenure group (>8 years as DR2) publishes 0.5 paper less on average than a newly promoted DR2 (TS estimates). Again, it appears that the incentives to publish become lower as the time spent in the status increases. But if we take into account the positive correlation of “talent” with tenure within the DR2 status, the effect of tenure as DR2 on productivity is greatly reduced and becomes statistically non significant.

Finally, according to the TS estimation, tenure has no effect on the productivity of the **CR** researchers, but it has a negative impact according to the TOTAL estimation. This suggest that the unobserved individual effects are negatively correlated to the tenure in the CR status. A discouragement effect could be a play according to which the CR with very low remaining perspectives of promotion have a low productivity.

Concerning the **laboratory effects**, the results are the following. There is a positive peer effect, that is an impact of the laboratory’s productivity (or colleagues’ productivity) on the number of articles published by the individuals. We find that, in the case of a 10% rise in the productivity at the laboratory level, a scientist would on average publish 0.27 paper more than the number he would publish otherwise (+10% on average). Interestingly, collaboration with foreign laboratories also has a strong positive impact on individual productivity. In the case of a 10% increase in the proportion of papers co-published with foreign scientists at the laboratory level, a member scientist of the lab would on average publish 0.8 articles more on average par year (+30% on average). This result suggest that forming centers of excellence in public research induce a better productivity of the member scientists.

The **size** of the lab itself does not matter comparatively to this peer effect. In the case of a 10% rise in the size of the lab, a scientist would on average publish 0.09 paper less than the number he would publish otherwise (-3% approximately on average). Nevertheless, our estimates suggests that "talented" researchers are more likely affiliated to larger laboratories. When we look at the dummy for small labs (DUMEF13), the TOTAL and TS estimates show a negative correlation between unobserved individual effects and this variable.

The result on influence of the **quality of the laboratory** (or colleague's productivity in terms of quality) is questioning. The peer effect in terms of quality is almost inexistent and not statistically significant in the TS regression. But the TOTAL estimate is strong, negative and significant (-0.254), that is in the case of a 10% rise in the quality at the laboratory level, a scientist would on average publish 0.25 paper *less* than the number he would publish otherwise (-10% approximately on average). One can wonder if there is a sort of substitution effect between quantity and quality, in the specific sense that if the lab stresses the importance of quality, a member of the lab is more likely to publish less but papers of greater impact. As we will see in the two next sections, individual productivity in terms of quality is positively influenced by quality at the laboratory level.

B. The determinants of the average quality of the journals of publication

The results are presented in table II.2. The most significant impact here is the one of the quality of the laboratory. In particular, the impact of individual characteristics and promotion variables is much less important then previously when the dependant variable was the mean number of articles.

The average quality of the journals of publication is negatively influenced by **age** and no obvious quadratic relation emerges from the estimates of the age groups. The oldest researchers aged 51 to 61 publish in journals that receive on average 0.3 citation less than journals of the youngest researchers. As previously, to refine our idea on the effect of age on impact factor, we did an annex regression in which the continuous variables age and age squared replaced the age groups. According to the annex TS regression in which time dummies are included, their is a negligible increase in the impact factor between 26 and 41, from 2.50 to 2.67. After 41, the average impact factor declines slightly and is equal to 2.35 at 61. The curves illustrating the age/publication relation according to the TS estimation and conditionally on the other variables is represented on *graph II.4.*

The **gender and "Grande Ecole"** variables are statistically significant and influence similarly the quality of the papers published, but with a much smaller order of magnitude than the one found in the previous regression on the number of articles. The women publish in journals that receive on average 0.10 citations less over two years than the journals into which men publish. Interestingly, the unmeasured individual effects and possibly what could be interpreted as "personal motivation for accessing recognized journals" are positively correlated to the gender variable "woman".

The **status and tenure** variables have no statistically significant impact on the average impact factor of the journals. Nevertheless, the TOTAL estimation shows as previously that "talent" is positively related both to status (whereby DR1 publishes 0.36 papers more on average than CR and DR2 0.15 on average) and to tenure among DR2.

The dominant effect here is the one of the **quality of the laboratory**. A 10% increase in the quality of the laboratory increases of 0.58 the impact factor within two years. The other laboratory effects are comparatively weak. A 10% increase in the **productivity of the laboratory** would decrease the impact factor of 0.05, so that a substitution effect reciprocal of the one identified in the previous paragraph is suggested. A productive environment may stimulate individual productivity to the detriment of individual quality.

