3-2007

The Determinants of Research Productivity: A Study of Mexican Researchers

Claudia Gonzalez-Brambila
Carnegie Mellon University

Francisco Veloso
Carnegie Mellon University, fveloso@cmu.edu

Published In
Research Policy, 36, 7, 1035-1051.

The Determinants of Research Productivity:

A Study of Mexican Researchers

Claudia Gonzalez-Brambila[a] and Francisco Veloso[b]

Department of Engineering and Public Policy, Carnegie Mellon University

Abstract

This paper uses a unique data set of Mexican researchers to explore the determinants of research productivity. Our findings confirm a quadratic relationship between age and productivity. However, productivity peaks when researchers are approximately 53 years old, 5 or 10 years later than what prior studies have shown. These results suggest that age is not very important in terms of research productivity. We also find that reputation matters for the number of citations but not publications. Results also show great heterogeneity across areas of knowledge. Interpretations of other aspects, such as gender, country of PhD, cohort effect, among others, are also discussed.

Key words: research productivity, life cycle, economics of science.

[a] Claudia Gonzalez-Brambila, Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, 129 Baker Hall, Pittsburgh, PA, 15213-3890, USA. Phone: + 412 268 5617. Fax: + 412 268 3757. E-mail: cgonzale@andrew.cmu.edu

[b] Corresponding author: Francisco Veloso. Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, 129 Baker Hall. Pittsburgh, PA 15213-3890, USA. Phone: + 412 268 4640. Fax. + 412 268 3757. E-mail: fveloso@cmu.edu

Introduction

The amount of funding for science and technology (S&T) has increased enormously since Vannevar Bush's manifesto, "The Endless Frontier" (Bush, 1945). In this seminal policy document, science is seen as the source of new knowledge that is published openly and drives the creation of new technologies that lead to economic growth. However, the last decades have shown that the social payoff for science and technology is neither direct nor clear. As a result, a necessity to asses the impact of S&T programs has grown. This has shaped a culture of evaluation and monitoring in research, where publications and citations are the most common ways to measure the contributions of a researcher or an organization.

Research output is evaluated and monitored at different levels and for different purposes. At the macro-level, governments have opted for increasing "project" funding of research, usually allocated on a competitive basis, in detriment of "institutional" funding[1] (OECD, 2002). To have access to this funding, researchers submit research proposals to a funding entity, which uses external referees and peer review committees to decide which are promising proposals that ought to be supported. In this evaluation process there is evidence that shows that past publications have an important effect on the expected level of grant funding (Arora et al., 1998 and Arora and Gambardella, 1998). At the micro-level, universities and research centers use publication and citation counts to monitor the

[1] Institutional funding refers to block funds that are allocated to universities or research centers on an annual basis and do not have strings.

performance of their researchers and to give raises and promotions. Moreover, in ranking universities' departments, one of the most important measures is the aggregate number of publications and citations of their faculty. Publications are also important as a channel of communication with industry. Arundel and Geuna (2001) found that publications review is the method of obtaining the results of public research most frequently cited as important among firms. Companies also use publications as a way of detecting expertise within universities, with subsequent hiring of faculty and graduates as consultants or employees.

Despite the increasing importance of publications and citations as a measure of research productivity, we know surprisingly little about the determinants of individual and organizational research productivity. Understanding productivity determinants can be of critical importance for administrators of universities and research laboratories, public or private. Being able to estimate expected productivity of researchers, taking into account individual characteristics, past history, and institutional variables, can help design policies to enhance productivity, or can plan for a balance in groups to compensate for the potential existence of age, cohort or other effects. It can be particularly relevant for policy makers in countries where most of the research system is financed with public funds, helping to design policies that enhance individual and institutional productivity.

In the pursuit for more in-depth knowledge of the factors that condition research productivity, it is crucial to analyze a diverse set of countries because of the different characteristics of the S&T community around the world, especially between the developed and the less developed world (Nelson, 1993). This aspect has been overlooked by most existing research on the topic, which has focused on the developed world. For example, Mexico, as an advanced developing country, has a relatively small scientific community, with 0.6 researchers per 1,000 employees, while France has 6, and the U.S. 8.1 (Conacyt, 2002). In Mexico, S&T expenditure is 0.4% of the Gross Domestic Product (GDP), against 2.76% in the U.S. and 2.15% in France (Conacyt, 2002). These figures reflect limited availability of research resources. Yet another important difference between countries is access to some journals, because of their selectivity or publication policies (Lawrence, 2003). Furthermore, the scientific community in many developing countries faces disadvantages due to differences in language and publication culture.

Another dimension that needs careful consideration in studies aiming at comparing the performance of researchers is the area of knowledge. A number of existing studies have documented significant heterogeneity in the pattern of publications across areas (Cole, 1979; Xie and Shauman 1998). Nevertheless, as with the contrast between countries, previous work has not made a systematic effort to explore how scientific area conditions research productivity.

This paper explores the determinants of research productivity using a data base of the most productive researchers in Mexico. Our main interest is to analyze the dynamics of productivity over the life cycle, and to explore differences among areas of knowledge. To the best of our knowledge, this is the most comprehensive study of its kind. First, the database is very complete, not only for the number of researchers in the sample, but also because it considers all areas of knowledge, researchers across an entire nation, and detailed information about each researcher, as well as a very long time span. Second, the study addresses how past success influences future performance, an issue hardly considered in previous studies but deemed of critical importance for the dynamics of productivity (Stephan, 1996). Third, it is the first study to look at those issues outside the developed world, focusing on an advanced developing economy, where the improvement of research productivity is a major policy issue (Conacyt, 2001).

This paper is organized in five sections. The first section presents a review of existing studies that address the determinants of productivity. Section 2 describes the data used in the study. In section 3 we formally present the model that is used to estimate the effects of the explanatory variables in international publications. Results are presented in Section 4, while conclusions and suggestions for future research are discussed in Section 5.

1. Literature Review

The dynamics of research productivity over the life cycle is one of the aspects of individual productivity that has received more attention. Gary Becker (1962) and Theodore Schultz (1963) are the pioneers exploring how life cycle may condition productivity in occupations where human capital plays an important role. Their models suggest that human capital investment declines over time due to the finiteness of life, which would lead individual productivity to follow an inverted U shape pattern (Stephan, 1996). Subsequently, a number of researchers has tried to test for such effects in different professional contexts, including the profession of researcher.

Using a sample of American chemical engineering departments, Bernier et al. (1975) find that publications and citations peak for those in the 40-44 age group. The study also correlates measures of quality, including number of citations, number of PhD's graduated, funds and peer evaluations of researcher quality, with publication volume. Overall, they find that quantity and quality are highly positively related. Analyzing a cross-section of American researchers from six different fields, Cole (1979) finds that age has a slight curvilinear relationship with both quality and quantity of scientific productivity. He suggests that the small increase in productivity through the thirties and a corresponding limited decrease over the fifties is explained by the scientific reward system. Since the reward system reduces the number of scientists who are actively publishing, those who continue publishing are the best members of their cohort.

One limitation of these earlier studies is the use of cross-sectional data, which make it possible that aging effects can be confounded with cohort effects (Stephan, 1996). For example, knowledge base is one of these potential cohort effects. If one assumes that in science the latest educated are the best educated (Levin and Stephan, 1991) and, as a result, better prepared to publish, regressions based on cross sections would attribute potential cohort effects to the younger age of researchers, biasing their productivity upwards. Another type of cohort effect is associated with cumulative advantage, where past publications are an important factor conditioning access to research resources. This would lead more experienced researchers to be the more productive, but not because of age but rather due to access to resources. This is especially important given the stratified distribution of the reward system in science discussed in classical papers by Merton (1968 and 1988), what he calls “The Matthew Effect in Science.” The Matthew effect consists “in the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark” (Merton, 1968). This increases even more the skew of publications and mainly of citations, underlining that beyond age, there are other critical factors at play in the dynamics of research productivity.

To counter these problems, more recent studies use longitudinal data on individual researchers. For example, studying the publishing activity of Berkeley mathematicians, Diamond (1986) finds a slight decline with age. Using a pooled model to estimate the rate of publications for 1000 Israeli scientists, Weiss and Lillard (1982), find a quadratic relation between the number of publications and the phase of the academic career.

One of the most important studies is Levin and Stephan (1991). They develop a model of scientific productivity that considers that scientists engage in research not only because of the future financial rewards associated with it, but also for the satisfaction of “solving the puzzle”. Using longitudinal data of American scientists, they find that life cycles effects are present in five of the six areas of physics and earth sciences studied. In their model B, which considers straight publication counts, the solid state and condensed matter physicists increase their publication rate until reaching a peak of 2 papers per year at age

45, declining after that. However, there are important variations across areas; the atomic and molecular physicists reach a peak at age 39 and geophysicists at 59. Adjusting for co-authorship, journal quality, or both generally causes the age peak to be reduced by 1 to 5 years. It is important to mention that they do not find strong evidence that the latest educated are the most productive.

