Individual Productivity Differences in Scientific Research: An Econometric Exploration of Publications by French Physicists

Laure Turner[a] and Jacques Mairesse[b]

In revision September 2007

Keyword: economics of science, scientific productivity, bibliometrics.
JEL Classification: C51, O32.

We are thankful to Serge Bauin and Michèle Crance of the Unité des Indicateurs de la Politique Scientifique (UNIPS) au Centre National de la Recherche Scientifique (CNRS) for their help in constructing the data base and for useful suggestions. We are also grateful to Bronwyn Hall, Richard Blundell, Paul David, Emmanuel Duguet, David Encaoua, Paula Stephan and Frank Windmeijer for advice and comments at various stages of our work, as well as to other participants in seminars and conferences where we presented it.

[a] CREST-ENSAE. Email: laure.turner@ensae.fr
[b] CREST-ENSAE; UNU-MERIT, University of Maastricht, Netherlands; and NBER, USA.
Email: mairesse@ensae.fr

Individual Productivity Differences in Scientific Research: An Econometric Exploration of Publications by French Physicists

Laure Turner and Jacques Mairesse

ABSTRACT

An empirical regularity has often been observed in economics of science: productivity differences among researchers are extremely large and persistent, and a prolific minority of scientists produces most of the publications and accounts for most of citations in almost all research fields. In this study, we investigate to what extent such dispersion and persistence can be accounted by three types of factors in so far as we can measure them: individual variables, mainly age and gender, career stage variables, and laboratory variables. Does individual scientific productivity significantly drop off as scientists become older and more or less advanced in their careers? Is it strongly related or not to career promotions and to the productivity and quality of the laboratories in which scientists work? Even if these factors prove quite significant, is it nonetheless the case that individual productivity differences remain largely unaccounted by them, and that they have to be mostly imputed to unobserved individual circumstances and characteristics, or so called "individual effects"? To answer such questions, we have put together a twelve year panel database for 465 condensed matter physicists working in the French public research organization CNRS, and we have specified and estimated a simple econometric model for both "productivity" and "quality", respectively measured by the number of publications per scientist per year and by the corresponding average citation impact of these publications.

Introduction

In the sociology of science, and more recently in economics of science, an empirical regularity has been found in numerous studies: research productivity is highly variable among scientists and productivity differences persist over the life cycle of given cohorts. Since Lotka's seminal article (Lotka, 1926), it has been repeatedly observed in almost all scientific fields that the distribution of publication counts is very dispersed, persistent and left-skewed, and that, consequently, a prolific minority of scientists produces most of the publications. In the data collected on French physicists presented below, the 25% more productive of them in a given year publish four or more articles, contributing to about ??% of the total number of publications, while the 25% less productive in that year publish no articles: see the frequency and cumulative frequency histograms and the concentration curve of the productivity distribution in Figure 1. Moreover, as can been seen in Table 1, about two third of the group of the 25% more productive researchers over the first six years 1986-1991 stay in this same group over the six following years 1992-1997, only less than 3% dropping down in the group of the 25% less productive, and likewise the same ranking stability is observed for the group of the 25% less productive researchers in the first six years period. (**Insert here a Footnote on cumulative advantage: Laure?% of publications of the 66% we remaining top quartile and low quartile in the two subperiods : do we see anything ??)**

Very dispersed and persistent distributions of productivity can be an issue for science policy and for the allocation of resources to research. The efficiency of knowledge production in public research has been challenged because up to half of the papers in a research field are published by a much smaller proportion of scientists (20% in our case-see concentration curve in Figure). Does the reason lie in an unequal distribution of talent among researchers? Are there some cumulative phenomena at play so that initial success (failure) leads to permanent high (low) productivity? Do research incentives influence the propensity to publish or are scientists pushed forward by some innate and unobservable aptitude for creativeness? A better understanding of the processes explaining scientific productivity is needed.

In order to gain this understanding, we built a longitudinal database containing the publications of French physicists for a 12 year period from 1980 to 1997. Our goal was to empirically explore hypotheses formulated in Economics of Science about the determinants of research productivity. A first hypothesis is that individual scientific productivity is related to the research incentives in science. Incentives in research are specific in that they are to a large extent non monetary and reputation based. Zuckerman(1992) identified the recipients of about 3000 North American science awards in the beginning of the 1990s. They received their prizes because they had been the first to discover and the first to publish, according to the "priority rule" in science(Merton 1957). Successful researchers are more likely to obtain further grants, be given more free time for their research and work in better laboratories than scientists who had not received a reward. In this approach, non monetary incentives explain the distribution of scientific outcomes by creating the conditions of a cumulative advantage (David 1994). Empirically assessing the role of these cumulative advantages is beyond the scope of this study, however, we will consider that promotions and laboratory affiliations "reward" past productivity and are likely to have a lasting impact on future publication. Our hypothesis is that promotion and membership in a dynamic laboratory with numerous research collaborations will fuel the process of cumulative advantage and thereby increase a researcher's propensity to publish.

A second approach uses laboratory variables to capture contextual 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) studied 80 laboratories belonging to a large French University and found a correlation between individual productivity and the way in which work within those laboratories was organized as measured by the ratio of permanent, teaching, doctoral and post-doctoral researchers. Mairesse and Turner (2002) showed that geographical proximity, size and the general productivity of laboratories positively influences co-publications in laboratory networks. In this paper we will look at the influence of a laboratory's global output on individual performance, as well as the effect of its size and of its share of international collaborations.

A third approach concerns the influence of individual variables on productivity such as age (Diamond (1984), Levin and Stephan (1991)), gender (Stephan, 1998), and training in prestigious university departments (Crane, 1965, Long, Alison and McGinnis, 1979). Diamond (1984) and Levin and Stephan (1991) built life cycle models of the quadratic relationship between age and productivity in order to explain why the number of publications tends to decrease towards the end of a career. These authors wanted to know if the aging of the United States scientific community was going to impact negatively on national scientific output. In addition to these observable differences between individuals, we will also attempt to account for "non measurables", that is, the innate ability, intuition or motivation which has been referred to as the "sacred spark" (Cole et Cole (1973)) or a taste for "puzzle-solving" (Levin and Stephan (1991)). Our model contains individual random effects in an attempt to test for this idea that the real drive for doing research lies inherently in a person's own creativeness.

This paper simultaneously analyzes the influence of individual, laboratory and career specific variables on scientific productivity. In the next section, we will describe the longitudinal database built for that purpose, specify our model and estimation methodology. In Section III we will present our results. The model shows that career incentives and laboratory dynamism explain variations in research productivity to the same extent as individual factors and that, consequently, science policy will make a difference to the competitivity of a research system. Section IV draws the conclusions of our work.