Finally, the **size** of the laboratory has a small negative impact on the quality of the publications of its members, but at a decreasing rate, as shown by the positive size squared estimate. Interestingly, the dummy for small labs has an important positive coefficient, which suggests that small labs favours higher impacts of individuals publications. This could be related to the fact that small labs are often non experimental labs but it is not obvious in what sense since the dummy for the Grenoble region where large experimental tools are is positive.

The effect of time is not linear as previously. The estimates of the time dummies are negative but shows that the impact factor decreased until 1988, then increased afterwards until 1997 where it approaches the level of 1986, except for 1991 and 1994.

C. The determinants of the average number of citations per two years

The results of the estimation are shown in *table II.3.* The results and the effects at play are more similar to the ones identified for the productivity measure in terms of publications than in terms of impact factors.

It appears that **age** have a negative impact on the average number of citations per two years, but this effect is not statistically significant, according to the within estimates. The life cycle effect is therefore not robust in a model where productivity is defined in terms of the annual number citations received.

Being a **woman** has a negative effect on citations, almost four times higher than on the impact factor measure but almost twice smaller than the effect on publications. A woman get 0.4 citations per article less than a man on average, all else being equal, according to the TS estimates. Again, we find a positive correlation between the unobservable effects and the dummy "woman", suggesting that something like "personal motivation for quality" is related to gender (or to its correlated unmeasured "sociological" variables).

The **"Grande Ecole"** effect is almost five times higher than in the impact factor model and similar to the effect on publications (0.536). We also find again the effects that we previously interpreted as "personal network" or "prestige" positive effects on productivity.

The effect of **status** is the same as in the model estimated on the number of articles. All other effects being equal, if "talent" was equally distributed a DR would be less likely to be cited than a CR (TS estimates – note that they are not statistically significant); but if we take into account the positive correlation of "talent" with the status variable, a DR1 receives 1.0 citation more on average per year than a CR scientist and a DR2 0.35 citation more (TOTAL estimates).

Concerning the **tenure** variables, the number of citations is negatively influenced by the time spent in the status DR1. On average, a DR1 in group 2 of experience in the grade receives on average 0.6 citations less in two years per paper than a newly promoted DR1, and the figure amounts to 1.2 for a DR1 in group 3 of tenure in the grade. Yet "talent" is positively correlated to longer tenure in DR status.

Among the laboratory variables, we find a positive effect of the **quality of the lab's** publications on individual productivity. If the quality of the laboratory's publications increased by

10%, the members would receive 0.3 citations more per article. The peer effect in terms of quantitative **productivity** has no statistically significant effect on the annual number citations received in the TS regression and a negative effect in the TOTAL regression. As for the measure in terms of impact factors, a sort of substitution effect might be at play such that the labs that publish the most are not the ones that obtain the more citations for their articles. We find a strong impact of the **international openness** of the lab: if the proportion of co-published work with foreign countries increased by 10%, the members would receive 0.8 citations more per article. Finally, the size effect is slightly positive and at an increasing rate (resp. marginal impacts: 0.053 and 0.018).

III. Conclusion

This work explores the differences in productivity among scientists in public research, both in terms of the number of articles and of the quality of the publications. We use a unique longitudinal data base, concerning French condensed-matter physicists between 1986 and 1997. Three sets of factors have been considered as determinants of the researchers' productivity. First, individual variables – age, cohort, gender, education. And second, variables that may be related to the incentive devices at work in the scientific institution – the status and tenure variables, and the laboratory variables to assess environmental effects.

We find a strong impact of the individual variables. In particular, we find a "life cycle effect" such that the mean number of publications decreases with age at the end of the career. But, using the citations as a measure of productivity, we have been able to see that age does not have a negative impact on the average number of citations received by the articles per researcher and per year. We also find a gender effect according to which women publish less than men and get less cited, and that a selective pre-doc education of the researchers is determinant to foster productivity. Some institutional factors seems to play a role as important. Individual productivity is stimulated in productive laboratories that are participating in international networks of co-publication. Finally, our results suggest that promotion plays as a research incentive. A long tenure in a status, and especially in the higher status, has a negative impact on productivity.

We must develop the work in several directions. One is to account for the fact that the promotion and laboratory variables may be endogenous. Another is to focus more precisely on certain variables. In Hall, Mairesse and Turner (2005) we concentrate on age, cohorts and period effects addressing the identification problem. The results about the gender effect need to be developed as well. If the estimates suggest that men are more productive than women, they do not tell neither about the reasons of this phenomena nor about the true abilities of women scientists. As we mentioned, several sociological reasons could explain this feature. It is intriguing that on average women access to the same journals in terms of quality than men do, whereas they get less citations on average by article. It could mean that it is less easy for them to build themselves a reputation or to be part of the citations networks. It could also mean that the hypothesis according to which past productivity impacts promotion and affiliation is gender biased. Finally, it would be of great interest to assess more deeply in an econometric model peer effects and environment effects in science.

