Does Gender Affect Scientific Productivity?

A Critical Review of the Empirical Evidence and a Panel Data Econometric Analysis for French Physicists

Jacques Mairesse*
Michele Pezzoni**

The paper revisits the issue of the gender gap in scientific productivity, often referred as the "productivity puzzle" by economists and sociologists of science. After providing a critical review of the empirical evidence, the paper presents the relevant results of a panel data econometric analysis of individual publication productivity differences for two large samples of French CNRS *and university physicists. We observe in both cases that the productivity of women in terms of number of publications is, by about one third in average, largely lower than that of men. We conclude, however, that female physicists appear as productive as their male colleagues in* CNRS*, and even more productive in French universities, when we take into account several factors, in particular unequal chances of promotion and frequent non-publishing spells, which can reflect strong family engagements.*

LES FEMMES SONT-ELLES MOINS PRODUCTIVES EN RECHERCHE ? UN SURVOL CRITIQUE DES ÉTUDES EMPIRIQUES ET UNE ANALYSE ÉCONOMÉTRIQUE SUR DONNÉES DE PANEL POUR LES PHYSICIENS DU CNRS ET DE L'UNIVERSITÉ EN FRANCE

Dans cette étude, nous reconsidérons la question de la moindre productivité scientifique des femmes, souvent désignée comme celle du « gender gap » ou du « productivity puzzle » par les économistes et sociologues de la science. Après une revue critique des travaux empiriques sur cette question, nous présentons les résultats d'une analyse économétrique conduite en parallèle sur deux échantillons de panels représentatifs des chercheurs en physique du CNRS *et des enseignant-chercheurs en physique des universités françaises. Nous observons que, dans les*

* Maastricht University (UNU-MERIT), CREST-ENSAE, EHESS and NBER. *Address:* CREST-ENSAE, 18, boulevard Gabriel Péri, 92245, Malakoff cedex. *Email:* mairesse@ensae.fr

** Université de Nice Sophia Antipolis (GREDEG) and Università Bocconi (CRIOS). *Address:* GREDEG Batiment 2, 2250, rue Albert Einstein 06560 Valbonne. *Email:* michele.pezzoni@unice.fr

We have benefitted from comments by participants to several seminars and conferences in the last three years, in particular: the "International RESUP Conference on Reforming Higher Education and Research" (Paris, 2011), the "International Conference on Intellectual and Institutional Innovation in Science" (Berlin, 2012), the "Conference on Micro Evidence on Innovation and Development" (Cape Town, 2012), "The Organisation, Economics and Policy of Scientific Research" (Torino, 2013), "CHER Annual Conference" (Lausanne, 2013), the "Biennial Atlanta Conference on Science and Innovation Policy" (Atlanta, 2013). We are especially thankful to James Evans, Thomas Heinze, Francesco Lissoni, Christine Musselin, Anne Pepin, Paula Stephan, and Fabiana Visentin for helpful remarks.

deux cas, la productivité des physiciennes en termes de publication est largement inférieure, d'environ un tiers en moyenne, par rapport à celle de leurs collègues masculins. Nous concluons cependant que cette différence de productivité disparaît pour le CNRS *et qu'elle s'inverse même pour les universités quand nous tenons compte de plusieurs facteurs, notamment ceux liés à des chances inégales de promotion et à des discontinuités notables dans les publications, qui peuvent refléter de forts engagements familiaux.*

JEL Codes*:* C23, I23, J40

INTRODUCTION

The lower productivity of women in science is a long-lasting research issue that has been explored in the past thirty years by economists and sociologists of science who often refer to it as the "gender productivity gap" or "gender productivity bias," or simply the "productivity puzzle" (Cole and Zuckerman [1984]). There is a wide unanimity in literature about the fact that the measured productivity is lower for female than for male scientists, almost in any discipline and whatever the productivity measure considered.[1] Why this is so, however, is still debated and largely remains a puzzle.

In the first part of our paper we propose a short critical review of the empirical evidence in the economic and sociological literature, based on a collection of papers that not only assess the extent of the gender productivity gap, but also investigate some of its possible determinants. We organize our review under the four general themes of family engagements, career path, personal characteristics and work environment, which have particularly attracted the attention of scholars, and summarize in Tables 1 and 2 the results we have found in selected articles on the effects of gender, motherhood and marital status on scientific productivity and on scientific career.