I. Specification and Estimation Methodology

To our knowledge, few panel databases have been built to study scientific productivity. One was built for the period 1991–2002 by C. Gonzalez-Brambila. It contains information on the age, gender, year of PhD and the research field of members of the Mexican National System of Researchers (SNI). Another was built by Levin and Stephan (1991) to study the North American age/productivity relationship for the period 1973-1979. Our database contains the publications and citations of 497 French physicists for the period 1986-1997 with information on laboratory affiliation, job status, age, gender, pre-doctoral formation and career promotions. These combined features make the data set relatively rich and original for studying scientific productivity. The data was obtained from the *Science Citation Index (SCI)* an internationally recognized source for bibliometric studies produced by the *Institute for Scientific Information (ISI).*

A. The Data

Productivity was measured using publication counts and two measures of publication quality averaged per year and per researcher. The first is the impact factor of the journal in which an article is published. A journal's impact factor is obtained by calculating the number of citations received during years T and T-1 by articles published during years T-1 and T-2, divided by the total number of articles in the sample. The second is the number of citations given to an article[1]. This number is traditionally considered in bibliometric studies as an indicator of an article's impact. These measures are related as follows: For individual i at date t, for journal j $\in$ J(it) and for the set of journals publishing papers from i at t, let *art* be the number of articles and *imp* be the journal impact factor. The quantitative measure of productivity in terms of article counts is $A_{it} = \sum_{j \in J(it)} art_{itj}$. The qualitative measure of productivity in terms of impact factors is $I_{it} = \sum_{j \in J(it)} art_{itj} * imp_{jt} / A_{it}$. And finally the qualitative measure of productivity in terms of citations is C_{it} = the sum of citations in year t, t+1 and t+2 received by the articles published by I at date t) / A_{it}. We surprisingly found that the citation and journal impact measures only correlated at 0.37 for our sample despite the fact that they are often viewed as substitutes for one another when assessing publication quality. Also, the correlation of the quantitative measure (article counts) with the impact factor quality measure is 0.36 and with the citation count quality measure is 0.26. We consequently decided not to use a weighted measure of article counts by quality because quantity and quality are clearly two distinct dimensions of scientific productivity.

The group of 497 physicists studied here represents almost all the CNRS[2] researchers working in condensed matter physics (654 in 1996). This field was chosen for two reasons. First, it is a field of pure basic science; journals with a strong reputation are clearly identifiable; the size of the field is clearly defined; and there is very little mobility among researchers from public research to teaching or to the private sector. Second, condensed matter research is growing quickly in France, was honoured by the award of the Nobel Prize for Physics to Pierre-Gilles de Gennes in 1991, and currently accounts for close to half of all French academic physics[3]. With respect to the scientists selected for inclusion in our sample, they were all born between 1936-1960, however, they entered the CNRS at different dates. 433 entered the CNRS before 1986, so we expected them to publish over the whole period of our study between 1986-1997. 32 entered the CNRS after 1986 but had started to publish before that date which meant that they too published articles over the entire period of our study. Between the

[1] We considered the citations received per article within two years. The period covered in the citation set is therefore 1986-1994. On average, an article receives approximately 40% of its citations within two years according to our data.
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] The Centre National de la Recherche Scientifique (CNRS) is a public organization with 25,000 employees (11,000 researchers and 14,000 engineers, technicians and administrative staff) Its mission is to carry out fundamental research in all areas of knowledge. University researchers often belong to CNRS labs, however, university research and CNRS research are institutionally two distinct areas of activity. therefore called "mixed units".
[3] 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 anchored in several experimental traditions (crystallography, diffusion of neutrons and electrons, magnetic resonance imagery, microscopy, etc.)

year of their first publication and their entry at the CNRS, they were assigned the job status that they received when hired. A final group of 32 researchers entered the CNRS too late to have published over the entire period and were consequently eliminated from the sample. So in the end our balanced sample contained 465 scientists for modeling publication count productivity. But for modeling citation count productivity, we were only able to find data for 352 scientists. The range of birth dates for these researchers was smaller than in the full sample, from 1936 to 1955 for the former and from 1936 to 1960 for the latter.

B. *Explaining Scientific Productivity*

Individual research productivity was measured by the mean number of publications per year per scientist, by the impact factor of journals averaged on the articles per year and per scientist, and by the number of citations per article averaged 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 published 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 are published in journals whose articles receive an average of 2.7 citations over two years. However, the quality of some journals is low – they receive almost no citations – whereas others receive up to 21.5 citations for an average two year period. Approximately 32 000 citations (within two years) were received by the publications of the scientists studied, which amounts to 3.5 citations per article, per researcher and per year on average over the period, with a standard error of 6.

Sections 1, 2 and 3 which follow discuss the relative impact on scientific productivity of individual factors, career incentives and context variables. Several methodological problems are identified which are then examined in detail in section 4.

1. Individual variables

---

and theories (static physics). It also has close ties with industry around materials used in electronics, plastics,

Age

The age-productivity relationship is a major issue for sociologists and economists of science. A negative relationship has been observed suggesting that young researchers are more productive than older ones. According to Lehman (1953), productivity is highest for scientists in their late thirties or early forties, a little sooner in Mathematics and Theoretical Physics, a little later in Biology and Geology (see also Cole, 1979). Life cycle models have been used by Diamond (1984), and Levin and Stephan (1991) to look both at how research effort is allocated over time and at productivity levels at the end of careers. Their conclusions confirm the decline of productivity over the life cycle. One reason lies in the fact that scientists engage in research because of the financial rewards associated with research activities. Retirement, of course, puts an end to that type of incentive. Finally, the negative age/productivity relationship was verified in six sub-fields of physics and earth science including solid state and condensed matter physics using panel data from the *SCI* and the *Survey of Doctorate Recipients* for the period 1973-1979[4]. Bonaccorsi and Daraio (2003) studied how the age structure of a research Institute affected its overall productivity. Evidence was once again found to show that when the average age of a research population increases the Institute's scientific productivity declines.

In our study, the age dispersion of the sample is high, with an average age of 44.6 years and a standard deviation of 8.0. The age/productivity relationship was studied for individual scientists and our results were compared to those obtained by Levin and Stephan (1991). The age variable used in our model is described in section 4.

Gender

Are rewards in science gender biased? It is appropriate to ask this question in our study because men represent 82% of our sample. Several studies have concluded that women publish less than men and that they earn less as well (see Stephan 1998). Zuckerman, Cole and Bruer (1991) show that a process of cumulative advantages might explain why women constantly appear in the "outer circle of science" because it amplifies an initial situation where women published less than men. But in empirical studies, the relationship between gender and outcome is often biased because estimations depend upon cross-sectional data which do not account for unmeasurable individual effects reflecting personal motivation, talent or some other similar individual variable. Moreover, the samples are generally not random but consist of successful scientists thereby introducing a selection bias as well. Using panel data allows us to avoid these biases. For example, when Levin and Stephan (1998) used panel data, they found that gender was not a significant determinant of salary changes in US academe

---

food or cosmetic gels, etc.

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

during the 1970's. Gender is introduced as a dummy variable in our model, and is equal to 1 when the scientist is a woman.

"Grande Ecole" dummy

A second dummy variable equal to 1 was used when a researcher had studied in a "Grande Ecole" in addition to graduating with a PhD[5]. 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, and the remainder to other Grandes Ecoles. We expect this dummy variable to have an effect on our measures of individual productivity since different studies have shown the importance of pre-doctoral training in explaining productivity differences in research (see for instance Long, Allison and McGinnis, 1979). We assume that the knowledge, values, and scientific performance criteria learned during that period will have a lasting positive impact on their work.

Individual heterogeneity

As mentioned above, a frequent hypothesis concerning scientific productivity is that it depends upon some individual quality such as a "sacred spark" or a specific aptitude to be creative. We've introduced random effects specific to individuals in order to test for what is often called in the literature the "individual heterogeneity effect". The estimation method we chose was determined by the existence of a correlation between the random individual effects and the explanatory variables.

2. Career incentives

The researchers were distributed according to the evolution of their career so that we could study the link between publication and promotion and more generally account for the role of incentives in producing science. A researcher with a typical career profile enters the CNRS as a "Chargé de Recherche" (CR), is then promoted to become a class 2 research director, (DR2) before becoming a class 1 research director, (DR1). However, many researchers in our sample were never promoted and remained 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 got promoted DR2. The most difficult promotion to obtain is DR1: only 10% of the sample succeeded in obtaining it.

[5] In the French educational system, after students graduate from high school, they can either go directly to University, or, if they have the grades, they can decide to write competitive exams for admission to France's elite schooling system, "les Grandes Ecoles". Students admitted to a "Grande Ecole" have therefore gotten over two hurdles: one concerning the quality of their high school grades which, if they are good enough, allow them to take part in a two year training program in order to prepare their competitive exams; the other the exam itself. .