References

ADAMS J., GRILICHES Z., 1996, “Research productivity in a system of universities”, NBER.

ALLISON P.D., PRICE D.J. deSolla, GRIFFITH B.C., MORAVESIK M.J. and STEWART J.A., 1976, "« Lotka’s Law : a problem in its interpretation and application », Social Studies of Science, 6.

ALLISON P.D., STEWART J.A., 1974, « Productivity differences among scientists : evidence for accumulative advantage », American Sociological Review, 39, august.

ALLISON P.D., LONG J.S., KRAUZE T.K., 1982, « Cumulative advantage and inequality in Science », American Sociological Review, 47, october.

ALLISON P.D., LONG J.S., 1990, “Departmental Effects on Scientific Productivity”, American Sociological Review, Vol. 55, No. 4, pp. 469-478.

AUERBACH F., 1910, « Geschischtafeln der Physik », Leipzig : J.A. Barth.

BARRE R., CRANCE M., SIGOGNEAU A., 1999, «La recherche scientifique française : situation démographique », Etudes et dossiers de l’OST, 1, avril.

BREUSCH T., MILZON G. and SCHMIDT P., 1989, «Efficient Estimation Using Panel Data », Econometrica, 57(3), May, 695-700.

CHAMBERLAIN G., 1982, "Multivariate Regression Models for Panel Data," Journal of Econometrics, 18, 5-46.

CHAMBERLAIN G., 1984, Panel Data, in *Handbook of Econometrics*, eds. Z.Griliches and M.Intriligator, North Holland, Amsterdam.

CRANE D., 1965, "Scientists at Major and Minor Universities: a Study of Productivity and Recognition", American Sociological Review, 30, 699-715.

COLE J. and COLE S., 1973, *Social stratification in science,* Chicago : U. of Chicago Press.

COLE S., 1979, "Age and scientific performance", American Journal of Sociology, 84(4), january.

DASGUPTA P., DAVID P., 1994, "Toward a new economics of science", Research Policy, 23(5).

DAVID P., 1994, "Positive feedbacks and research productivity in science: reopening another black box", in Ove Granstrand, Amsterdam: Elsevier Science B. V (Eds.), *The economics of technology*.

DIAMOND A., 1984, « An economic model of the life-cycle research productivity of scientists », Scientometrics, 6(3), may.

DUGUET E., MONJON S., 2001, « Creative destruction and the Innovative Core : is Innovation Persistant at the firm level ?" mimeo, Cahier de la MSE, 2001.

HALL B.H., JAFFE A., TRAJTENBERG M., 1999, «Market value and patent citations : a first look », mimeo, june.

HALL B., MAIRESSE J., TURNER L., 2005: "Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists", mimeo.

HAUSMAN J., HALL B., GRILICHES Z., 1984, « Econometric Models for Count Data with an Application to the Patent R&D Relationship », Econometrica, 52(4), July, 909-937.

HAUSMAN J., TAYLOR, 1981, "Panel Data and Unobservable Individual Effects", Econometrica, 49(6)

HECKMAN, 1981, "Statistical models for discrete panel data", in Structural Analysis of Discrete Data with Economic Applications, eds C.Manski and D.McFadden, the MIT Press, Cambridge Massachusetts.

LEVIN S., STEPHAN P., 1991, "Research productivity over the life cycle: evidence for academic scientists", American Economic Review, 81(1), march.

LONG S., 1978, "Productivity and Academic Position in the Scientific Career", American Sociological Review, 43, 889-908.

LONG, J.S., 1992, "Measures of sex differences in scientific productivity", Social Forces, 71, 159-178.

LONG S., ALLISON P., McGINNIS R., 1979, "Entrance to the Academic Career", American Sociological Review, 44, 816-830.

LOTKA A., 1926, "The frequency distribution of scientific productivity", Journal of the Washington Academy of Sciences, 97, september.

MAIRESSE J. and L.TURNER, 2002, « Measurement and Explanation of the Intensity of Co-Publication in Scientific Research: An Analysis at the Laboratory Level », in « New Frontiers in the Economics of Innovation and New Technology », Essays in honour of Paul David", eds C.Antonelli, D. Foray, B.Hall and E.Steinmueller, Oxford University Press, forthcoming.

MERTON R., 1957, "Priorities in Scientific Discoveries: A Chapter in the Sociology of Science", American Sociological Review, 22(6), December, 635-659.