In a very recent study, Tuner and Mairesse (2003) analyze the impact of research productivity relative to age, gender and education of French condensed matter physicists. They find that there is a quadratic relation between the age of the scientists and the number of publications, with researchers' productivity increasing before 50 and then declining after 51. The results using citations are not significantly different from those obtained with publications. Finally, the results suggest that graduates from the French Grande Ecoles publish more and that a woman publishes on average almost 0.9 papers less than a man per year.

This last aspect of measuring sex differences in scientific productivity has also captured the attention of researchers. Several studies have found that female scientists publish at lower rates than male scientists. Using a sample of American biochemists, Long (1992) finds that sex differences in the number of publications and citations are bigger during the first decade of the career but are reversed later. He attributes the lower productivity of females to their overrepresentation among non-publishers and their underrepresentation among the extremely productive. In a more recent study, Xie and Shauman (1998) find that sex differences in research productivity have declined, with the female-to-male ratio increasing from about 60 % in the late 1960’s to 75 to 80% in the late 1980’s and early 1990’s. Their research also uses a sample of American scientists.

A slightly different perspective on understanding productivity dynamics is presented by Allison and Stewart (1974). Instead of looking at productivity over the life cycle, they use a cross-section of chemists, physicists, and mathematicians in the U.S., and find that the highly skewed distributions of productivity among researchers can be explained by a process of cumulative advantage. Highly productive researchers maintain or increase their productivity because they receive recognition and resources, while those scientists who do not, become less productive or have to leave their career as researchers. This inequality becomes increasingly pronounced as career age increases. In an extension of this study, Allison et al. (1982) examine cohorts of biochemists and chemists, and they confirm that increasing inequality is observed for counts of publications but not for counts of citations to all previous publications, suggesting that scientist’ older publications are cited with less inequality than their more recent work.

Buchmueller et al. (1999) develop a productivity model function where publications depend on the graduate school, the type of the first job after graduating, some personal characteristics, and unobserved factors. They empirically test their model using data of American PhD economists. Their findings suggest that graduates of the top ten programs who had some experience as research assistants are more productive. When they take into account the initial job, they find that economists employed in research universities are also more productive.

Table 1 presents a summary of the determinants of scientific productivity that have been studied and the key findings.

- Table 1 -

The methodologies, size of the samples, period of time of the study, as well as limitations vary greatly among these studies. As mentioned above, one of the most important limitations of the early work was the use of cross sectional data, since it does not allow separating age and cohort effects. Moreover, most existing studies analyze cohorts before the 1990's and the dynamics of publishing activity have changed dramatically during the last twenty years. Among other factors, the pressure to publish and the number of journals have increased substantially during this period. In addition, some of the longitudinal studies analyze publications in a short period of time (between 2 and 6 years), so some conclusions could be weak and inconclusive because of the limited data. The source of publications also varies among the studies. Some studies use the field abstracts, and others use self reported number of publications. Furthermore, as was mention before, all of these studies focus on the developed world.

This paper tries to correct for most of the critical limitations of previous work, while expanding into new directions. First, we are using a recent and extensive panel, with close to 15,000 researchers and publication data for over 20 years. These data includes all areas of knowledge and detailed information about researchers across an entire nation.

This enables better controls and a more robust estimation of the critical aspects already identified by previous authors, especially the dynamics of productivity over the life cycle, as well as a comparison among different research areas of knowledge. Our analysis looks at differences in scientific productivity considering gender, country where PhD was earned, cohort and total number of researchers and publications in the same area of knowledge. The study also addresses how past success influences future performance, a critical issue for the dynamics of productivity. Finally, it is the first research to look at scientific productivity outside the developed world, focusing on Mexico, an advanced developing economy.

2. The data

We had access to information on 14 328 researchers, in all fields of knowledge, who have been part of the Mexican National System of Researchers (SNI), for at least one year, from 1991 to 2002. The National System of Researchers was created in 1984 to enhance the quality and productivity of researchers in Mexico. It gives pecuniary compensation, as a complement of salary, to the most productive researchers[2]. SNI grants represent on average 30% of the income of researchers in the program. In 2001, 31% of researchers in Mexico were in SNI (Conacyt, 2002). It is important to note that the researchers in SNI

[2] To be included in the system, researchers apply to the program, then are evaluated by peer committees that determine which researchers can receive the benefits of the program and the appropriate level. The program ranks researchers in 4 levels, with compensations depending on level (Candidate, Level 1, 2, and 3, where Level 3 is the highest).

are not randomly distributed and do not represent the whole population of researchers in Mexico. On the contrary, given the characteristics of the program, it is expected that the most productive researchers are in the system.

The data are classified in two categories:

1. Characteristics of the researchers:
    - Gender
    - Institution
    - Age
    - Years since PhD was earned
    - Country where PhD was earned
    - Area and discipline[3]
    - Papers published and citations

2. SNI variables:
    - The researchers' presence (or absence) and level in system for each year
    - Budget and lagged budgets of Conacyt[4]
    - Total number of researchers in SNI by area and discipline
    - Total number of publications by SNI researchers by area and discipline.

[3] We use a broad classification of fields of knowledge as area. In this paper, we are using 6 different areas: Exact Sciences, Biology and Chemistry, Health Sciences, Social and Humanities, Agricultural Sciences and Biotechnology, and Engineering. We refer to discipline as a field of knowledge; for example, Exact Sciences includes Mathematics, Physics, Earth Sciences, Astronomy, and Material Science.

[4] Conacyt is the National Council for Science and Technology in Mexico. This agency is in charge of the science and technology policy in Mexico and manages the SNI program among other programs to support science and technology activities.

Our measures of productivity are publications and citations in the Science and Social Sciences Citation Indexes, produced by the Institute of Scientific Information (ISI). The publications were obtained by matching the data base of the researchers in SNI with Mexican articles from the ISI database from 1981-2002. We also have access to all the citations that were made to each of the papers until 2002 (ISI, 2003).

The mean of publications per year for the entire population is 0.174 with standard deviations of 0.62 (overall), 0.32 (between) and 0.52 (within) [N= 14 328][5]. The distribution of publications is highly skewed. The lower extreme of productivity is represented by those who never have published [n=5900][6]. Lack of publication may represent at least three types of activity: people may no longer be active in science; people are devoting most of their time in other types of activity such as teaching or

[5] The statistics for the entire population are: 73% are males ; 95% are in public institutions; the distribution by level is the following: 7% in level 3, 15% in level 2, 54% in level 1 and 24% are candidates; 49% got their PhD in Mexico, 17% in the United States, and 30% in Europe. The average age is 41.11 years old (std. dev. 10 years). Related to areas of knowledge, the largest area is Social and Humanities with 25%, followed by Biology and Chemistry with 17%, and Exact Sciences, Agricultural Sciences and Biotechnology, and Engineering with 16% each; finally Health Sciences has the smallest number of researchers with 10%.

[6] This subsample has similar distribution to the entire population. The average age is 41.34 years (std dev 10.1), 72% are males [there is no overrepresentation of nonproductive women as in the case of the Long study], 94% are in public institutions, 4% are in level 3, 11% are in level 2, 53% are in level 1, and 32% are candidates; 17% got their PhD in Mexico, 17% in the United States and 34% in Europe. 7% are in Exact Sciences, 7% in Biology and Chemistry, 3% in Health Sciences, 45% in Social and Humanities, 21% in Agricultural and Biotechnology and 17% in Engineering.

consulting; it may also mean that they have other types of output such as books or publications in journals not in ISI. Because the main purpose of this study is the analysis of the dynamics of paper productivity (and respective citations) over the life cycle, we limited our study to those researchers who have at least one publication in ISI[7]. Given that we do not have information on those researchers who were in the system before 1991, we are considering eleven years of publications from 1991 to 2001. As a result, our final baseline sample consists of 7793 researchers in all disciplines.

- Table 2 -

The mean of publications per year of this subsample is 0.489 with standard deviation of 0.965. The average age (40.3 years old) is slightly lower than that of the entire population (41.11). The proportion of men (73%) is the same as the entire population. We have 51% of researchers who got their PhD in Mexico, while 17% did so in the United States and 27% in Europe[8]. In addition, most of the researchers in our sample got their PhD after 1992 (56%), 28% got their PhD between 1982 and 1991, 10% between 1973 and 1981, and only 5% before 1972[9]. Biology and Chemistry is the area of knowledge that concentrates the largest number of researchers (24%), followed by Exact Sciences that has 22% of the researchers. On the other hand, Social and Humanities is the smallest area with 10%. Table 2 shows descriptive statistics for the variables used in the model.

[7] The bias that is caused because of this situation affects the peak of publications but not the shape of the curve.

[8] There is 27% of missing data, so it is possible to have a bias in the estimation of this parameter.

[9] There is 45% of missing data, so it is possible to have a bias in the estimation of this parameter.

Figure 1 presents the mean of publications per year for the entire population, from 1981 to 2001, showing the differences among areas of knowledge. As it can be seen, there has been an enormous increase in the mean of publications by Mexican researchers in SNI over the period studied, especially after 1990. Health Sciences is the area of knowledge that concentrates the most productive researchers, followed by Exact Sciences, and Biology and Chemistry. On the other hand, Social and Humanities is the area with less productive researchers.