A descriptive study (Turner, 2003) of our sample has shown, for each year, a positive relationship between publication and job status, however, there was no evidence of a linear relationship between past publication and current position.. This is surprising because the CNRS tends to reward publication with promotion[6]. Looking at the inverse relationship to see if promotion constitutes an incentive to publish more did not produce any conclusive results. On average researchers are just as productive after their promotion as they were before it. However, their individual productivity varies as a function of status, age and time with older scientists publishing less after promotion and younger scientists publishing more after a DR2 promotion.

Explanatory variables relating to career paths and promotion profiles were introduced into our econometric model in order to determine the influence on publishing behaviour of being or *not* being promoted. The variables were chosen to capture observed changes: The productivity of scientists who were never promoted and who remained CR over the entire period of our study first increased with time and then decreased as a result of possible discouragement. DR2 productivity remained constant over time and was neither influenced by promotion incentives nor by their position in the job organization of the CNRS. Finally, DR1 productivity decreased with tenure in the grade. Consequently, we retained tenure in each grade as the career path variables and used dummies for the grades DR2 and DR1[7]. These decisions will be explained in more detail in section 4.

3. Context variables: laboratory and peer group effects

A second descriptive study of our data was carried out on collaboration among scientists (Mairesse and Turner, 2002). We found that the dynamism of a laboratory as measured by its size, overall productivity, quality of publications and intensity of international collaborations, influenced the laboratory's co-publication level. In this study, the impact of these same laboratory variables was studied but, this time, our focus was on individual productivity. We assumed that belonging to a dynamic laboratory will stimulate a researcher's individual propensity to publish and consequently be a factor of cumulative advantage. Because scientists will more easily gain peer recognition by working in this type of laboratory, they will likely receive as well more resources for publishing.

The laboratories where researchers were working during 1997, the last year of our study, were used to define their working context. The laboratory variables defined below were therefore time

[6] Traiter la rq du referee suivante : The paper would benefit from contrasting the French system with the U.S. system that is often studied. For example, it is my understanding that there is less of a relationship between salary and productivity in France than in the U.S

[7] Several tests of our data were carried out in order to control for the total number of years a scientist had been working at the CNRS and for the impact of ending one's career. Both factors were of less importance in explaining productivity than the effect of tenure in a grade.

invariant and some aggregated over the entire 12 year period of our study. More precisely, the laboratory variables are the following:

- Size: the size of the laboratory is the total number of researchers in that laboratory in 1997 including University scientists affiliated to the lab. The size variable is also centered and squared in order to measure quadratic effects.
- Productivity of the laboratory : this was determined by taking the total number of publications published by a laboratory for the whole period 1986-1997 and then subtracting from that sum the number of publications over the period which were produced by each researcher considered individually. We took the logarithm of this variable which we then used as a proxy measure for defining the dynamism of the researcher context.
- Quality of the laboratory: this was determined by taking the average impact factor obtained by all the articles published by a laboratory over the whole period 1986-1997 and then subtracting from that number the average impact factor obtained for the articles published individually by each researcher for the period 1986-1997. Once again, we took the logarithm of this variable and used it as a proxy measure for defining the quality of a researcher environment.
- Intensity of international cooperation: it is the proportion of articles written by the laboratory for the whole period which are 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 were introduced to qualify the laboratories in 1997 because most of the physicists in our sample reside in those two regions and most of the publications over the period were produced there. The Grenoble and Paris regions include all the towns containing CNRS laboratories which are geographically situated less than 100 km from these two cities.

The mobility of researchers is obviously important in correctly evaluating a "laboratory effect" and, consequently, using only the 1997 laboratory affiliation of a researcher might seem inappropriate. However, three reasons justify our decision. First, the mobility of researchers is in fact 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, we analysed the sub-set of all those scientists who never changed laboratories (55% of the total sample) and obtained the same results as those obtained on the full sample. Finally, our model was confirmed by calculating the average productivity, quality and international collaboration intensity of a laboratory for 6 years and comparing the results with those obtained for 12 years. The results were the same in both cases. We nevertheless added a dummy equal to one when the number of changes is more than one.

4 Data limitations : identification and endogeneity issues

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

Our statistics show a steady, yearly increase in the number of publications to the end of the period which is age independent. We introduced time dummies in our regressions in order to account for work environment or state of the art changes in condensed matter physics which might explain this observation.

Levin and Stephan (1991) have argued that vintage effects should also be taken into account when estimating the influence of age on productivity. In a model which did not account for individual heterogeneity, they included vintage dummies that correspond to PhD cohorts. A cohort was composed of all the researchers who received their PhD in a given year in order to capture the influence of changes in research interests and employment opportunities on cohort productivity. For our study, no such data was available so we replaced PhD cohorts with birth cohorts on the assumption that individuals finish their doctoral training at approximately the same age in France. We then estimated simultaneously the effect of age, time and birth on productivity but the three variables are collinear and the identification is spurious. This well-known identification problem is hard to resolve especially when using fixed effects, which is our case in this study. Consequently, we decided not to control for the cohort effect in this paper. However, in another paper, Hall, Mairesse and Turner (2005) have dealt specifically with this issue and have produced a solution using the same data.

Another identification problem arises when using age, tenure and career path variables. Colinearity exists between age and tenure when researchers enter the CNRS at the same age. It also exists between time and tenure in status because a significant number of researchers were never promoted. This colinearity was particularly evident when the two variables were considered in deviation from the means. Moreover, when fixed effects are considered, we can't simultaneously estimate correctly time dummies with age. We dealt with this identification problem by forming groups of age and groups of tenure within job status. Taking the DR1 status as an example, the three groups of tenure were 0-1, 2-5 and more than 5 years as DR1, the first group serving as the reference[8] (see table I.1). In this way we were able to estimate the impact on productivity of long appointments as CR, DR2 or DR1. The age groups were: group 1 (reference group), 26<=age<=38; group 2, 38<age<=45; group 3, 45<age<51; group 4, 51<=age<=61. These groupings went a long way in breaking the colinearity between age and time and between tenure and time and we were consequently able to estimate the influence of these two variables over time. However, in order to address the specific issue of productivity peaks as a function of age we adopted another model[9]. This model uses

[8] For instance we estimate the effect on productivity of the dummy "spend 2-5 years as DR1" and "spend more than 5 years as DR1" when the reference is "is a newly promoted DR1 (0-1 year as DR1)".

[9] When we use age and age squared instead of age groups, the tenure estimates are similar and the tenure effects are the same.

age and age squared variables as continuous variables and tests individually the impact of career path variables and time dummies.

The endogeneity issue

Career stage variables and laboratory variables are likely endogenous. We have made several econometric attempts to address this issue that are detailed in the next section of the paper. However, the magnitude of the endogeneity bias is hard to determine. We will show that when using regressions without laboratory variables the model's estimates are robust. Similarly, career stage variables appear to interact with age and time variables (likely because of the identification problem mentioned above), but other estimates are robust to alternative models with or without career stage effects (Tables II.4 and II.5). Although our results are non-conclusive, their discussion helps to shed light on the endogeneity issue thereby illustrating the usefulness of our econometric model.

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 ran the estimations on the hypothesis that the explanatory variables are strictly exogenous with respect to errors.