MUNDLAK Y., 1978, "On the Pooling of Time Series and Cross Section Data", Econometrica, 46(1), 69-85.

PRICE, D.deSolla, 1976, « A general theory of bibliometrics and other cumulative advantage processes » Journal of the American Society for Information Science, 27 (5/6).

STEPHAN P., 1996, "The economics of science", Journal of Economic Literature, 31, september.

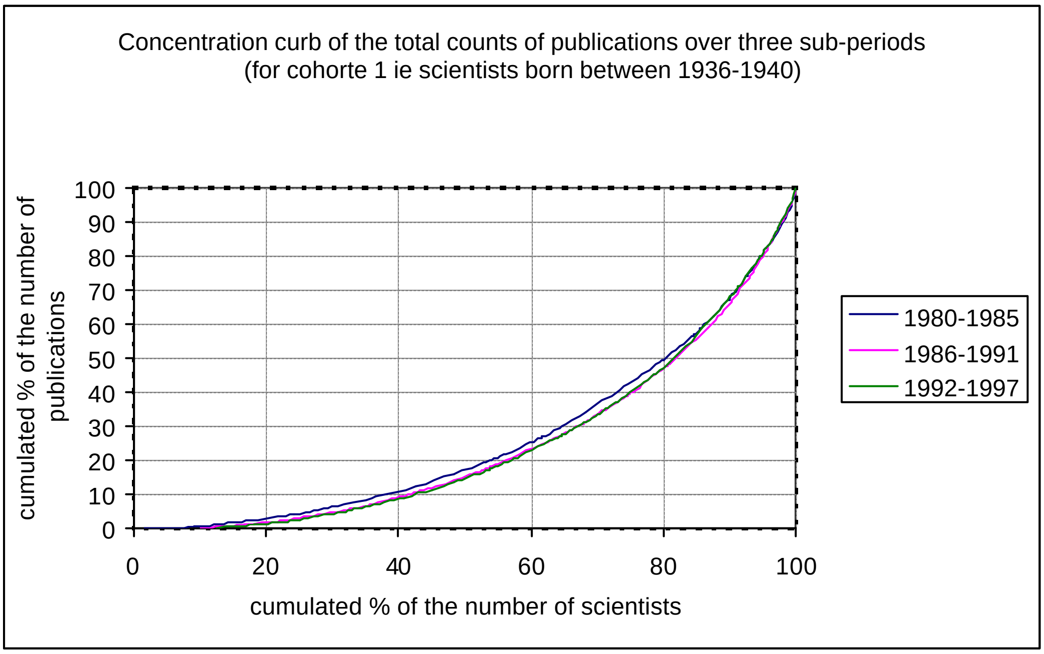
STEPHAN P., 1998, « Gender Differences in the Rewards to Publishing in Academe: Science in the 1970's », Sex Roles, Vol.38, 11/12.

TURNER L., 2003, « La recherche publique dans la production des connaissances : contributions en économie de la science », PhD dissertation, Université Paris 1.

ZUCKERMAN H., 1992, "The Proliferation of Prizes: Nobel Complements and Nobel Surrogates in the Reward System of Science", Theoretical Medicine, 13, 217-231.

ZUCKERMAN H., COLE J., BRUER J. (eds.), 1991, *The Outer Circle : Woman in the Scientific Community*, New York : W.W.Norton.

<u>Figure 1</u>



- *The graphs for the other cohorts are not reproduced since they show the same results.*
- *Cohorts of scientists are in this study "age cohorts" and not "PhD cohorts" as in most studies. For more details see section I.A.*

<u>Figure 2</u>

1992/1997 1986/1991	First quartile of researchers' productivity : the most productive researchers	Second quartile of researchers' productivity	Third quartile of researchers' productivity	Last quartile of researche rs' productivity : the less productive researchers	Total
Quartile 1	**65.5%**	23.3%	8.6%	2.6%	100%
Quartile 2	22.3%	**44.7%**	22.3%	10.7%	100%
Quartile 3	8.2%	28.1%	**40.5%**	23.1%	100%
Quartile 4	2.6%	9.5%	21.6%	**66.4%**	100%
Total	100%	100%	100%	100%	100%

- *Similarly persistence is found when we cross periods 1980/1985 and 1986/1992, and also periods 1980/1985 and 1992/1997.*