- Figure 1-

3. The model

To study the dynamics of productivity over the life cycle, we will assume that the function determining publishing proficiency P_{it} is given by:

$P_{it} = F(X_{it}, Z_i, c_i, u_{it})$, i identifies researchers and t year.

Z_i: Variables that are stable across time but not across researchers:

area of knowledge, gender, institution, country of PhD, cohort.

X_{it}: Variables that vary in both dimensions:

age, age^2, level, stock of past publications, budget, total number of researchers in the same area, total number of publications in the same area.

c_i: Individual unobserved effect which is stable across time but not across researchers.

u_{it}; Unobserved effect that varies in both dimensions.

We use the Negative Binomial fixed effects model proposed by Hausman, et al. (1984) because of the panel nature of our data. This model allows for both the possibility of permanent unobserved individual effect as well as the possibility that some unobserved effects may be correlated with publications and other explanatory variables. We chose the Negative Binomial distribution over the Poisson because the latter imposes a constant variance. This is not true for the data used in our study where the variance of productivity far exceeds the mean. One drawback of the Negative Binomial distribution is that our conclusions may be less precise since the estimated standard errors tend to be larger than in the alternative Poisson model. Another drawback comes from using the fixed effects estimator because it does not take into account the between estimator, which considers the variation between the cross section observations. Specifically the model that we use is:

$$E(y_{it}|X_{it}, Z_i) = \exp(\mu + \gamma Z_i + \beta X_{it} + c_i + u_{it}) \qquad (1)$$

With y_{it} as a Negative Binomial

We assume that the errors are not serially correlated and allow for correlation between the unobserved time variant effect and the explanatory variables[10]. In addition, we

[10] This assumption is supported by the studies of Zener (1968, 1970) who suggests that productivity is the result of many “mental factors” that vary across time.

assume that the specific individual constant effect is not correlated with the individual characteristics of the researcher[11]. Finally, we assume that there is no multicollinearity among the explanatory variables that vary across time and among time invariant explanatory variables. Specifically our assumptions are:

$$E(X'_{it}\, u_{it}) \neq 0$$
$$E(Z'_{i}\, c_{i}) = 0$$
$$E\,(u_{it} \mid X_{it}, c_{i}) = 0,\ t = 1,2,...,T$$
$$\text{Rank}\ [E(X'_{it}X_{it})] = K$$
$$E\,(\, c_{i} \mid Z_{i}) = 0$$
$$\text{Rank}\ [E(Z'_{i}Z_{i})] = K$$

We estimate β by the Conditional Maximum Likelihood Estimation (CMLE) proposed by Hausman, et al. (1984), and γ by the Non Linear Least Square Method in a second step proposed by Turner and Mairesse (2003), replacing β by its CMLE estimate and estimating equation (2).

$$y_{it} / \exp\,(\beta\text{hat}\ X_{it}) = \exp\,(\mu + \gamma\ Z_{i} + c_{i}) \qquad (2)$$

Since the method that we are using estimates variables in deviation from their mean, time dummies variables and the age variable tend to have collinearity problems. Therefore,

[11] The only variable that could be endogenous is the institution, but since there is almost no mobility across institutions, and currently there are no vacancies in the biggest institutions, few researchers can choose the institution where they would like to work.

because one of our main interests is the analysis of the dynamics of productivity over the life cycle[12], we decided not to include time dummies[13].

Because data are missing in some of the variables and in order to use all the observations in the subsample, binary indicators were created when the corresponding value of the variables were missing. The models were run using these binary variables as instruments along with the rest of the variables (Wooldridge, 2002).

The same model and assumptions are used to adjust for paper quality. In this case the dependent variable is the number of publications plus the number of citations in the four years subsequent to publication. The four year citation window was chosen trying to balance the loss of observation years in our panel due to the citation window, against the desire to include as much as possible this measure of quality and impact. Besides, on average, publications receive 70% of cites in this window, with the remaining 30% received in the following 10 years. Alternative approaches would be possible[14]. One could be to consider the average number of citations per year; however, one drawback of this approach is that the more recent publications could have smaller average number of citations. This could cause an important bias in our results. Another approach could be to

[12] We also tried the alternative approach of analyzing the dynamics of productivity by considering years since PhD was earned instead of age. The results are similar to the ones reported here. We chose to report only the results using age because they allow better comparison with previous studies.

[13] We also ran the same models including time dummies and the change in the variables was not significant. So, we report only the regressions without considering time dummies.

[14] Regressions were run using these alternative approaches, and the results were very similar.

use the expected number of citations produced by ISI. We decided to use the number of citations per four years instead of ISI factor because there is not enough evidence that the ISI method would yield a better indicator of the quality and significance of a given publication[15].

4. Results

In the first section, we begin by presenting the results obtained in the fixed model for the time variant variables. Special emphasis is put on the interpretation of the age effects. In the second section, we turn to the analysis of the results of the second step where the coefficients of the time invariant variables were obtained. In the following section we present the interpretation of the results for citations. Finally, in the last section, we present the interpretation for reputation. The results of the regressions using publications as a productivity measure are shown in Table 3[16]. Table 5 reports the results for citations.

- Table 3 -

[15] It should be noted that the two indicators are quite close. The correlation between the number of citations per four years and the expected number of citations produced by ISI is 0.6.

[16] In preliminary work we estimated publication regressions by Poisson and Least Squares. The qualitative results, i.e., the sign and significance of key parameters, did not vary substantially across the different specifications. Considering that the negative binomial is the best model given the nature of our data, we restrict our attention and our comments to the negative binomial estimates.

4.1 Age Effects

The age coefficients of Table 3 and Figure 2, which represent the predicted value of publications over the life cycle, confirm the quadratic relationship between age of researchers and publications per year that other authors have also found. According to our model, researchers are productive, in terms of publishing, between 30 and 79 years old[17], reaching a peak of 1.76 papers per year by the time they are 53 years old. It should be noted that while a quadratic relationship is also found when considering areas of knowledge, there is a wide variation among different areas.

- Figure 2 -

The decline of productivity in Exact Sciences researchers does not occur until they are 57 years old, even though they receive their PhD when they are younger (33.5 years old with a standard deviation of 5.6 years). They reach a maximum of 2.1 publications per year. Researchers in Biology and Chemistry reach the highest number of publications per year among all areas, about 2.17; the decline starts when they are 54 years old, and the average age of graduation is 34.6 (5.5)[18]. Researchers in Health Sciences have the smallest dispersion of all disciplines. They reach a maximum of 2.04 publications per year at 51 years old. They obtain their PhDs on average at 36.4 years of age (6.8). Researchers in Social and Humanities have the lowest productivity, reaching a maximum

[17] On average, they get their PhD when they are 36 years old.

[18] Standard deviations are presented in parentheses.

of 0.95 publications per year at the same age as the overall sample, although they graduate later than everybody else, at 39.2 years old on average (8.2). Researchers in Agricultural Sciences and Biotechnology reach a maximum of 1.27 publications per year, the decline starts when they are 54, and the average age of graduation is 36.9 (5.7). Finally, researchers in Engineering reach a maximum of 1.24 publications per year, the decline starts when they are 52, and the average graduation age is 34.2 (5.3). Table 4 summarizes the life cycle productivity dynamics for all areas considered in the estimation.

As can be observed, researchers in Biology and Chemistry are the most productive, not only because they reach the highest peak of publications (2.17 per year), but also because their productive cycle is the longest (56 years). Besides, researchers in Biology and Chemistry, and in Health Sciences are the ones to start publishing at a younger age, 27 years old. The other three areas of knowledge, Social and Humanities, Agricultural Sciences and Biotechnology, and Engineering are less productive. Researchers in Social and Humanities and Engineering have the shortest productive cycle (38 years) and are the ones that have the latest start in publishing (36 and 34, respectively).

- Table 4 -

The observed differences in productivity behavior across areas could suggest that knowledge in Exact Sciences, Biology and Chemistry, and Health Sciences is more easily codified, making its transmission through published papers easier. Thus, researchers can publish more, for more years, and from a younger age. Another possibility is that researchers in Social and Humanities, Agricultural Sciences and Biotechnology, and Engineering work in research topics that are mostly of local interest (regional or country level), and their results tend to be diffused in local journals that are not part of the ISI databases, or in other media. On the other hand, knowledge in Exact Sciences, Biology and Chemistry, and Health Sciences tends to be more universal and, therefore, more easily captured in our measure of productivity.

Although the analysis of all areas of knowledge is interesting, we decided to analyze Physics in detail to compare our results with those of Levin and Stephan (1991), and Turner and Mairesse (2003) [19].

Physicists are increasingly productive until they are 55 years old, reaching a maximum of 2.06 publications per year, and their average graduation age is 32.9 (5.2). These figures are quite different from the results of Levin and Stephan (1991) (model B in their paper), where they find that solid state and condensed matter physicists reach a peak of 2 papers at 45. It is also different from Turner and Mairesse (2003), where condensed matter physicists reach a peak of 2.9 publications at 52 years old[20].