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

$$\begin{aligned} &E(y_{it} \mid X_{it}, Z_i) = \exp(\mu + Z_i\gamma + X_{it}\beta + \alpha_i) \\ &\textit{with}\ \ \mathrm{y}_{\mathrm{it}} \sim \textit{Poisson} \end{aligned} \qquad (1)$$

$$E(\alpha_i \mid X_{it}) \neq 0,\ E(\alpha_i \mid Z_i) = 0 \qquad (H1)$$

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 Hausman tests show that we cannot assume that individual effects are uncorrelated with the explanatory variables. To solve this problem of the correlated unobserved individual effects with the explanatory variables, we treat them as fixed effects in our estimation. Moreover, 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[10]. 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 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. Yet we believe that the descriptive power of the variable remains high.

A consistent two step estimation was therefore 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 :

$$y_{it} / \exp(X'_{it} \hat{\beta}) = \exp(\mu + Z'_i \gamma) + \varepsilon_{it} \qquad (2)$$

We obtain consistent estimates of γ and μ. We ran 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:

$$\log\ (y_{it}) = dum(y_{it} = 0) + \mu + Z_i \gamma + X'_{it} \beta + \delta_i + u_{it} \qquad (3)$$

[10] We would not have been obliged to make (H1) hypothesis had our data provided satisfactory instrumental variables - which it did not. Hausman and Taylor (1981) and Breusch, Mizon and Schmidt (1989) show that a model like ours with individual effects exogenous but correlated with the regressors and time-invariant variables can be estimated by the IV method using as intruments : [$W_I X1$, $W_I X2$, $B_I X1$, Z1], where X1 is a subset of variables X exogenous and uncorrelated with the individual effects, W_I is the « within » operator and B_I is the « between » operator, X2 are exogenous but correlated with the individual effects, and Z1 is a subset of variables Z exogenous and uncorrelated with the individual effects ; and with the number of X1 variables greater than the number of Z2 (exogenous but correlated with the individual effects) variables.

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 :

$$\log\ (y_{it}) - X'_{it}\hat{\beta} = dum(y_{it}=0) + \mu + Z'_i\gamma + \delta_i + u_{it} \qquad (4)$$

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

As mentioned, the career stage variables as well as the laboratory variables are likely endogenous and we consequently made several attempts to address this issue. In particular we assumed that our regressors were predetermined. During the first step in our estimation, we estimated the coefficients of the time-varying variables X by the GMM on model (1) previously quasi-differentiated (Crépon and Duguet 1997, Blundell et al. 2000). The estimates of the time-invariant variables Z should then also take into account the correlation with the individual effects. Yet the only instruments available are transformations (lags) of our variables, and age and tenure are not informative. In any case, it is likely that the dependent variable is sufficiently persistent to produce the weakness of our instruments. For these reasons no other results are presented in this paper to control for endogeneity, only those which concern regressions without carrier stage and laboratory variables. These results serve specifically for discussing individual variable coefficients and age estimates (*Table II.4 and II.5*).

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

II. Results

A. The determinants of publication

This section describes successively the impact of individual, career stage and laboratory variables on individual productivity assessed by the mean number of articles per year and per researcher. We first describe the results obtained with the full model where age and career stage variables are used as group dummies. We then focus on the coefficients of individual variables and age estimates to discuss regression results which use neither career stage nor laboratory variables.

1. Results for the full model

When looking at the **age** groups, the estimation suggests a quadratic relationship between the age of the scientists and the average number of their publications per year. According to the estimation, a researcher's productivity increases between the first and the third age group, that is before 50, and then declines after 51. More precisely, compared to the group of young researchers aged 26 to 38, the group of researchers aged 39 to 45 publish an average of 0.26 papers more per year, the group aged 46 to 50 publish 0.36 papers more, while the oldest group of researchers, aged 51 to 61, publish only 0.13 papers more per year.

The age effect is reinforced by a **time effect** according to which productivity increases with time for all 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 might suggest that the scientists are under increasing pressure to publish or that they have more opportunities to do so because of an increase in the number of scientific journals.

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

The **gender** effect is important. All other variables being equal, a woman publishes almost 0.9 papers less than a man on average per year according to our estimate. This result needs to be examined in more detail by doing a specific study on gender. If the estimates suggest that men are more productive than women, they tell us nothing about why this might be the case nor about the true abilities of women. And our estimation method does not allow us to distinguish the pure effect of being a woman from all the related unmeasured "sociological" facts (number of children, marital status, etc.) which impact as well on the value of the estimated coefficient.

The result of **pre-doctoral training** is puzzling. The scientists who were educated in a Grande Ecole publish 0.7 papers more than the others per year on average, according to the TS regression. But this figure amounts to 0.3 in the TOTAL estimation. In other words, if all other observable effects are held equal, and all unmeasured effects are held equal as well, scientists who have been to a "Grande Ecole" will publish 0.7 articles more each year than the other scientists. However, when unmeasured effects are not held equal, a negative correlation is found between them and "Grande Ecole" and only 0.3 more papers are produced per year. We postulate that the "prestige" of the "Grande Ecole" system in France is an important element in explaining this finding. Members of this system benefit from Alumni networks which facilitate collaboration, promotion and mobility and probably make it easier for them to publish. Of course, this hypothesis needs further investigation.

The TOTAL and TS estimations give a different picture of the **career stage variables** 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 papers more. Whereas in the TS estimation, the DR2 coefficient is not statistically significant (-0.1), and the DR1 coefficient is negative, the DR1 publishes on average 0.8 papers less than a CR. Reaching the DR1 grade has a negative impact on productivity according to the TS regression which 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 grade 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. This means that the impact on productivity of being DR1 diminishes as the number of years spent in the grade increases. Consequently, the incentives to publish appear to be lower with time for those researchers who have reached the higher status. Although one reason is no doubt an increased number of administrative tasks, this result holds all other variables being equal, in particular the **talent** among the DR1. The higher TOTAL estimates suggest that talent is positively correlated with tenure among DR1, which is what one would expect given that time in the grade obviously means early promotion, all other things being equal. Taking this correlation into account reduces the effect of tenure by almost two times: 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 can be observed for the **DR2**. A scientist who has been in that grade for 4 to 8 years publishes 0.3 papers less on average than a newly promoted DR2, and a DR2 of the third tenure group (>8 years as DR2) publishes 0.5 papers less on average than a newly promoted DR2 (TS estimates). Again, it appears that the incentive to publish becomes lower as time spent in a grade increases. But if we take into account the positive correlation of "talent" with tenure within the DR2 status, the effect of tenure as a 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 suggests that the individual heterogeneity effects are negatively correlated to tenure as a CR. A discouragement effect could be at play, small hope for promotion decreases propensity to publish.

Concerning the **laboratory** effects, the results are the following. An individual is likely to be more productive when the level of his colleagues' productivity in the laboratory is high. We found that in the case of a 10% rise in peer productivity, a scientist publishes an average of 0.27 papers more than the number he would have published otherwise (+10% on average). Collaboration with foreign laboratories also has a strong positive impact on individual productivity. When a laboratory has a 10% rise in the proportion of papers co-published with foreign scientists, members of that laboratory individually publish an average of 0.8 articles more per year (+30% on average). This result suggests that forming centers of excellence in public research is likely to induce a better productivity of the member scientists.

The **size** of a laboratory does not compensate for the peer effect. If a laboratory's size increases by 10%, a member of that laboratory will publish on average 0.09 papers less than the number he would have published otherwise (-3% approximately on average). Nevertheless, our estimates suggest that "talented" researchers are more likely to be affiliated with 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 influence of the **quality of a laboratory** variable (or colleague's productivity in terms of quality) is open to question. The peer effect in terms of quality is almost non-existent 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 of a laboratory, a scientist could be expected to publish on average 0.25 papers *less* than the number he would have published otherwise (-10% approximately on average). Perhaps this finding suggests a substitution effect between quantity and quality, in the specific sense that if a laboratory stresses the importance of quality, a member of that laboratroy is more likely to publish less but his papers will have greater impact. As we will see in the next two sections, the quality of an individual's production is positively influenced by the overall quality of the laboratory's production.