Table I.1

	Mean	Standard Error	Median	1st Qrt	3rd Qrt
Dependent variables					
Number of articles per year (ART)	2.69	3.21	2	0	4
Average impact factor per researcher and per year (NOT_I)	2.66	2.30	2.54	0	3.8
Average number of citations (within 2 years) per researcher and per year (MCIT)	3.50	6.10	2	0	4.8
Extensions					
Average number of authors per article (harmonic) NOTA	3.23	2.57	3.33	0	4.90
Average number of pages per article NOT_P	5.48	4.68	5.4	0	7.83
Individual variables + Time Dummies					
AGE	44.65	8.03	45	38	51
Age cohort 1 (AGE1) 26<=age<=38	0.25	0.44	0	0	1
Age cohort 2 (AGE2) 38<age<=45	0.25	0.43	0	0	0
Age cohort 3 (AGE3) 45<age<51	0.23	0.42	0	0	0
Age cohort 4 (AGE4) 51<=age<=61	0.27	0.44	0	0	1
Education in a “Grande Ecole” (ECOLE)	0.17	0.38	0	0	0
Gender (WOMAN)	0.18	0.39	0	0	0
More than one mobility (DUMCH23)	0.13	0.34	0	0	0
Promotion variables					
Status (DR2_0)	0.08	0.28	0	0	0
Status (DR1_0)	0.33	0.47	0	0	1
Tenure in status CR cohort 1 (C1ANCCR)	0.21	0.41	0	0	0
Tenure in status CR cohort 2 (C2ANCCR)	0.19	0.39	0	0	0
Tenure in status CR cohort 3 (C3ANCCR)	0.18	0.39	0	0	0
Tenure in status DR2 cohort 1 (C1ANCDR2)	0.13	0.33	0	0	0
Tenure in status DR2 cohort 2 (C2ANCDR2)	0.10	0.3	0	0	0
Tenure in status DR2 cohort 3 (C3ANCDR2)	0.10	0.3	0	0	0
Tenure in status DR1 cohort 1 (C1ANCDR1)	0.03	0.18	0	0	0
Tenure in status DR1 cohort 2 (C2ANCDR1)	0.02	0.15	0	0	0
Tenure in status DR1 cohort 3 (C3ANCDR1)	0.03	0.16	0	0	0
Laboratory variables					
Size of the laboratory in logarithm LOGNBCH	3.23	1.39	3.64	3	4.09
Productivity of the laboratory in logarithm (LOGPROD)	0.72	0.42	0.79	0.44	0.99
Quality of the laboratory publications in logarithm (LOGQUAL)	1.09	0.45	1.27	1.09	1.33
Proportion of the laboratory articles with foreign co-authors (MPETR)	0.04	0.03	0.04	0.02	0.06
Dummy for the Grenoble region (ADUMGREN)	0.26	0.44	0	0	1
Dummy for the Paris region (ADUMPAR)	0.36	0.48	0	0	1
Dummy for laboratory with less than 3 researchers (DUMEF13)	0.14	0.34	0	0	0

- *Number of individuals = 465, Number of years = 12, Number of Observation = 5580*

Table II.1

	POISSON ON ART		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE2	0.205*** (0.025)	0.098*** (0.029)	0.553	0.263
AGE3	0.233*** (0.025)	0.137*** (0.032)	0.628	0.368
AGE4	0.072*** (0.025)	0.086*** (0.028)	0.193	0.233
WOMAN	-0.273*** (0.024)	-0.33*** (0.046)	-0.736	-0.89
ECOLE	0.118*** (0.021)	0.26*** (0.043)	0.317	0.701
DUMCH23	0.041* (0.025)	-0.024 (0.051)	0.11	-0.064
DR1	0.543*** (0.041)	-0.287*** (0.096)	1.463	-0.772
DR2	0.174*** (0.031)	-0.045 (0.06)	0.47	-0.121
C2ANCCR	-0.094*** (0.029)	0.028 (0.054)	-0.252	0.077
C3ANCCR	-0.291*** (0.032)	0.054 (0.079)	-0.786	0.145
C2ANCD2	0.012 (0.031)	-0.129*** (0.042)	0.031	-0.347
C3ANCD2	-0.029 (0.034)	-0.2*** (0.062)	-0.078	-0.539
C2ANCD1	-0.05 (0.057)	0.023 (0.07)	-0.134	0.063
C3ANCD1	-0.287*** (0.058)	-0.494*** (0.091)	-0.774	-1.33
MPETR	2.942*** (0.432)	3.01*** (1.07)	7.928	8.112
LOGNBCH	-0.131*** (0.015)	-0.091** (0.036)	-0.131	-0.091
LOGNBCH2	0.016*** (0.006)	0.016 (0.011)	0.016	0.016
LOGPROD	0.233*** (0.032)	0.274*** (0.067)	0.233	0.274
LOGQUAL	-0.254*** (0.069)	0.018 (0.144)	-0.254	0.018
ADUMGREN	0.173*** (0.026)	0.179*** (0.053)	0.465	0.483
ADUMPAR	-0.055** (0.026)	0.009 (0.047)	-0.147	0.026
DUMEF13	-0.66*** (0.099)	-0.109 (0.202)	-1.78	-0.293