In the second part of the paper we present the relevant results of a panel data econometric analysis of individual publication productivity differences for two large samples of French CNRS and university physicists. We observe a large and significant average productivity gap for women relative to men in terms of number of publications (weighted or not by the five-year impact factor of the journals) for both CNRS and universities over our study period 1982-2005. We also observe two facts that can be viewed as two sources of differences in accounting for the publication gender gap under similar conditions for male and female physicists. Women are less likely than men to be promoted to higher ranks, even controlling for past publication productivity, and they have time spells with no (or very low-quality) publications that are more frequent than men, a fact that could be in part related to stronger engagements in time consuming activities external to research, such as motherhood, child care and other family and household duties. A striking and original contribution of our econometric analysis follows: if we take into account the probability of promotion and the probability

1. The scientific productivity measures considered by scholars are usually based on bibliometric data, and refer to simple publication counts or publications weighted by average impact factors based on number of citations. In addition to publications, scholars tend increasingly to consider other outcome measures, such as patents, training and placement of students.

of having no (or very low-quality) publishing time spells, we find that female physicists appear as productive as their male colleagues in CNRS, and even more productive in French universities. These results remain if we control, as we do, for coauthorship and to initial conditions that proxy for unobserved factors of productivity. Our findings are also robust to specific analyses that focus on differences on publication quality.

A CRITICAL REVIEW OF THE LITERATURE

Although many studies have documented the "gender productivity gap" in science, much fewer have investigated its possible causes and have been successful in assessing empirically to what extent different factors can account for it. Somewhat surprisingly, the puzzle still largely remains.

In their book *Who succeeds in Science? The gender dimension*, Sonnert and Holton [1995] suggest that the explanations for women lower productivity can be classified in two categories associated with what they call the "*difference"* and the "*deficit*" models. The difference model states that women act differently because they are different from men, in particular with respect to motivation and commitment to scientific career. These differences could *a priori* be innate and/or social and cultural. Various studies in sociology have disposed of the idea of innate differences and stressed the importance of cultural and social influences on women to make certain educational choices, to allocate their time between research and family activities, and to take decisions about their career.

By contrast, the deficit model states that, although women have the same goals and aspirations as men, they are treated differently. Their lower performance is mainly due to lower opportunities, higher difficulties in their career, more problems in raising research funds and in collaborating with other scientists. Such obstacles prevent women from having similar career advancements as men, and hinder significantly their scientific productivity.

Although Sonnert and Holton [1995] present separately the two models, they rightly and strongly point out that they are fully compatible and that in fact "structural obstacles (barriers that exist as a feature of the social system of science) and internal obstacles (barriers that exist in the form of women's attitude and values) [...] reinforce each other".

In practice, the prime distinction between the factors expected to explain differences between female and male scientists' productivity is whether they are measured and can be considered in a multivariate econometric model or they are not and fall in the mix of unobserved potential direct or indirect causes of these differences. Although motherhood, career status, quality of the work environment and scientist's personal characteristics are in principle all observable and measurable (or can be well approximated or proxied by other variables), the existing literature has essentially focused selectively on a few of these factors, and to our knowledge extensive investigations aiming at controlling for many of them simultaneously have not been implemented, most probably due to a lack of availability within the same corpus of data.

We can list other methodological problems affecting gender studies such as the lack of longitudinal dimension, the tendency to over-generalize the results and the non-representative samples that are often analyzed.

We thus propose here a short review of the results of the empirical studies of the factors that contribute to the gender productivity gap in scientific research under the four following headings: family engagements, career path, scientists' unmeasured personal characteristics, and work environment.

Family Engagements

Family engagements are perhaps the most frequently proposed explanations for the productivity puzzle. Among them, motherhood is of particular interest for scholars because it is an easily identifiable event that may explain temporary shortfalls in the publication productivity of young women. Ward and Wolf-Wendel [2004] in their article on "Academic motherhood: Managing complex roles in research universities" put forward four explanations or theories of why family engagements affect scientific productivity of women.

Their first theory, *role conflict*, states that research and family activities are mutually exclusive and result inevitably in a trade-off in time and effort between the two. Women are expected to be the principal caregivers to children and they have to substitute time dedicated to them to the time for research. The second one, *ideal worker*, centers on the ways scientists tend to work. Research work is endless and requires a considerable effort and flexibility, especially when a scientist is young and is on a tenure-track. There is little time left for a child. This claim is supported by the evidence that, as observed by Zuckerman *et al.* [1991] a high share of women scientists do not have children: "… [the] climate of opinion meant that women determined to have serious research careers did not marry [and did not have children]". Finkel and Olswang [1996] find in their survey that about 30% of women faculties do not have children, half of them declaring this was largely influenced by incompatibilities with their career plans. The third theory, *male clockwork*, stresses that academic career is based on male model assuming freedom from competing responsibilities (such as family duties). Academic career success is based on a consistent and stable publication flow that does not allow for periods of suspended or lowered productivity. Motherhood often happens during the initial stage of the scientists' career, when a productivity slowdown can affect strongly the pursuit of the career. Such a slowdown may result in a long term effect of cumulative (dis)advantage (Merton [1968]; Allison and Stewart [1974]; Allison *et al.* [1982]; David [1994]). It can considerably amplify the difference between women and men publication scores and influence their career paths. Zuckerman *et al.* [1991] refer to this idea as the "theory of limited differences" applied to the productivity puzzle in science.