2. Further results on the individual and age effects

As we explained above, different regressions were run in order to check for identification and endogeneity biases when using career stage and laboratory variables on the one hand, and in order to attempt to precisely locate at what age productivity peaks on the other hand. In these regressions, continuous age and age-squared variables were used instead of age groups, and estimates of age, gender and education were made without career and laboratory variables. In such a model, the time dummies can no longer be correctly estimated, however, our interest was in the age estimates as we said previously. The results are in *Table II.4 and II.5*.

The first result is that the estimates of gender and education are robust to any changes in the specification. As a matter of fact, dropping the career stage and lab variables has no significant effects on the other coefficients. The lab variables also have no effects on the age estimates.

Now if we look at the productivity peak issue, both career stage variables and the way we enter time dummies matter. The full model and the model without career stage, time and lab variables (age only model) predicts quite precisely the peak location: the age-productivity relation is quadratic and peaks at 54. The regression without the career stage variables results in a lower estimation of this peak that occurs at 50. The regression without the time dummies estimates the location of the peak at 58. The curves illustrating the age/publication relation according to the TS estimation and conditionally on the other variables are represented on *graph II.1*. It includes the mean point of 2.7 papers at 44.6 years old. The "age only" model gives age estimates for which the closest are the estimates given by the model without time dummies. All in all, we would conclude that the full model is the more satisfactory for predicting the age period around the productivity peak but likely under-estimates a little the age effect at the beginning of the career. For this reason, dropping the career stage variables doesn't help to improve the estimate, whereas the way time dummies are entered into the model has an impact (see *graph II.1*).

We compared our TS results to the ones obtained by Levin and Stephan (1991) for 182 scientists in solid state and condensed matter physics 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 squared 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 career state of scientists and the count nature of the data which contain a high proportion of zeros.

Finally, in all specifications it appears that the estimates in the TOTAL are lower than in the TS. The correlation between unmeasured effects and age is positive. If unmeasured effects reflect ability, then our estimates do not confirm the more or less common belief that ability to do innovative work decrease with age.

B. *The determinants of the average quality of journal publications*

This section describes successively the impact of individual, career stage and laboratory variables on individual productivity assessed by the mean impact factors averaged per year and per researcher.

The results are presented in table II.2. The most significant impact on the average quality of journal publications is produced by the quality of the laboratory variable. In particular, individual characteristics and career stage effects appear to be less important then previously observed when the dependant variable was the mean number of articles.

The average quality of journal publications 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 an average of 0.3 citations less than journals used by the youngest researchers. As previously, in an effort to better qualify this finding, we present the results of the regressions in which the continuous age and age squared variables replaced the age groups, with and without the carrier stage and laboratory variables, focusing on the coefficients of individual variables and the age estimates. The TS regression for the full model shows a negligible increase in the impact factor between 26 and 37, from 2.52 to 2.66. After 37, the average impact factor declines slightly and is equal to 1.90 at 61. When the career stage variables are dropped, it has almost no effect on the age-productivity relation, except that the peak occurs earlier at 35. Dropping the lab variables has no significant impact. Dropping the time dummies changes the age-productivity relationship in a way similar to the "age only" model and sets the productivity peak at 41 (the peak is at 39 with the "age only" model). The curves illustrating the age/publication relationship according to the TS estimations and conditionally on the other variables are represented on *graph II.2*.

The **gender and "Grande Ecole"** variables are statistically significant and both influence in the same way 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. Women publish in journals that receive on average 0.10 citations less over two years than the journals in which men publish. Interestingly, however, a positive correlation exists between the gender variable "woman" and "the individual heterogeneity effects" that could possibly be interpreted as "personal motivation for accessing recognized journals".

The **status and tenure** variables have no statistically significant influence 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 more) 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 the average journal impact factor of the members of that laboratory by 0.58 within two years. The other laboratory effects are comparatively weak. A 10% increase in the **productivity of the laboratory** decreases the impact factor by 0.05, suggesting a substitution effect which is symmetrical to the one identified in the previous section. 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 laboratories has a high positive coefficient, which suggests that being in small laboratories might favour the production of higher quality publications. This could be related to the fact that small laboratories are often not equipped to do experimental work and therefore produce more theoretical papers, however, this interpretation needs additional study because the dummy for the Grenoble region is also positive and in this region, the infrastructure for experimentation is highly developed.

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

C. The determinants of citations

The results of the estimation are shown in *table II.3* [11].

It appears that **age** has a negative impact on the average number of citations per article, per scientist and per year (received within 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 of citations received.

[11] As mentionned, the citations data have been available for a reduced sample. To compare what is comparable, we ran the regression of the three productivity measures on the reduced sample. The results for the measure in

Being a **woman** has a negative effect on citations, almost four times higher than on the impact factor measure but almost two times smaller than the effect on publications. A woman gets 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 individual heterogeneity 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 positive effects on productivity that we previously interpreted as being due to "personal network" or "prestige".

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 citations more on average per paper per year than a CR scientist and a DR2 0.35 citations more (TOTAL estimates).

Concerning the **tenure** variables, the number of citations is negatively influenced by the time spent in the DR1 grade. The number of citations over a two year period for a paper by a DR1 in the tenure group 2 will be on average 0.6 citations less than the number of citations received by a paper from a newly promoted DR1 (tenure group 1), and as high as 1.2 for a DR1 in group 3. 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 laboratory's** publications on individual productivity. An increase of 10% in the quality of a laboratory's publications means that members of that laboratory will receive 0.3 citations more per article per year. The peer effect in terms of quantitative **productivity** has no statistically significant effect on the annual number of citations received in the TS regression and a negative effect in the TOTAL regression. As was the case for the impact factor measure, we might be observing a substitution of quantity for quality because the laboratories that are publishing the most are not the ones getting the highest number of citations. A strong impact of **international openness** can be noted: when the proportion of co-published work with foreign countries increases by 10%, laboratory members receive 0.8 citations more per article per year. Finally, the **size** effect is slightly positive and at an increasing rate (resp. marginal impacts: 0.053 and 0.018).

---

terms of articles and for the measure in terms of impact factor are the same than those which have been

III. Conclusion

This article has explored the differences in productivity among scientists in public research, both in terms of the number of articles and the quality of 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 research productivity: individual variables such as age, gender and education; variables such as status and tenure describing the incentive mechanisms at work in a scientific institution; and, finally, laboratory variables for assessing context and work environment.

Individual variables have a strong impact on productivity. The mean number of publications tends to decrease with age at the end of a career, however, this “life cycle effect” does not seem to have a negative impact on the average number of citations received per year. Gender also has an impact – women publish less than men and get cited less as well – so does the fact of being admitted to a highly selective pre-doctoral education program. That said, institutional factors can be just as important as individual factors. Being a member of a highly productive laboratory which co-publishes frequently in international networks clearly stimulates an individual’s propensity to publish. Finally, our results suggest that promotion might be an incentive to publish whereas long tenure in a grade, and especially in high grades, has a negative impact on productivity. Yet we must take into account the endogeneity of that variable to assess the causality.

The work presented here is currently being developed in several directions. One is to account for the fact that career stage and laboratory variables may be endogenous. discuss further the econometric issues of identification and of endogeneity biases, and the limits of the present exercise 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 gender issue needs a closer examination as well, in order to better isolate what is specifically feminine in scientific practice and to better understand how women become a part of a citation network, how they build a reputation for themselves and the extent to which their past performance impacts upon their promotion and institutional affiliations. Finally, our last direction of research is aimed at building an econometric model for assessing the influence of peer and laboratory context on individual scientific productivity.

described.