[19] Since we are considering all physics and not the subdisciplines of physics, this is not totally comparable. Besides, Levin and Stephan (1991) use a pooled cross sectional Tobit model with two years of publications in the 1970's. In the case of Turner and Mairesse (2003), they use a Poisson model with publications from 1986 to 1997.

[20] This result was gotten from an annex regression in which age and age squared replaced the age cohorts.

All the results suggest that Mexican researchers reach their publishing peak 5 to 10 years later than what other studies have shown. Moreover, the decline is quite subtle until reaching the retirement age (65 in Mexico), where they are as productive as when they were 43 and more than when, on average, they got their PhD.

We believe that at least three different factors, not mutually exclusive, can help explain this dissimilarity. The first one is that Mexican researchers are responding to the incentives created by the National System of Researchers, which encourages researchers to continue publishing throughout their lives. This is related to the fact that, in Mexico, the base salary of researchers is about one third of what they actually receive. The other two thirds are given in the form of pecuniary compensations directly from SNI and from the university where researchers work, also heavily dependent on SNI's appraisal of the candidate's proficiency. Moreover, the system does not guarantee 'tenure' and its payments are not considered as salary when establishing a person's retirement pension. Evaluation is done regularly and the researcher can be demoted if the panel evaluating him or her assesses lower research productivity. Thus, researchers are reluctant to retire and they have to continue showing some productivity in order to receive these complements of salary.

A second explanation could be that the eldest researchers of a research group or a laboratory tend to appear as coauthors in the publications of their colleagues more frequently than what would be found in more developed research systems such as in the

U.S.[21]. Yet, a third explanation could be that Mexican researchers get their PhD when they are older, at least in comparison to the U.S., the country used for most previous productivity studies. In our sample, the average age of PhD graduation is 36.1 years old with a standard deviation of 6.9 years. Thus, it is possible that Mexican researchers start and finish their careers some years later than their colleagues in the developed world.

The first potential explanation for the lack of a substantial age effect on publishing, related with the salary structure in Mexico, cannot easily be tested because of the lack of an adequate control sample. The second one, suggesting that the eldest researchers of a research group or laboratory tend to appear as coauthors in the publications of their colleagues, will be tested in future research. The third one, which supposes that Mexican researchers start and finish their careers some years later than their colleagues in other countries, can be easily tested in a number of ways with the dataset. If indeed the cycle is conditioned by age of PhD completion, one would expect that those who get their PhD earlier would, on average, peak sooner and decline earlier than the rest.

To test this idea, we divided the sample into two groups: those researchers who got their PhD when they were 30 years old or younger, and those who were above 30[22]. The

[21] This means that, because the system is very small and hierarchical, less senior researchers would tend to include older researchers in the department as coauthors just because of their seniority, even if they do not contribute to the work.

[22] 30 years old is a conservative upper bound for the PhD completion age of a person that does all degrees straight one after the other, including a masters degree.

results[23] suggest that the decline for those in the first group happens six years later than the second group. In addition, those completing their PhD earlier reach a peak of 0.5 publications per year more than the second group. These results are not a regularity across areas of knowledge. This evidence contradicts our hypothesized explanation, and suggests that PhD completion age is not the critical factor determining the life cycle research productivity effects.

4.2. Other Time Variant Effects

4.2.1. Level in SNI

The results of the regression show that researchers in level 3, the more prestigious and better rewarded level in SNI, produce 0.39 papers more than candidates. Researchers in level 2 publish 0.45 papers more, and researchers in level 1 publish 0.42 papers more than the same baseline. However, the differences in the coefficients are not statistically significant, except for the difference between level 3 and level 2, which at 8% of confidence suggest that level 2 is more productive than level 3.

The average age of researchers in level 3 is 58.7 (10.6), in level 2 is 50.6 (8.9), in level 1 is 44.4 (7.9), and in the candidate level is 34.8 (3.8). Therefore, it is possible to have confounded effects between the level in the SNI program and age. To test for this, we ran a model excluding the level variables. The changes in the remaining coefficients, as well as in the log likelihood were very small (See Model 2 in Table 6).

[23] Not shown here but available from the authors upon request.

4.2.2.Critical Mass

Our estimates suggest that although small, there is a positive and significant externality effect in the productivity when we consider the total number of publications in the same area of knowledge in all areas except in Physics. In contrast, it seems that there is a small system saturation effect, reflected in the coefficient for the total number of researchers in the same area of knowledge. This negative result is significant, even for individual areas of knowledge, but its effect is small. Still, this is a surprising result since the scientific community in Mexico is relatively small and one could, in principle, expect the presence of some type of critical mass effect as the system grows. It is important to note that the evaluation was done at a very aggregate level, which may not be the best way to observe research critical mass effects. Researchers tend to cooperate more with colleagues in the same institution or in the same region, not necessarily in the broad scientific area as defined in the paper. Hence, this variable may instead be measuring some degree of competition for resources. Although the Mexican community is small, researchers compete for scare resources needed for publishing in a particular area[24], so that more people means less individual funding and, as a consequence, lower productivity. Nevertheless, we feel that any potential interpretation of this variable should be done with care. Future work will specifically address potential critical mass aspects, and will use specific controls for institution and area to better identify these aspects.

[24] Mexico has a particularly small level of funding for science and technology, even when compared to peer countries.

4.2.3.Budget

We also tried to understand the impact of the Science and Technology (S&T) budget[25] on the productivity of researchers. It is reasonable to believe that budget has an impact on overall productivity with a lag, though it is hard to determine what lag to use. Therefore, several regressions considering different lags were run; the only significant one was the one with a previous year lag. Our results suggest that there is a significant and positive relation between the budget of the previous year and publications. However, this variable does not make a distinction among the resources that are invested in each area of knowledge or, even better, the resources that each researcher has. So, our ability to draw any strong conclusion on this topic is limited.

4.3. Time Invariant Effects

4.3.1. Gender

Gender effects can be identified from the female coefficient and its respective marginal effect in the time invariant portion of the regression shown in Table 3. Results suggest that there is not a big gender difference in scientific productivity for our sample. Mexican female scientists are not overrepresented among the non publishers, and they produce only slightly fewer papers (0.08 paper) than men on average per year. The proportion of female scientists varies a lot among areas of knowledge. The areas with greater representation are Health Sciences with 39%, and Social and Humanities with 38%,

[25] We are using the annual budget of Conacyt, since is the main agency in Mexico that supports S&T activities.

followed by Biology and Chemistry with 36%. In Agricultural Sciences and Biotechnology 22% are women, and in Exact Sciences, 15%. Although only 12% of engineers are women, our results suggest that women in Engineering are slightly more productive than men[26]. The biggest gap is found in Health Sciences where women publish 0.25 paper less than men. Women physicists publish 0.19 paper less than men per year. We can compare this result with that of Turner and Mairesse (2003), who found that a woman publishes almost 0.9 paper less than a man on average per year. In their study, the proportion of women is 18% and in our case it is 11%. However, we do not have any explanation about this situation. As was mentioned by Cole and Zuckerman (1984), sex differences in research productivity continues to be "the productivity puzzle."

4.3.2. Country of PhD

Our results suggest that researchers who got their PhD in the United States or Europe publish about 0.2 paper less than those who got their PhD in Mexico or in other countries. However, there are some caveats that it is important to note. First, there is 27% of missing data associated with this parameter. Second, our database does not include publications that might have been published by the researcher in an institution outside Mexico. That is, we are not taking into account those publications that were published when the researcher was studying for his PhD abroad[27]. To avoid this bias, we decided to include a dummy variable, in the first step, that takes the value of one if the publication(s)

[26] In Social and Humanities the parameter is also positive but it is not significant at 10%.

[27] Indeed, we only have 303 observations of researchers who got their PhD abroad and published with a Mexican address before they graduated.

was published before the year of graduation. This new regression considers only those cases where it is known not only the country where the PhD was earned but also the year of graduation. This subsample includes 4,193 researchers.

The results of this new regression suggest that there is no significant difference depending on the country where the PhD was earned. This result could be interpreted in several ways. First, it could be that the most productive researchers who got their PhD outside Mexico stay in other countries (“brain drain”). Another possibility could be that the advantages of gaining access to bigger networks when someone studies abroad are compensated if it is easier for those researchers who got their PhD in Mexico to find partners in Mexico with whom to publish, or have a better knowledge of how to obtain research resources. One last possible explanation could be that the peer committees of SNI do not make any distinction depending on the country where the PhD was earned, so that all researchers have to show the same productivity no matter where they graduated. These suggestions must definitely be interpreted very carefully since there is not enough evidence to support them.

To better understand the results of this variable, we consider it important to note that there is a difference in the average age of graduation depending on the country where PhD was earned. For those who got their PhD in Mexico, the average age of graduation is 38.2 (7.4), and for those who graduated from American or European universities, the average age is 34.8 (5.4) and 33.6 (5.7), respectively. There are also differences in the proportion of researchers in each area of knowledge who got their PhD abroad. The two most productive areas, Biology and Chemistry, and Health Sciences, have the lowest proportion of researchers who got their PhD outside of Mexico, 30% and 32%, respectively. The largest proportion is in Engineering (69%), followed by Exact Sciences, and Agriculture Sciences and Biotechnology with 58% each. In Social and Humanities 48% of the researchers got their PhD abroad.