A87	0.117*** (0.046)	0.114** (0.046)	0.315	0.307
A88	0.203*** (0.045)	0.219*** (0.045)	0.548	0.589
A89	0.336*** (0.043)	0.362*** (0.044)	0.905	0.974
A90	0.191*** (0.045)	0.254*** (0.046)	0.514	0.684
A91	0.227*** (0.044)	0.329*** (0.046)	0.612	0.885
A92	0.262*** (0.044)	0.393*** (0.046)	0.707	1.06
A93	0.457*** (0.042)	0.599*** (0.045)	1.232	1.615
A94	0.388*** (0.043)	0.55*** (0.046)	1.046	1.481
A95	0.297*** (0.044)	0.482*** (0.048)	0.8	1.298
A96	0.393*** (0.043)	0.583*** (0.047)	1.059	1.572
A97	0.373*** (0.043)	0.575*** (0.048)	1.005	1.55
C	0.95*** (0.105)	0.372** (0.19)	2.561	1.003
Log-vraisemblance R^2-Adj for Step 2	-14009.8	-8978.58 0.03		

- *Number of individuals = 465, Number of years = 12, Number of Observation = 5580*
- *Standard Errors computed from analytic second derivatives (Newton) for the TOTAL and First Step and from quadratic form of analytic first derivatives (Gauss) for the Second Step.*

Table II.2

	LOGLINEAR ON NOT_I		MARGINAL IMPACTS	
Variables	Total	Within	Total	Within
AGE2	-0.041* (0.021)	-0.002 (0.029)	-0.109	-0.004
AGE3	-0.104*** (0.026)	-0.06 (0.042)	-0.276	-0.16
AGE4	-0.144*** (0.03)	-0.111** (0.055)	-0.383	-0.295
WOMAN	-0.022 (0.016)	-0.04** (0.016)	-0.058	-0.106
ECOLE	0.032* (0.017)	0.042** (0.016)	0.085	0.111
DUMCH23	-0.012 (0.019)	-0.01 (0.019)	-0.031	-0.026
DR1	0.135*** (0.042)	-0.054 (0.074)	0.36	-0.143
DR2	0.058** (0.026)	-0.041 (0.046)	0.155	-0.109
C2ANCCR	-0.004 (0.02)	-0.02 (0.04)	-0.01	-0.054
C3ANCCR	0.015 (0.028)	-0.011 (0.056)	0.041	-0.029
C2ANCD2	0.041 (0.026)	-0.021 (0.031)	0.11	-0.055
C3ANCD2	0.063** (0.029)	0.039 (0.046)	0.166	0.105
C2ANCD1	0.033 (0.052)	0.025 (0.059)	0.089	0.066
C3ANCD1	0.006 (0.05)	0.069 (0.077)	0.016	0.184
MPETR	0.031 (0.326)	0.121 (0.327)	0.082	0.322
LOGNBCH	-0.033*** (0.012)	-0.025** (0.012)	-0.033	-0.025
LOGNBCH2	0.021*** (0.004)	0.024*** (0.004)	0.021	0.024
LOGPROD	-0.046* (0.024)	-0.049** (0.024)	-0.046	-0.049
LOGQUAL	0.57*** (0.051)	0.579*** (0.05)	0.57	0.579
ADUMGREN	0.029 (0.019)	0.044** (0.019)	0.077	0.118
ADUMPAR	0.009 (0.018)	0.025 (0.018)	0.023	0.067
DUMEF13	0.501*** (0.074)	0.551*** (0.073)	1.335	1.467

A87	-0.051* (0.03)	-0.051* (0.028)	-0.135	-0.135
A88	-0.123*** (0.03)	-0.122*** (0.028)	-0.327	-0.325
A89	-0.109*** (0.03)	-0.103*** (0.029)	-0.29	-0.274
A90	-0.1*** (0.03)	-0.085*** (0.03)	-0.265	-0.227
A91	-0.123*** (0.03)	-0.104*** (0.031)	-0.328	-0.277
A92	-0.076** (0.03)	-0.054* (0.032)	-0.203	-0.143
A93	-0.191*** (0.031)	-0.167*** (0.033)	-0.509	-0.444
A94	-0.087*** (0.031)	-0.065* (0.034)	-0.232	-0.172
A95	-0.057* (0.031)	-0.03 (0.035)	-0.151	-0.08
A96	-0.036 (0.032)	-0.007 (0.037)	-0.095	-0.019
A97	-0.062** (0.032)	-0.035 (0.038)	-0.166	-0.093
DUMMY (ART=0)	-1.119*** (0.014)	-1.093*** (0.016)	-2.979	-2.91
C	0.566*** (0.077)	0.516*** (0.071)	1.506	1.373
Log likelihood R^2-Adj	-3482.77 0.557	-2941.03 0.602		