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.

LEHMAN H.C., 1953, *Age and Achievement*, Princeton, NJ: Princeton University Press.

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

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

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).

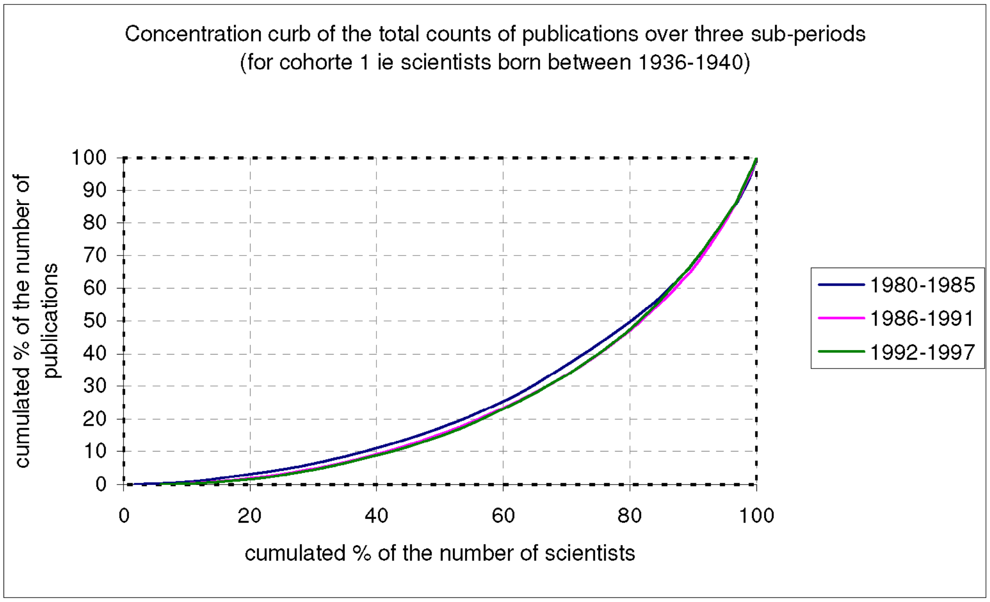
STEPHAN P., 1996, "The economics of science", Journal of Economic Literature, 31, september.

TURNER L., 2003, « La recherche publique dans la production des connaissances : contributions en économie de la science », PhD dissertation, Université Paris 1, http://www.crest.fr/pageperso/lei/laure.turner/these.htm.

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.

Figure 1



- *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.*

Figure 2

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 researchers' 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	Min	Max
Dependent variables							
Number of articles per year	2.69	3.21	2	0	4	0	62
Average impact factor per article per researcher and per year	2.66	2.30	2.54	0	3.8	0	21.48
Average number of citations (within 2 years) per article per researcher and per year	3.50	6.10	2	0	4.8	0	161
Extensions							
Average number of authors per article (harmonic)	3.23	2.57	3.33	0	4.90	1	25
Average number of pages per article	5.48	4.68	5.4	0	7.83	0	58
Individual variables + Time Dummies							
AGE	44.65	8.03	45	38	51	26	61
Age group 1 26<=age<=38	0.25	0.44	0	0	1	0	1
Age group 2 38<age<=45	0.25	0.43	0	0	0	0	1
Age group 3 45<age<51	0.23	0.42	0	0	0	0	1
Age group 4 51<=age<=61	0.27	0.44	0	0	1	0	1
Education in a "Grande Ecole"	0.17	0.38	0	0	0	0	1
Gender	0.18	0.39	0	0	0	0	1
More than one mobility	0.13	0.34	0	0	0	0	1
Career stage variables							
Status (CR_0)	0.59	0.49	0	1	1	0	1
Status (DR2_0)	0.08	0.28	0	0	0	0	1
Status (DR1_0)	0.33	0.47	0	0	1	0	1
Tenure in status CR group (0-10 years)	0.21	0.41	0	0	0	0	1
Tenure in status CR group 2 (11-18 years)	0.19	0.39	0	0	0	0	1
Tenure in status CR group 3 (>18 years)	0.18	0.39	0	0	0	0	1
Tenure in status DR2 group 1 (0-3 years)	0.13	0.33	0	0	0	0	1
Tenure in status DR2 group 2 (4-8 years)	0.10	0.3	0	0	0	0	1
Tenure in status DR2 group 3 (>8 years)	0.10	0.3	0	0	0	0	1
Tenure in status DR1 group 1 (0-1 year)	0.03	0.18	0	0	0	0	1
Tenure in status DR1 group 2 (2-5 years)	0.02	0.15	0	0	0	0	1
Tenure in status DR1 group 3 (>5 years)	0.03	0.16	0	0	0	0	1
Laboratory variables							
Size of the laboratory (number of researchers)	46.44	26.34	26	43	60	3	98
Productivity of the laboratory in logarithm	0.72	0.42	0.79	0.44	0.99	-1,1	1.69
Quality of the laboratory in logarithm	1.09	0.45	1.27	1.09	1.33	0	1.67
Proportion of the laboratory articles with foreign co-authors	0.04	0.03	0.04	0.02	0.06	0.11	7.59
Dummy for the Grenoble region	0.26	0.44	0	0	1	0	1
Dummy for the Paris region	0.36	0.48	0	0	1	0	1
Dummy for laboratory with less than 3 researchers	0.14	0.34	0	0	0	0	1

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

Table II.1

- *Number of individuals = 465, Number of years = 12, Number of Observation = 5580*
- *Dependant variable: Number of articles per year per researcher*

	POISSON ON Number of articles per year per scientist (ART)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE group 2 (38<age<=45)	0.205*** (0.025)	0.098*** (0.029)	0.553	0.263
AGE group 3 (45<age<51)	0.233*** (0.025)	0.137*** (0.032)	0.628	0.368
AGE group 4 (51<=age<=61)	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
Education in a “Grande Ecole”	0.118*** (0.021)	0.26*** (0.043)	0.317	0.701
Dummy More than one mobility	0.041* (0.025)	-0.024 (0.051)	0.11	-0.064
Status (DR1_0)	0.543*** (0.041)	-0.287*** (0.096)	1.463	-0.772
Status (DR2_0)	0.174*** (0.031)	-0.045 (0.06)	0.47	-0.121
Tenure in status CR group 2 (11-18 years)	-0.094*** (0.029)	0.028 (0.054)	-0.252	0.077
Tenure in status CR group 3 (>18 years)	-0.291*** (0.032)	0.054 (0.079)	-0.786	0.145
Tenure in status DR2 group 2 (4-8 years)	0.012 (0.031)	-0.129*** (0.042)	0.031	-0.347
Tenure in status DR2 group 3 (>8 years)	-0.029 (0.034)	-0.2*** (0.062)	-0.078	-0.539
Tenure in status DR1 group 2 (2-5 years)	-0.05 (0.057)	0.023 (0.07)	-0.134	0.063
Tenure in status DR1 group 3 (>5 years)	-0.287*** (0.058)	-0.494*** (0.091)	-0.774	-1.33
Proportion of the laboratory articles with foreign co-authors	2.942*** (0.432)	3.01*** (1.07)	7.928	8.112
Size of the laboratory in logarithm	-0.131*** (0.015)	-0.091** (0.036)	-0.131	-0.091
Size of the laboratory in logarithm squarred	0.016*** (0.006)	0.016 (0.011)	0.016	0.016
Productivity of the laboratory in logarithm	0.233*** (0.032)	0.274*** (0.067)	0.233	0.274
Quality of the laboratory in logarithm	-0.254*** (0.069)	0.018 (0.144)	-0.254	0.018
Dummy for the Grenoble region	0.173*** (0.026)	0.179*** (0.053)	0.465	0.483
Dummy for the Paris region	-0.055** (0.026)	0.009 (0.047)	-0.147	0.026
Dummy for laboratory with less than 3 researchers	-0.66*** (0.099)	-0.109 (0.202)	-1.78	-0.293