We also ran new regressions by area of knowledge to capture the bias that is caused for not having all the publications before graduation. The results suggest that researchers in Exact Sciences and Social and Humanities who got their PhD abroad are slightly more productive. In contrast, engineers who got their PhD in Mexico are more productive. Researchers in Health Sciences, and in Agricultural Sciences and Biotechnology who got their PhD in Europe are slightly less productive. No significant difference is found in researchers in Biology and Chemistry.

4.3.3. Cohort Effects

The cohorts were created by looking at the evolution of Mexican publications in ISI since 1966[28]. Cohorts were created assuming that every time that there was a disruptive change in the trend of publications, this implied an important change in the Mexican scientific knowledge base. Four cohorts were created. The first one includes researchers who got their PhD before 1972, the second one includes researchers who got their PhD between 1973 and 1981, the third one consists of those who got their PhD between 1982 and 1991, and the last one those who graduated after 1992.

Our results suggest that the latest educated are the best educated due to the change in the knowledge base. We find that researchers who graduated before 1971 and between 1972-1983 publish about 0.2 paper less than those in the cohort 1992-2002. It may also reflect that this cohort is under greater pressure to publish, and since we are analyzing publications in the 1990's, it is possible that nowadays researchers at the beginning of their careers tend to publish more than previous generations because of the new culture of monitoring and evaluation.

In the analysis of areas of knowledge, our results vary a lot in significance for this control. The areas of knowledge where we get significant results in most of the cohorts are Exact Sciences and Social and Humanities. For researchers in Exact Sciences our results support the general presumption that the latest educated are the best educated, and thus the most productive. The cohort 1992-2002 is more productive than any other cohort, the greatest difference being with the cohort that got their PhD between 1973 and 1981. This latter cohort publishes 0.5 paper less. The same result is seen in Physics. In Social and Humanities our results suggest that the cohort 1992-2002 is more productive than the cohorts 1973-1981 (0.26 paper) and 1982-1991 (0.07 paper). However, there is no significance difference with the cohort that graduated before 1972.

Considering that the results for the other areas of knowledge are not significant and that we are analyzing publications from the nineties, we believe that the cohort effect that is found is not very strong and can not be supported in all cases.

---

[28] 1966 is the first year that Mexican publications appear in the "Web of Knowledge" of ISI.

4.4 Results for citations

To adjust for quality of publications, the number of citations per four years was used and he results of the regressions are shown in Table 5[29]. When accounting for citations over four years, the baseline data is reduced to a sub-sample of 5658 researchers and 7 years of publications (1991-1997).

- Table 5 -

The estimation also confirms a quadratic relation between age of researchers and citations that was found using publications. The peak of citations per four years is 0.42 and is reached when researchers are 56 years old; this is three years older than what is found for publications. Breaking down by area of knowledge, it is observed that researchers in Exact Sciences, Biology and Chemistry, and Health Sciences reach a peak at 56 years old; and researchers in Social and Humanities, Agricultural Sciences and Biotechnology, and Engineering reach a peak a little bit younger, 54, 53 and 52, respectively. The greatest difference is seen in researchers in Health Sciences who reach a peak of productivity six years later when citations instead of publications are taken into account. These results could suggest that SNI researchers continue publishing not only because of the incentives of receiving the support of the program but also for the satisfaction of "solving the puzzle." (See Levin and Stephan, 1991).

[29] In preliminary work we dropped those publications that have more than 50 citations over 4 years. However, the results do not vary substantially.

Our results also suggest that publications of researchers in level 3 receive 0.28 more citations than those of candidates, and researchers in level 2 and in level 1 receive 0.21 and 0.14 more citations than candidates, respectively. Testing for the difference of the coefficients confirms that researchers in level 3 receive in all cases more citations than their peers in level 2 or 1. When the areas of knowledge are analyzed, the most important difference among levels in the program is seen in level 3 researchers in Exact Sciences who receive 0.52 more citations than candidates. Level 2 researchers receive 0.45 and level 1 receive 0.31 more citations than the lowest level in the program. On the other hand, the area with less difference is Agricultural Sciences and Biotechnology. In this case, researchers in level 3 receive 0.12 more citations, in level 2 0.08 and in level 1 0.06, than candidates. This is an important result because as researchers become more senior, it is not the count of publications that matters, but rather the impact of those publications. This could be associated with the Matthew Effect, which tends to give the credit to the most prestigious researchers, as will be seen for the results of professional reputation in the next section.

Researchers in Health Sciences receive the largest number of citations per four years, getting 0.44 more citations than researchers in Social and Humanities. Researchers in Exact Sciences and Biology and Chemistry received 0.33 and 0.34 more citations than Social and Humanities researchers, respectively. Engineers and researchers in Agriculture Sciences and Biotechnology received only 0.15 and 0.16 more citations, respectively, considering the same reference. These results could suggest, as in the case of publications, that knowledge in Exact Sciences, Biology and Chemistry, and Health

Sciences is more codified and universal, and its transmission is easier; and that researchers in Social and Humanities, Agricultural Sciences and Biotechnology, and Engineering choose research topics that are of local interest (regional or country level), and their results tend to be diffused in other media.

As in the case of publications, there is only a slight difference between Mexican female scientists and their male colleagues. They receive only 0.06 fewer citations. The area of knowledge that has the largest gap is Health Sciences where women receive 0.14 fewer citations than men. However, in Social and Humanities, as well as in Engineering, women receive slightly more citations than men, 0.02 and 0.04, respectively.

Our results also suggest that there is almost no difference in citations depending on the country where PhD was earned[30]. However, there is some difference depending on the cohort. Those researchers who got their PhD before 1972 receive 0.1 less citations than those who graduated after 1992; those who got their PhD between 1973 and 1981 receive 0.15 more citations, and those who got their PhD between 1982 and 1991 receive 0.13 more citations. As was mentioned earlier, this could also suggest that at the beginning of their careers researchers tend to publish many papers in order to get more prestige, and later on their concern is more about the quality than the quantity in their publications. When looking at the areas of knowledge, we find similar results.

[30] This result does not change if we consider only those researchers whose country of PhD and year of graduation is known and we correct for publications that were published before the year of graduation.

Although there are some differences in the number of citations depending on the cohort, these differences are very small. Thus, we believe that there is not strong evidence that support cohort effects.

4.5. Reputation

One aspect that is important to include in studying the determinants of research productivity is the reward structure of science (Merton, 1968 and 1988; and Stephan, 1996). The count of publications and citations are the most common way to measure a scientist's contribution. However, as mentioned earlier, in science the distribution of recognition is influenced by a "class structure" (Merton, 1968) that skews in favor of those researchers who already have reputation. It is possible that the Matthew Effect is even more critical in countries like Mexico where the scientific community is small. Thus, a potential consequence is that the access to the means of scientific production is different, and past publications and citations are a critical determinant of research productivity.

- Table 6 -

Since there is no standard way to measure reputation, we used several alternative specifications to explore how reputation influences future productivity. First, we included as an additional time variant explanatory variable the stock of publications (model 4 in Table 6) or the stock of publications and citations (model 5 in Table 6) in the past 10 years. Our results show that prior publications have a negative impact on current publications[31], but a positive effect in the number of citations. An alternative way to proxy for reputation would be to depreciate the stock of publications or use the stock of citations (see Model 3 in table 6). For example, Henderson and Cockburn (1996) use the depreciated stock of own patents as an explanatory variable of patents in a firm; and Agrawal and Henderson (2002) use the depreciated stock of patents and papers to explore the relationship between patenting and publishing among faculty. Likewise, the stock of publications plus citations can be used (see model 5 in table 6). Table 6 shows the results of the regressions using these alternative variables. As can be seen, the one that best explains the importance of reputation in publications as well as in citations is the stock of publications over 10 years.

These results suggest that, in Mexico, a record of past publications signals less future papers but higher impact. When looking at the effect by areas of knowledge, we find similar results. For all areas of knowledge the effect of prior publications in publication count is negative, and the effect in citation count is positive only for Exact Sciences, Biology and Chemistry, and Health Sciences[32]. We find evidence to support the Mathew Effect in science. Researchers that have gained some repute will tend to receive more citations in the future. However, at least in the Mexican scientific community, this inequality is relatively small, considering the marginal effect of the coefficients.

[31] To test for possible cyclical effects we lagged this variable; however, the coefficients were still negative.

5. Conclusions

This work analyzes the determinants of researcher productivity using a database of the most productive researchers in Mexico. We find that, although there is a quadratic relationship between age and researcher productivity, the effect of age is not very important: SNI researchers at 65 years old are as productive as those at 43 and more than when they finish their PhD. This result is consistent not only for count in publications but also for citations. Thus, we can conclude that researchers at retirement age have the same satisfaction of “solving the puzzle” and the same economic incentive as their colleagues who are 25 years younger. This could either suggest that research activity over the life cycle is not investment-motivated, or that it is, and the system in Mexico has created incentives for people to continue such investment until the end of their working lives. This could be also related to Levin and Stephan’s (1991) conclusions that age explains research productivity very little, suggesting that at least some of the processes at work are state dependent (Stephan, 1996). These findings could have important implications for the design of public policy, mainly in countries where the research system is funded by public funds, because it seems that it is possible to create programs to enhance productivity until the end of researchers’ working lives.