- *Number of individuals = 465, Number of years = 12, Number of Observation = 5580*

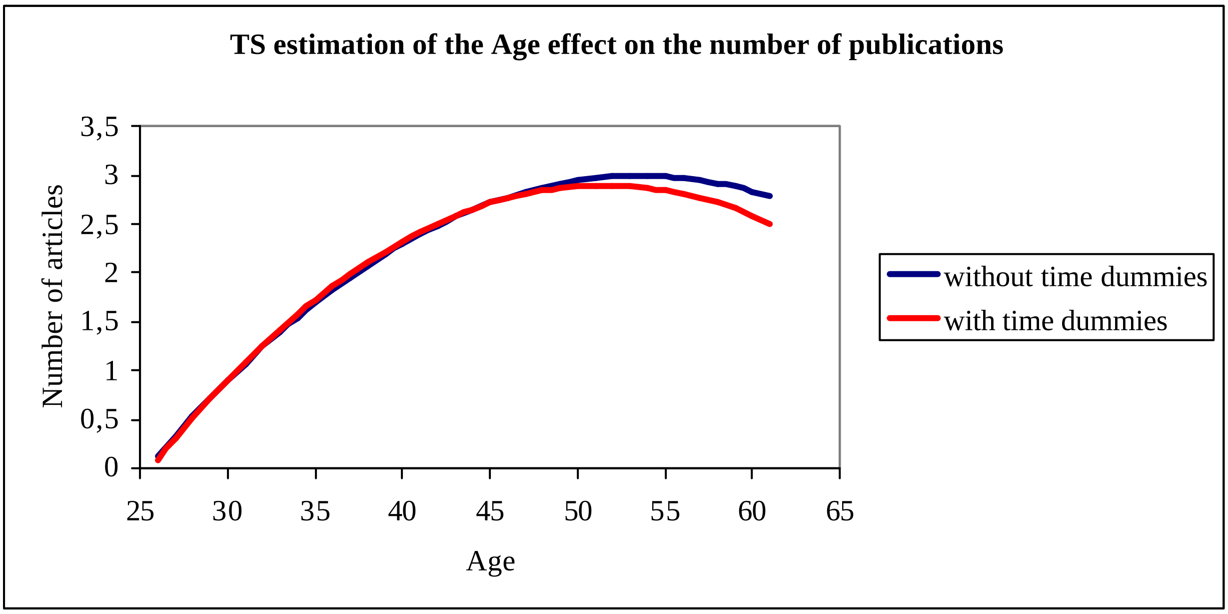
Table II.3

	LOGLINEAR ON MCIT_2		*MARGINAL IMPACTS*	
Variables	TOTAL	TWO STEP	TOTAL	TWO STEP
AGE2	-0.022 (0.036)	0.006 (0.05)	-0.077	0.023
AGE3	-0.103** (0.043)	-0.028 (0.073)	-0.357	-0.097
AGE4	-0.172*** (0.049)	-0.019 (0.096)	-0.597	-0.066
WOMAN	-0.049* (0.029)	-0.111*** (0.03)	-0.172	-0.385
ECOLE	0.062* (0.033)	0.154*** (0.033)	0.215	0.536
DUMCH23	0.078** (0.033)	0.083** (0.034)	0.27	0.29
DR1	0.306*** (0.084)	-0.164 (0.134)	1.065	-0.571
DR2	0.101** (0.045)	-0.057 (0.067)	0.352	-0.197
C2ANCCR	-0.054 (0.038)	0.015 (0.052)	-0.187	0.052
C3ANCCR	-0.011 (0.05)	0.016 (0.078)	-0.038	0.054
C2ANCD2	-0.022 (0.044)	-0.13** (0.053)	-0.077	-0.451
C3ANCD2	0.023 (0.048)	-0.072 (0.083)	0.081	-0.251
C2ANCD1	-0.051 (0.096)	-0.183** (0.093)	-0.177	-0.638
C3ANCD1	-0.01 (0.097)	-0.355** (0.148)	-0.034	-1.234
MPETR	1.756*** (0.597)	1.944*** (0.62)	6.109	6.766
LOGNBCH	0.011 (0.021)	0.053** (0.022)	0.011	0.053
LOGNBCH2	0.015** (0.007)	0.018** (0.007)	0.015	0.018
LOGPROD	-0.104** (0.043)	-0.056 (0.044)	-0.104	-0.056
LOGQUAL	0.28*** (0.09)	0.301*** (0.093)	0.28	0.301
ADUMGREN	0.128*** (0.035)	0.141*** (0.037)	0.446	0.489
ADUMPAR	0.031 (0.033)	0.082** (0.035)	0.107	0.284
DUMEF13	0.29** (0.13)	0.544*** (0.133)	1.009	1.894