In their fourth theory and in contrast with the other three, Ward and Wolf-Wendel offer a positive view of the impact of motherhood on productivity. They point out that being involved in *multiple roles*, as scientist and parent, may have positive effects on self-esteem and can also bring about greater social support. Women who have children may have also better "health, energy and stamina than women without children". The idea of a positive effect on productivity of motherhood through self-selection is supported by Fox and Faver [1985] and

Fox [2005]. An alternative explanation, also proposed by Fox and Fever, is the tendency for women scientists to have children when their career paths are well established and when taking care of children will have less consequences. Finally, Stack [2004] makes the point that raising children could motivate scientists, women and men alike, to work harder in order to provide them better living conditions.

The evidence provided by the empirical literature is as mixed than the survey observations and a priori reflections. The effects of family engagements are not very strong and often disappear when scientists' personal characteristics, discipline specificities, work environment are taken into account. Moreover several studies, as reported in Table 1, compare in fact the effects of motherhood on productivity between women with and without children, but not with men.

Long [1990], Kyvik [1990], and Kyvik and Teigen [1996] find a significant negative impact of having young children on women's scientific productivity, while Toren [1991] finds no effect. Fox and Faver [1985] find on the contrary that women with young children are more productive than women with no children. The last result is in line with the one found by Fox [2005] for whom, women with pre-school children show an higher average productivity if compared to women without children or with school-aged children. Zuckerman *et al.* [1991], on the basis of their analysis of the publication records of one hundred and twenty eminent scientists actually conclude that "[…] science and motherhood do mix […]" with little effects on productivity.

Another family related factor that has been often analyzed in literature is the researcher's marital status and the spouse occupation. Once more the empirical evidence is mixed. Fox and Faver [1985] do not find any effect of being married. However, Mary Frank Fox [2005], with more detailed information on the family status, finds that the type of marriage and the spouse occupation are important determinants. According to Fox, subsequent marriages or being married with another scientist have a positive effect on female scientist's productivity. Sax *et al.* [2002] find a positive effect of marriage as well. Long [1990] find a positive (indirect) effect of marriage on productivity at the time of the completion of the doctoral training. The positive effect is determined by the higher chances of collaboration with the mentor when the female scientist is married. Zuckerman *et al.* [1991] find a little impact of being married on scientific productivity.

Studies listed in Table 1 show that the most likely result in gender studies is a significant bias in favor of men productivity and that controlling for the presence of children and for the marital status is usually not enough to explain the gender gap.

Information on child and motherhood are often obtained by mean of *ad-hoc* questionnaires. This strategy of data collection often force researchers to work on relatively small cross-sectional samples. An alternative approach indirectly suggested by Long [1992] is to consider maternity leaves as possible causes of prolonged non-publishing time spells. To the best of our knowledge no one before Long has focused on the effect of prolonged non-publishing time spells as a possible explanation of the productivity puzzle. According to him the reasons for not publishing for long periods can be mainly three: no papers are submitted or accepted for publications during that period; scientist may be active in research but may not be in publication; scientists may no longer be active in science for

Table 1. *A selection of literature contributions on the effects of gender, motherhood and marital status on scientific productivity*