We also find that there are significant differences in research productivity among areas of knowledge, not only in the peak of publications and citations but also in the productivity cycle. In addition, the proportion of non-productive researchers is also dissimilar. These

[32] The coefficients for the remaining 3 areas (Social and Humanities, Agricultural Sciences and

results suggest that researchers in different areas of knowledge have singular incentives, at least to publish in ISI journals. Besides, it seems that the nature of knowledge is different. There are some areas, such as Exact Sciences, Biology and Chemistry, and Health Sciences, where knowledge is more codified and universal, so that the outcomes of research could be published and diffused more extensively than knowledge in other areas, such as Engineering, Social and Humanities, and Agricultural Sciences and Biotechnology.

When comparing with similar studies in the U.S. and France, we found that the difference in the peak of publications between Mexican and American physicists is only 0.06 publication per year. However, when comparing with the French, Mexicans are 0.84 paper per year less productive. Considering that the study for the U.S. was done with data of the 1970's and the one in France with data of the late 1980's and 1990's, we could conclude that the pressure for publishing has increased in the last decade, both in developed and developing countries.

Our findings related to age, in conjunction with the results of the other variables, mainly the cohort effects, have important implications in terms of the planning of higher education institutions. For example, the concern that the average age of researchers has increased, could be something that is not a major issue in terms of count of publications. We also find that although the marginal effect of reputation is small, it matters for the number of citations but not for the number of publications.

Biotechnology, and Engineering) are not significant at 10%.

However, these results must not be considered completely conclusive. As we mentioned before, our sample includes only the most productive researchers. In addition, it must also be taken under consideration that publications are not the only product of research. Although publications and citations are the most commonly used measures, other factors such as the number of coauthors, author position, and other types of productivity output such as books and patents, need to be considered to get stronger results.

Furthermore, it is possible that the reward structure of the research establishment responds to initiatives like SNI, and researchers choose to allocate time to those activities that maximize future rewards. However, it is possible that those activities are not necessarily the ones that produce the greater social payoff. For example, a researcher can choose between allocating time to publishing or to teaching. Perhaps these activities complement each other, but if not, the current reward system could be giving fewer incentives to those activities that produce higher social payoff.

It would be useful to study the determinants of research productivity in other countries, especially outside the more developed world, to determine whether the patterns hold more generally. Studying publications during the 1970’s and 1980’s might be especially interesting to reveal changes in the patterns of publications over time.

Acknowledgements

We are grateful for helpful comments from Paula Stephan, Bronwyn Hall, Granger Morgan, Robert Lowe, and Steven Kleeper. Support from the Mexican Council for Science and Technology (Conacyt) is gratefully acknowledges.

References

Agrawal, A. & Henderson, R. (2002). Putting Patents in Context: Exploring Knowledge Transfer from MIT. *Management Science*, 48 (1), 44-60.

Allison, P., Long, S. & Krauze, T. (1982). Cumulative Advantage and Inequality in Science. *American Sociological Review*, 47 (5), 615-625.

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

Arora, A., David, P, & Gambardella, A. (1998). Reputation and Competence in Publicly Funded Science. *Annales d'Economie et de Statistiques*, 49/50, 163-196.

Arora, A. & Gambardella, A. (1998). *The Impact of NSF Support for Basic Research in Economics*. Working paper, April.

Arundel, A. & Geuna, A. (2001*). Does Proximity Matters for Knowledge Transfer from Public Institutes and Universities to Firms?*. SPRU, University of Sussex.

Becker, G. (1962). Investment in Human Capital: A Theoretical Analysis. *The Journal of Political Economy*, 70 (5), 9-49.

Bernier, C., Gill, W. & Hunt, R. (1975). Measures of Excellence of Engineering and Science Departments: A Chemical Engineering Example. *Chemical Engineering Education*, 194-197.

Buchmueller, T., Dominitz, J. & Hansen, L. (1999). Graduate Training and the Early Career Productivity of PhD Economists. *Economics of Education Review*, 14, 65-77.

Bush, V. (1945). *Science, The Endless Frontier*. Washington, D. C., Office of Scientific Research and Development.

Cole, S. (1979). Age and Scientific Performance. *The American Journal of Sociology*, 84 (4), 958-977.

Cole, J. & Zuckerman, H. (1984). The Productivity Puzzle: Persistence and Change in Patterns of Publications of Men and Women Scientists. *Advances in Motivation and Achievement*, 2, 217-258.

CONACYT. (2001). *Programa Especial de Ciencia y Tecnología 2001-2206*. Mexico.

CONACYT. (2002). *Indicadores de Actividades Científicas y Tecnológicas 2002*. Mexico.

Diamond, A. M. (1986). The Life-Cycle Research Productivity of Mathematicians and Scientists. *Journal of Gerontology*, 41, 520-525.

Hausman, J., Hall, B. & Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica*, 52, 909-938.

Henderson, R. & Cockburn, I. (1996). Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery. *The Rand Journal of Economics*, 27 (1), 32-59.

Institute of Scientific Information (2003, April 18). *National Citation Report for Mexico*. Articles from ISI database years 1981-2002 cited in the period 1981-2002. Project 2487.

Lawrence, P.A. (2003). The Politics of Publication. *Nature*, 422 (20), 259-261.

Levin, S. & Stephan, P. (1991). Research Productivity Over the Life Cycle: Evidence for Academic Scientists. *The American Economic Review*, 81, 114-132.

Long, S. (1992). Measures of Sex Difference in Scientific Productivity. *Social Forces*, 71 (1), 159-178.

Merton, R. (1968). The Matthew Effect in Science. The Reward and Communication Systems of Science are Considered. *Science*, 159 (3810), 56-63.

Merton, R. (1988). The Matthew Effect in Science, II. Cumulative Advantage and the Symbolism of Intellectual Property. *ISIS*, 79, 606-623.

Nelson, R.(Editor). (1993). *National Innovation Systems: A Comparative Analysis*. Oxford University Press.

OECD. (2002). Ad Hoc Working Group. Steering and Funding of Research Institutions. *Governance of Science Systems: New Challenges*. Detailed Outline of the Final Project Report. Paris.

Schultz, T.W. (1963). *The Economic Value of Education*. New York: Columbia University Press.

Stephan, P. (1996). The Economics of Science. *Journal of Economic Literature*, XXXIV, 1199-1235.

Turner, L. & Mairesse, J. (2003). Explaining Individual Productivity Differences in Scientific Research Productivity: How important are Institutional and Individual Determinants? An Econometric Analysis of the Publications of French CNRS Physicists in Condensed Matter (1980-1997). *Submitted to Annales d'Economie et de Statistiques* for the special issue in honor of Zvi Griliches.

Weiss, Y. & Lillard, L. (1982). Output Variability, Academic Labor Contracts, and Waiting Times for Promotion. *Research in Labor Economics*, 5, 157-188.

Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Xie, Y. & Shauman, K. (1998). Sex Differences in Research Productivity: New Evidence About an Old Puzzle. *American Sociological Review*, 63, 847-870.

Zener, C. (1968). *An Analysis of Scientific Productivity*. Texas A&M University, College Station.

Zener, C. (1970). Statistical Theories of Success. *Proceedings of the National Academy of Sciences*, 66 (1) 18-24.

Tables and Figures

Table 1. Main Findings of Studies Related to the Determinants of Research Productivity.

Determinant	Author	Main findings
Age	Bernier et al. (1975)	Life cycle with a peak in the 40-44 group.
	Cole (1979)	Life cycle with a peak before 50's
	Levin and Stephan (1991)	Life cycle with a peak at 45
	Turner and Mairesse (2003)	Life cycle with peak at 50
Gender	Cole and Zuckerman (1984)	Women publish 57% as many papers as men.
	Long (1992)	Males publish between 26 and 91% more than women.
	Xie and Shauman (1998)	Female-to-male ratio from 60 to 80%
	Tuner and Mairesse (2003)	Women publish 0.9 fewer paper per year.
Cumulative advantage	Allison and Stewart (1974)	Inequality between productive and unproductive researchers increases with age
	Cole (1979)	Inequality between productive and unproductive researchers increases with age
Education	Buchmueller et al. (1999)	Graduates from top schools, with research assistant experience and employed in research universities, are more productive
	Turner and Mairesse (2003)	Graduates from Grande Ecoles are more productive
Income	Stephan (1996)	Salary is positively related to both article and citation counts.
Cohort effect	Levin and Stephan (1991)	No evidence