Dummy 1987	0.117*** (0.046)	0.114** (0.046)	0.315	0.307
Dummy 1988	0.203*** (0.045)	0.219*** (0.045)	0.548	0.589
Dummy 1989	0.336*** (0.043)	0.362*** (0.044)	0.905	0.974
Dummy 1990	0.191*** (0.045)	0.254*** (0.046)	0.514	0.684
Dummy 1991	0.227*** (0.044)	0.329*** (0.046)	0.612	0.885
Dummy 1992	0.262*** (0.044)	0.393*** (0.046)	0.707	1.06
Dummy 1993	0.457*** (0.042)	0.599*** (0.045)	1.232	1.615
Dummy 1994	0.388*** (0.043)	0.55*** (0.046)	1.046	1.481
Dummy 1995	0.297*** (0.044)	0.482*** (0.048)	0.8	1.298
Dummy 1996	0.393*** (0.043)	0.583*** (0.047)	1.059	1.572
Dummy 1997	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-likelihood 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

- *Number of individuals = 465, Number of years = 12, Number of Observation = 5580*
- *Dependant variable: Average impact factor per article per researcher and per year*

	LOGLINEAR ON Average impact factor per article per researcher and per year (NOT_I)		MARGINAL IMPACTS	
Variables	Total	Within	Total	Within
AGE group 2 (38<age<=45)	-0.041* (0.021)	-0.002 (0.029)	-0.109	-0.004
AGE group 3 (45<age<51)	-0.104*** (0.026)	-0.06 (0.042)	-0.276	-0.16
AGE group 4 (51<=age<=61)	-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
Education in a "Grande Ecole"	0.032* (0.017)	0.042** (0.016)	0.085	0.111
Dummy More than one mobility	-0.012 (0.019)	-0.01 (0.019)	-0.031	-0.026
Status (DR1_0)	0.135*** (0.042)	-0.054 (0.074)	0.36	-0.143
Status (DR2_0)	0.058** (0.026)	-0.041 (0.046)	0.155	-0.109
Tenure in status CR group 2 (11-18 years)	-0.004 (0.02)	-0.02 (0.04)	-0.01	-0.054
Tenure in status CR group 3 (>18 years)	0.015 (0.028)	-0.011 (0.056)	0.041	-0.029
Tenure in status DR2 group 2 (4-8 years)	0.041 (0.026)	-0.021 (0.031)	0.11	-0.055
Tenure in status DR2 group 3 (>8 years)	0.063** (0.029)	0.039 (0.046)	0.166	0.105
Tenure in status DR1 group 2 (2-5 years)	0.033 (0.052)	0.025 (0.059)	0.089	0.066
Tenure in status DR1 group 3 (>5 years)	0.006 (0.05)	0.069 (0.077)	0.016	0.184
Proportion of the laboratory articles with foreign co-authors	0.031 (0.326)	0.121 (0.327)	0.082	0.322
Size of the laboratory in logarithm	-0.033*** (0.012)	-0.025** (0.012)	-0.033	-0.025
Size of the laboratory in logarithm squarred	0.021*** (0.004)	0.024*** (0.004)	0.021	0.024
Productivity of the laboratory in logarithm	-0.046* (0.024)	-0.049** (0.024)	-0.046	-0.049
Quality of the laboratory in logarithm	0.57*** (0.051)	0.579*** (0.05)	0.57	0.579
Dummy for the Grenoble region	0.029 (0.019)	0.044** (0.019)	0.077	0.118
Dummy for the Paris region	0.009 (0.018)	0.025 (0.018)	0.023	0.067
Dummy for laboratory with less than 3 researchers	0.501*** (0.074)	0.551*** (0.073)	1.335	1.467

Dummy 1987	-0.051* (0.03)	-0.051* (0.028)	-0.135	-0.135
Dummy 1988	-0.123*** (0.03)	-0.122*** (0.028)	-0.327	-0.325
Dummy 1989	-0.109*** (0.03)	-0.103*** (0.029)	-0.29	-0.274
Dummy 1990	-0.1*** (0.03)	-0.085*** (0.03)	-0.265	-0.227
Dummy 1991	-0.123*** (0.03)	-0.104*** (0.031)	-0.328	-0.277
Dummy 1992	-0.076** (0.03)	-0.054* (0.032)	-0.203	-0.143
Dummy 1993	-0.191*** (0.031)	-0.167*** (0.033)	-0.509	-0.444
Dummy 1994	-0.087*** (0.031)	-0.065* (0.034)	-0.232	-0.172
Dummy 1995	-0.057* (0.031)	-0.03 (0.035)	-0.151	-0.08
Dummy 1996	-0.036 (0.032)	-0.007 (0.037)	-0.095	-0.019
Dummy 1997	-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

- *Number of individuals = 352, Number of years = 9, Number of Observation = 3168*
- *Dependant variable: Average number of citations (within 2 years) per article per researcher and per year*

	LOGLINEAR ON Average number of citations (within 2 years) per article per researcher and per year (MCIT2)		*MARGINAL IMPACTS*	
Variables	TOTAL	TWO STEP	TOTAL	TWO STEP
AGE group 2 (38<agc<=45)	-0.022 (0.036)	0.006 (0.05)	-0.077	0.023
AGE group 3 (45<age<51)	-0.103** (0.043)	-0.028 (0.073)	-0.357	-0.097
AGE group 4 (51<=age<=61)	-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
Education in a “Grande Ecole”	0.062* (0.033)	0.154*** (0.033)	0.215	0.536
Dummy More than one mobility	0.078** (0.033)	0.083** (0.034)	0.27	0.29
Status (DR1_0)	0.306*** (0.084)	-0.164 (0.134)	1.065	-0.571
Status (DR2_0)	0.101** (0.045)	-0.057 (0.067)	0.352	-0.197
Tenure in status CR group 2 (11-18 years)	-0.054 (0.038)	0.015 (0.052)	-0.187	0.052
Tenure in status CR group 3 (>18 years)	-0.011 (0.05)	0.016 (0.078)	-0.038	0.054
Tenure in status DR2 group 2 (4-8 years)	-0.022 (0.044)	-0.13** (0.053)	-0.077	-0.451
Tenure in status DR2 group 3 (>8 years)	0.023 (0.048)	-0.072 (0.083)	0.081	-0.251
Tenure in status DR1 group 2 (2-5 years)	-0.051 (0.096)	-0.183** (0.093)	-0.177	-0.638
Tenure in status DR1 group 3 (>5 years)	-0.01 (0.097)	-0.355** (0.148)	-0.034	-1.234
Proportion of the laboratory articles with foreign co-authors	1.756*** (0.597)	1.944*** (0.62)	6.109	6.766
Size of the laboratory in logarithm	0.011 (0.021)	0.053** (0.022)	0.011	0.053
Size of the laboratory in logarithm squarred	0.015** (0.007)	0.018** (0.007)	0.015	0.018
Productivity of the laboratory in logarithm	-0.104** (0.043)	-0.056 (0.044)	-0.104	-0.056
Quality of the laboratory in logarithm	0.28*** (0.09)	0.301*** (0.093)	0.28	0.301
Dummy for the Grenoble region	0.128*** (0.035)	0.141*** (0.037)	0.446	0.489
Dummy for the Paris region	0.031 (0.033)	0.082** (0.035)	0.107	0.284
Dummy for laboratory with less than 3 researchers	0.29** (0.13)	0.544*** (0.133)	1.009	1.894