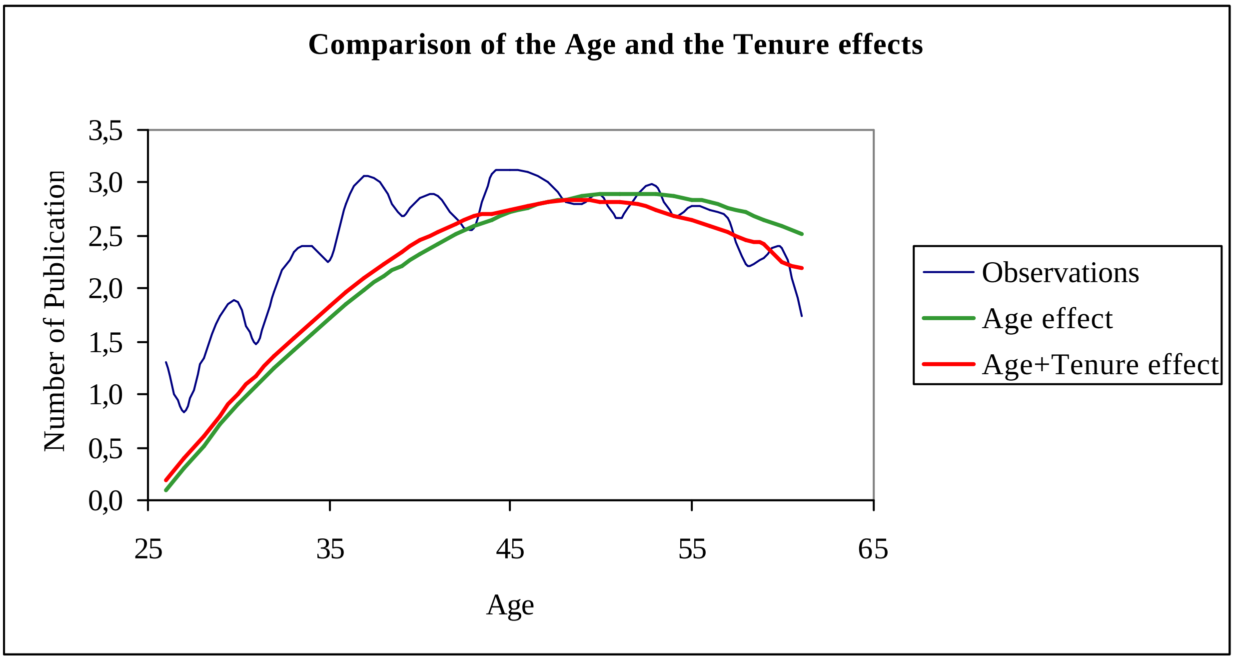
A87	0.042 (0.046)	0.046 (0.043)	0.147	0.16
A88	0.0001 (0.046)	0.008 (0.044)	0.0002	0.028
A89	0.01 (0.046)	0.022 (0.046)	0.036	0.077
A90	0.006 (0.047)	0.027 (0.05)	0.019	0.095
A91	0.0004 (0.047)	0.032 (0.053)	-0.001	0.11
A92	0.01 (0.048)	0.056 (0.057)	0.034	0.195
A93	-0.349*** (0.048)	-0.309*** (0.061)	-1.215	-1.073
A94	-0.209*** (0.049)	-0.17*** (0.065)	-0.729	-0.593
DUMMY (MCIT_2=0)	-1.233*** (0.024)	-1.095*** (0.027)	-4.29	-3.808
C	0.835*** (0.135)	0.549*** (0.133)	2.904	1.909
Log likelihood R^2-Adj	-3223.11 0.50	-2728.45 0.58 (step1) 0.46 (step2)		

- *Number of individuals = 352, Number of years = 9, Number of Observation = 3168*

Graph II.1


TS estimation of the Age effect on the number of publications
Number of articles
3,5
3
2,5
2
1,5
1
0,5
0
25
30
35
40
45
50
55
60
65
Age
without time dummies
with time dummies

Graph II.3



Graph II.4