Figure 1. Evolution of the mean of ISI publications per year by SNI authors

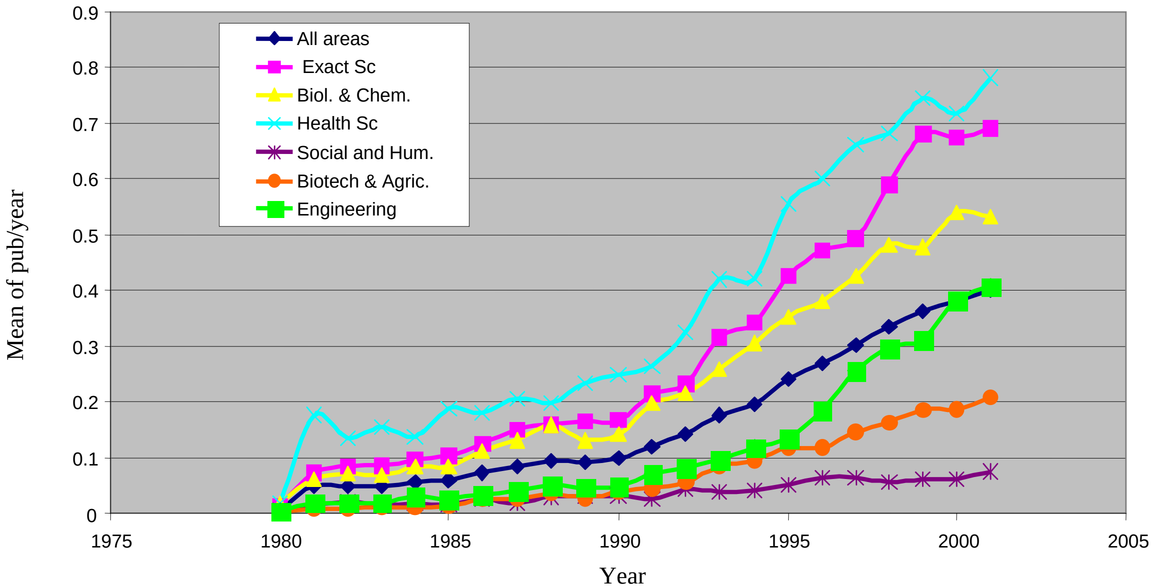

Figure 2. Research Productivity Over the Life Cycle: All Areas of Knowledge

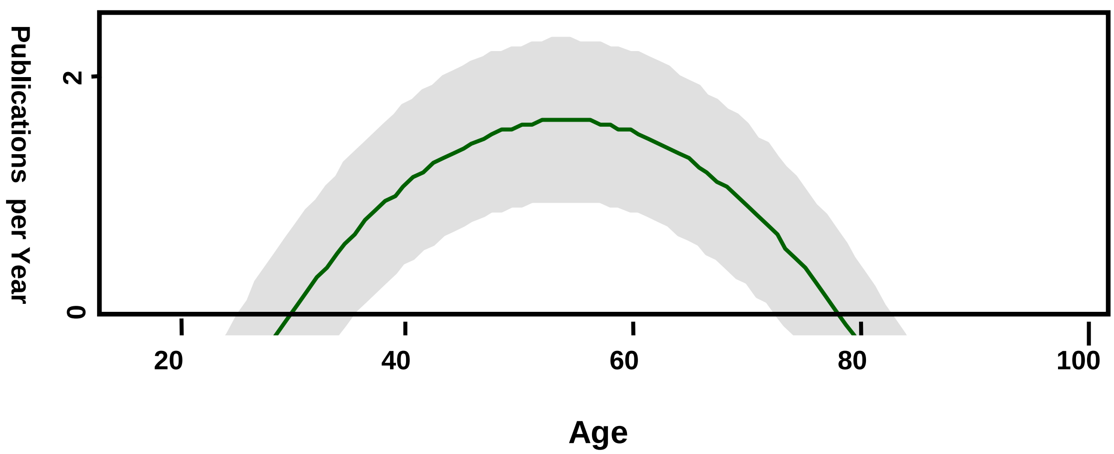

Table 2. Descriptive Statistics

Variable	Mean	Std. Dev.
Time variant variables		
Publications per year	0.489	0.965
Age	40.321	9.817
Age Square	1722.187	884.546
Level 3	0.079	0.270
Level 2	0.171	0.376
Level 1	0.538	0.499
Level Candidate	0.211	0.408
Budget in Millions of dollars (2001)	332.351	58.593
Budget -1 year in Millions of dollars (2001)	308.723	82.168
Budget -2 years in Millions of dollars (2001)	286.996	101.046
Time invariant variables		
Exact Sciences	0.221	0.415
Biology and Chemistry	0.244	0.429
Health Sciences	0.171	0.376
Social and Humanities	0.097	0.296
Agriculture and Biotechnology	0.118	0.322
Engineering	0.150	0.357
Male	0.730	0.444
Female	0.270	0.444
Public Institution	0.967	0.178
Private institution	0.032	0.177
PhD in Mexico	0.508	0.500
PhD in USA	0.173	0.378
PhD in Europe	0.271	0.445
Cohort before 1972	0.049	0.216
Cohort 1973-1981	0.103	0.304
Cohort 1982-1991	0.284	0.451
Cohort 1992-2002	0.564	0.496
Total # of years in SNI (1991-2002)	6.762	3.890

N= 7793. Total number of publications (1991-2001) >0

Table 3. Publications. Regressions for all Areas of Knowledge.

Variables	*Coef.*	*Std. Err.*		*Marginal Impact*
Time variant variables				0.9814
Age	0.2637	0.0093	***	0.2588
Age Square	-0.0025	0.0001	***	-0.0024
Level 3	0.3950	0.0475	***	0.3877
Level 2	0.4624	0.0304	***	0.4538
Level 1	0.4302	0.0198	***	0.4222
Total # of pub	0.0005	0.0000	***	0.0005
Total # of researchers	-0.0003	0.0000	***	-0.0003
Budget -1 year in Millions of dollars (2001)	0.0009	0.0001	***	0.0009
Constant	-5.9628	0.2117	***	
Time invariant variables				
Exact Sciences	0.3706	0.0376	***	0.3637
Biology and Chemistry	0.3719	0.0371	***	0.3750
Health Sciences	0.4670	0.0389	***	0.4683
Agriculture and Biotechnology	0.0370	0.0438		0.0363
Engineering	0.1621	0.0408	***	0.1591
Female	-0.0771	0.0202	***	-0.0757
PhD in USA	-0.2180	0.0302	***	-0.2139
PhD in Europe	-0.1885	0.0250	***	-0.1850
Cohort before 1972	-0.1669	0.0594	***	-0.1638
Cohort 1973-1981	-0.1837	0.0425	***	-0.1803
Cohort 1982-1991	-0.0109	0.0253		-0.0107
Constant	-2.0293	0.0612	***	-1.9916

The base is a male researcher in Social and Humanities, who got his PhD in Mexico or other country between 1992 and 2002.
N=7793, T=1991-2001
*** Significant at 1%, ** significant at 5%, * significant at 10%.

Table 3. Regressions for Individual Scientific Areas

	Exact Sciences			Biology and Chemistry			Heath Sciences		
	Coef.		Marginal Impact	Coef.		Marginal Impact	Coef.		Marginal Impact
Time variant variables			1.059			1.336			1.350
Age	0.2535	***	0.268	0.2737	***	0.366	0.2838	***	0.383
Age Square	-0.0023	***	-0.002	-0.0025	***	-0.003	-0.0027	***	-0.004
Level 3	0.9264	***	0.981	0.1550		0.207	-0.0377		-0.051
Level 2	0.9509	***	1.007	0.2183	***	0.292	0.0908		0.123
Level 1	0.8149	***	0.863	0.2187	***	0.292	0.2025	***	0.273
Total # of pub	0.0001		0.000	0.0003	**	0.000	0.0008	***	0.001
Total # of researchers	0.0001		0.000	-0.0004		-0.001	-0.0008	***	-0.001
Budget -1	0.0012	***	0.001	0.0008		0.001	0.0009		0.001
Constant	-6.1644	***	-6.526	-5.4228	***	-7.244	-5.8518	***	-7.899
Time invariant variables									
Female	-0.1487	***	-0.157	-0.1153	***	-0.154	-0.1840	***	-0.249
PhD in USA	-0.2287	***	-0.242	-0.2445	***	-0.327	-0.1014		-0.137
PhD in Europe	-0.1559	***	-0.165	-0.1317	**	-0.176	-0.4360	***	-0.589
Cohort before 1972	-0.4419	***	-0.468	-0.2408		0.322	-0.0086		-0.012
Cohort 1973-1981	-0.4757	***	-0.504	0.0459		-0.061	-0.0536		-0.072
Cohort 1982-1991	-0.1803	***	-0.191	0.0464		0.062	0.1150		0.155
Constant	-2.0618	***	-2.183	-2.4772	***	-3.309	-1.5865	***	-2.142