Dummy 1987	0.042 (0.046)	0.046 (0.043)	0.147	0.16
Dummy 1988	0.0001 (0.046)	0.008 (0.044)	0.0002	0.028
Dummy 1989	0.01 (0.046)	0.022 (0.046)	0.036	0.077
Dummy 1990	0.006 (0.047)	0.027 (0.05)	0.019	0.095
Dummy 1991	0.0004 (0.047)	0.032 (0.053)	-0.001	0.11
Dummy 1992	0.01 (0.048)	0.056 (0.057)	0.034	0.195
Dummy 1993	-0.349*** (0.048)	-0.309*** (0.061)	-1.215	-1.073
Dummy 1994	-0.209*** (0.049)	-0.17*** (0.065)	-0.729	-0.593
DUMMY (MCIT2=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*

Table II.4

- *Number of individuals = 465. Number of years = 12. Number of Observation = 5580*
- *Dependant variable: Number of articles per year per researcher*

Full model

	POISSON ON Number of articles per year per scientist (ART)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.019***	0.028***	-0.051	0.075
AGE²	-0.0021***	-0.0015***	-0.006	-0.004
WOMAN	-0.280***	-0.307***	-0.753	-0.826
"Grande Ecole"	0.081***	0.376***	0.218	1.011

Without the stage variables

	POISSON ON Number of articles per year per scientist (ART)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.0012	0.020***	-0.003	0.054
AGE²	-0.0019***	-0.0018***	-0.005	-0.005
WOMAN	-0.331***	-0.317***	-0.890	-0.853
"Grande Ecole"	0.180***	0.296***	0.484	0.796

Without the time variables

	POISSON ON Number of articles per year per scientist (ART)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.0046*	0.045***	-0.012	0.121
AGE²	-0.002***	-0.0018***	-0.005	-0.005
WOMAN	-0.268***	-0.283***	-0.721	-0.761
"Grande Ecole"	0.109***	0.445***	0.293	1.197

Without the lab variables

	POISSON ON Number of articles per year per scientist (ART)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.019***	0.028***	-0.051	0.075
AGE²	-0.0021***	-0.0015***	-0.006	-0.004
WOMAN	-0.249***	-0.323***	-0.670	-0.869
"Grande Ecole"	-0.099***	0.386***	-0.266	1.038

Without the career stage, time, and lab variables (age only)

	POISSON ON Number of articles per year per scientist (ART)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	0.006***	0.037***	0.016	0.100
AGE²	-0.0017***	-0.0019***	-0.005	-0.005
WOMAN	-0.295***	-0.276***	-0.794	-0.742
"Grande Ecole"	0.225***	0.368***	0.605	0.990

Table II.5

- *Number of individuals = 465. Number of years = 12. Number of Observation = 5580*
- *Dependant variable: Average impact factor per article per researcher and per year*

Full model

	LOGLINEAR ON Average impact factor per article per researcher and per year (NOT_I)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.008***	-0.008***	-0.021	-0.021
AGE²	-0.00029***	-0.0005***	-0.001	-0.001
WOMAN	-0.022	-0.042*	-0.059	-0.112
"Grande Ecole"	0.031*	0.032**	0.082	0.085

Without the stage variables

	LOGLINEAR ON Average impact factor per article per researcher and per year (NOT_I)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.003***	-0.008***	-0.008	-0.021
AGE²	-0.00019**	-0.0004***	-0.001	-0.001
WOMAN	-0.031*	-0.040*	-0.082	-0.106
"Grande Ecole"	0.050***	0.035**	0.133	0.093

Without the time variables

	LOGLINEAR ON Average impact factor per article per researcher and per year (NOT_I)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.007***	-0.002***	-0.019	-0.005
AGE²	-0.00018*	-0.0003***	0.000	-0.001
WOMAN	-0.022	-0.038*	-0.059	-0.101
"Grande Ecole"	0.032**	0.053***	0.085	0.141

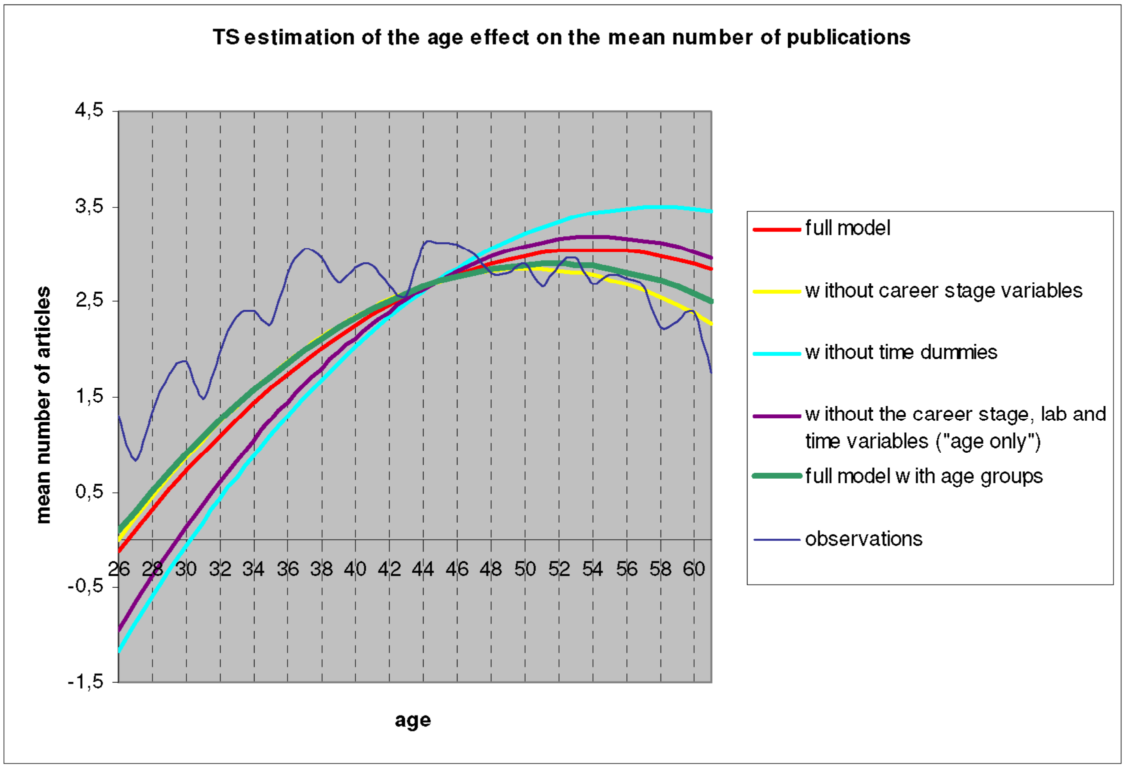
Without the lab variables

	LOGLINEAR ON Average impact factor per article per researcher and per year (NOT_I)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.008***	-0.008***	-0.021	-0.021
AGE²	-0.0003***	-0.0005***	-0.001	-0.001
WOMAN	-0.043***	-0.066***	-0.114	-0.176
"Grande Ecole"	0.037**	0.042*	0.098	0.112

Without the career stage, time and lab variables (age only)

	LOGLINEAR ON Average impact factor per article per researcher and per year (NOT_I)		MARGINAL IMPACTS	
Variables	TOTAL	Two Step	TOTAL	Two Step
AGE	-0.003***	-0.003*	-0.008	-0.008
AGE²	-0.0001	-0.00025**	0.000	-0.001
WOMAN	-0.054***	-0.058***	-0.144	-0.154
"Grande Ecole"	0.058***	0.060***	0.154	0.160

Graph II.1



Graph II.2