*** Significant at 1%, ** significant at 5%, * significant at 10%.

Table 3. Continuation

	Social and Humanities Coef.		Social and Humanities Marginal Impact	Agricultural Sc and Biotechnology Coef.		Agricultural Sc and Biotechnology Marginal Impact	Engineering Coef.		Engineering Marginal Impact
Time variant variables			0.457			0.690			0.451
Age	0.2274	***	0.104	0.2228	***	0.154	0.2426	***	0.109
Age Square	-0.0022	***	-0.001	-0.0022	***	-0.001	-0.0024	***	-0.001
Level 3	0.6382	***	0.292	-0.0386		-0.027	0.2672	**	0.120
Level 2	0.6438	***	0.294	0.0689		0.047	0.5503	***	0.248
Level 1	0.5762	***	0.263	0.1790	***	0.123	0.5661	***	0.255
Total # of pub	0.0050	***	0.002	0.0023	***	0.002	0.0016	***	0.001
Total # of researchers	-0.0005	**	0.000	0.0002		0.000	-0.0008	***	0.000
Budget -1	0.0002		0.000	0.0004		0.000	-0.0005		0.000
Constant	-5.6422	***	-2.579	-5.2671	***	-3.632	-5.3475	***	-2.410
Time invariant variables									
Female	0.0616		0.028	-0.0951		-0.066	0.1293	**	0.058
PhD in USA	0.0369		0.017	-0.2478	***	-0.171	-0.3844	***	-0.173
PhD in Europe	-0.0676		-0.031	-0.2970	***	-0.205	-0.3569	***	-0.161
Cohort before 1972	0.0168		0.008	0.1066		0.074	0.0491		0.022
Cohort 1973-1981	-0.5638	***	-0.258	0.0343		0.023	0.0601		0.027
Cohort 1982-1991	-0.1566	*	-0.072	0.3573	***	0.246	0.0904		0.041
Constant	-1.6442	***	-0.751	-1.6338	***	-1.127	-1.5846	***	-0.715

*** Significant at 1%, ** significant at 5%, * significant at 10%.

Table 3: Continuation

	Physics Coef.		Physics Marginal Impact
Time variant variables			1.147
Age	0.2219	***	0.255
Age Square	-0.0021	***	-0.002
Level 3	0.8565	***	0.983
Level 2	0.8602	***	0.987
Level 1	0.7866	***	0.902
Total # of pub	-0.0002		0.000
Total # of researchers	0.0003		0.000
Budget -1	0.0018	***	0.002
Constant	-5.5221	***	-6.335
Time invariant variables			
Female	-0.1785	**	-0.205
PhD in USA	-0.0517		-0.059
PhD in Europe	-0.0803		-0.092
Cohort before 1972	-0.3938	***	-0.452
Cohort 1973-1981	-0.4594	***	-0.527
Cohort 1982-1991	-0.2083	***	-0.239
Constant	-1.7545	***	-2.012

*** Significant at 1%, ** significant at 5%, * significant at 10%.

Table 4. Publishing Dynamics in the Different Areas of Knowledge.

Area of knowledge	Researchers are productive between: (years old)	The number of publications per year peaks at:	The peak in productivity is at age: (years old)
All	30-79	1.76	53
Exact Sciences	31-85	2.10	56
Bio. & Chem.	27-83	2.17	53
Health Sciences	27-78	2.04	50
Social and Hum.	36-74	0.95	53
Agric. & Biotech.	32-74	1.27	53
Engineering	34-72	1.24	51
Discipline			
Physics	29-83	2.06	55

Table 5. Citations per four years

Variables	*Coef.*	*Std. Err.*		*Marginal Impact*
Time variant variables				0.2509
Age	0.1446	0.0107	***	0.0363
Age Square	-0.0014	0.0001	***	-0.0004
Level 3	1.1161	0.0584	***	0.2800
Level 2	0.8348	0.0419	***	0.2094
Level 1	0.5699	0.0287	***	0.1430
Total # of pub	0.0007	0.0001	***	0.0002
Total # of researchers	-0.0002	0.0000	***	0.0000
Budget -1 year in Millions of dollars (2001)	0.0012	0.0001	***	0.0003
Constant	-5.9997	0.2336	***	
Time invariant variables				
Exact Sciences	1.2938	0.0239	***	0.3246
Biology and Chemistry	1.3664	0.0238	***	0.3428
Health Sciences	1.7367	0.0241	***	0.4357
Agriculture and Biotechnology	0.6451	0.0271	***	0.1619
Engineering	0.6139	0.0263	***	0.1540
Female	-0.2268	0.0093	***	-0.0569
PhD in USA	-0.0299	0.0122	**	-0.0075
PhD in Europe	-0.2755	0.0112	***	-0.0691
Cohort before 1972	0.2994	0.0226	***	0.0751
Cohort 1973-1981	0.5987	0.0146	***	0.1502
Cohort 1982-1991	0.5231	0.0100	***	0.1312
Constant	-1.3172	0.0375	***	-0.3305

The base is a male researcher in Social and Humanities, in a private institution, who got his PhD in Mexico or other country between 1991 and 2002.
N=5658, T=1991-1997
*** Significant at 1%, ** significant at 5%, * significant at 10%.

Table 6. Determinants of Research Productivity.

	Model 1		Model 2		Model 3		Model 4		Model 5	
Dependent variable: Publications. Negative Binomail, fixed effects										
Variables										
Time variant variables										
Age	0.2637	***	0.3161	***	0.2743	***	0.3209	***	0.2877	***
	(0.0093)		(0.0090)		(0.0094)		(0.0097)		(0.0096)	
Age Square	-0.0025	***	-0.0029	***	-0.0026	***	-0.0027	***	-0.0027	***
	(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Level 3	0.395	***		***	0.4842	***	0.9446	***	0.5683	***
	(0.0475)				(0.0486)		(0.0503)		(0.0489)	
Level 2	0.4624	***		***	0.5205	***	0.7572	***	0.5516	***
	(0.0304)				(0.0310)		(0.0309)		(0.0308)	
Level 1	0.4302	***		***	0.4507	***	0.4829	***	0.4474	***
	(0.0198)				(0.0198)		(0.0194)		(0.0196)	
Total # of pub	0.0005	***	0.0006	***	0.0005	***	0.0008	***	0.0005	***
	(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Total # of researchers	-0.0003	***	-0.0002	***	-0.0003	***	-0.0003	***	-0.0003	***
	(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Budget -1 year in	0.0009	***	0.0008	***	0.0009	***	0.0005	***	0.0009	***
Millions of dollars (2001)	(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Depreciated knowledge					-0.0020	***				
					(0.0002)					
Prior publications							-0.0826	***		
							(0.0023)			
Prior publications and citations									-0.0024	***
									(0.0002)	
Constant	-5.9628	***	-7.0731	***	-6.108	***	-7.0303	***	-6.3161	***
	(0.2117)		(0.2060)		(0.2140)		(0.2219)		(0.2169)	
Log likelihood	-51784.996		-52032.292		-51741.966		-51057.696		-51667.447	

Standard errors in parentheses
*** Significant at 1%, ** significant at 5%, * significant at 10%.
N=7793, T=1991-2001
We use a depreciated rate of 30%, which is the result of an analysis of the number of citations over time
Prior publications considers 10 years of past publications.
Prior publications and citations considers 10 years of past publications plus citations

Table 6. Continuation

	Model 1		Model 2		Model 3		Model 4		Model 5	
Dependent variable: Publications plus citations per 4 years. Negative Binomail, fixed effects										
Variables										
Time variant variables										
Age	0.1446	***	0.2095	***	0.1458	***	0.1364	***	0.1447	***
	(0.0108)		(0.0102)		(0.0107)		(0.0107)		(0.0107)	
Age Square	-0.0014	***	-0.0019	***	-0.0014	***	-0.0013	***	-0.0014	***
	(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Level 3	1.1161	***		***	1.1435	***	0.9262	***	1.1179	***
	(0.0584)				(0.0619)		(0.0648)		(0.0628)	
Level 2	0.8348	***		***	0.8494	***	0.7381	***	0.8357	***
	(0.0419)				(0.0433)		(0.0443)		(0.0443)	
Level 1	0.5699	***		***	0.5749	***	0.5420	***	0.5702	***
	(0.0287)				(0.0289)		(0.0289)		(0.0288)	
Total # of pub	0.0007	***	0.0007	***	0.0007	***	0.0006	***	0.0007	***
	(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Total # of researchers	-0.0002	***	-0.0001	***	-0.0002	***	-0.0002	***	-0.0002	***
	(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Budget -1 year in Millions of dollars (2001)	0.0012	***	0.0011	***	0.0012	***	0.0012	***	0.0001	***
	(0.0001)		(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Depreciated knowledge					-0.0003					
					(0.0002)					
Prior publications							0.0182	***		
							(0.0027)			
Prior publications and citations									-0.0001	
									(0.0001)	
Constant	-5.9628	***	-7.5011	***	-6.0183	***	-5.8278	***	-6.0011	***
	(0.2336)		(0.2255)		(0.2342)		(0.2241)		(0.2343)	
Log likelihood	- 39076.860		- 39393.163		- 39075.964		- 39055.151		- 39076.857	

Standard errors in parentheses
*** Significant at 1%, ** significant at 5%, * significant at 10%.
N=5658, T=1991-1997
We use a depreciated rate of 30%, which is the result of an analysis of the number of citations over time
Prior publications considers 10 years of past publications.
Prior publications and citations considers 10 years of past publications plus citations