Article / book	Year(s) of analysis	Cross-sectional / longitudinal data (sample size)	Type of analysis: promotion / productivity	Gender (only in case of W-M comparison)	Motherhood	Marital status	Discipline (Country)	Control list
(Fox and Faver, 1985)	1981	Cross-sectional (300)	Productivity	Negative effect	Positive and significant effect of having young children	No effect	Social work academics (US)	Personal Characteristics; Rank; Location; Research Collaboration;
(Long,1990)	1950-1967	Longitudinal (1159)	Productivity	Negative effect	Negative effect	Positive effect	Biochemistry (US)	Research Collaboration; Location; PhD Characteristics
(Kyvik,1990)	1982	Cross-sectional (1569)	Productivity	Negative effect	Negative effect of having young children	No effect	Cross discipline study based on in 4 universities (Norway)	Rank; Personal Characteristics;
(Zuckerman *et al.*, 1991)	Samples from PhD cohorts: 1920-59 / 1960-69 / 1970-79	Longitudinal (120)	Productivity	NA (Focus only on women)	Effect slightly negative over whole career but increasing after the birth of a child	Little/no effect	Mathematics; biology; physics; economics; psychology (US)	No Controls
(Toren, 1991)	1983	Cross-sectional (42)	Productivity	NA (Focus only on women)	No effect	No effect	Natural sciences; humanities and social sciences (Israel)	No controls
(Long, 1992)	1956-1963	Longitudinal (1159)	Productivity	Negative effect	NA	NA	Biochemistry (US)	No controls

Article / book	Year(s) of analysis	Cross-sectional / longitudinal data (sample size)	Type of analysis: promotion / productivity	Gender (only in case of W-M comparison)	Motherhood	Marital status	Discipline (Country)	Control list
(Kyvik and Teigen, 1996)	1992	Cross-sectional (1277)	Productivity	Negative effect	Negative effect of having young children	NA	27 disciplines including social, natural, humanities and medical sciences (Norway)	Personal Characteristics; Rank; Research Collaboration;
(Shauman and Xie, 1996)	1969; 1973; 1988; 1993	4 Cross-sectionals (14000; 9000; 1000; 3000)	Productivity	Small negative effect	NA	NA	Biology; engineering; mathematics; physics; social sciences (US)	Personal Characteristics; Rank; Funding Resources
(Sax *et al.*, 2002)	1998-1999	Cross-sectional (8544)	Productivity	Negative effect	Little/no effect	Weak positive effect of being married with a scientist	Cross discipline study based on in 57 universities (US)	Personal Characteristics; Rank; Location; Salary; Discipline;
(Stack, 2004)	1995	Cross-Sectional (11231)	Productivity	Negative effect	Positive and significant effect of having young children (except for pre-school children)	NA	Biology; mathematics; engineering; medical sciences; physics (US)	Personal Characteristics; Rank; Structural Factors; Location;
(Fox, 2005)	1993-1994	Cross-sectional (1215)	Productivity	Negative effect	Positive effect of pre-school children	Positive effect of being married with a scientist	Computer science; chemistry; electrical engineering; microbiology; physics (US)	No controls

Article / book	Year(s) of analysis	Cross-sectional / longitudinal data (sample size)	Type of analysis: promotion / productivity	Gender (only in case of W-M comparison)	Motherhood	Marital status	Discipline (Country)	Control list
(Turner and Mairesse, 2005)	1986-1997	Longitudinal (497)	Productivity	Negative effect	NA	NA	Physics (FR)	Personal Characteristics; Rank; Structural Factors; Location; Research Collaboration; Mobility Indicators;
(Fox and Mohapatra, 2007)	1993-1994	Cross-sectional (1215)	Productivity	No effect (effect is given by team composition)	NA	NA	Computer science; chemistry; electrical engineering; microbiology; physics (US)	Research Collaboration; Work Practices; Departmental Climate;
(Lissoni *et al.*, 2011)	1975-2005	Longitudinal (3600)	Productivity	Negative effect	NA	NA	Physics (FR,IT)	Personal characteristics; Rank; Structural Factors; Research Collaboration; Discipline;

several reasons including family engagements. Long shows that, in the field of biochemistry, overall differences in publications between women and men pertaining to the same 1956 cohort are consistently reduced if only scientists who actively publish are considered. In the empirical analysis reported in section 3, we relate prolonged non-publishing time spells to possible maternity leaves of young women. We observe a substantial reduction in the productivity puzzle when we measure productivity by excluding prolonged stops in publications.[2]

Career Path

Although the productivity puzzle has been widely studied, it is not the only puzzle connected to gender. Similarly to productivity, scholars have shown that women are rewarded less than men for their research achievements. Women with comparable level of scientific productivity and reputation, have lower wages and their career advancement takes longer time than men (Fox, [1981]; Levin and Stephan [1998]; Long *et al.* [1993]; Pezzoni *et al.* [2012]). In several European academic systems, such as in France and Italy, wages are not negotiable and promotion to higher ranks constitutes a major component of the economic reward for scientists. Ideally the only criterion to assign the reward should be based on scientist's merit, often proxied by the scientist's publication score (Dasgupta and David [1994]; Merton [1973]). Any other determinant such as gender, is a signal that the scientific community is allocating the reward differently from the ideal model. Table 2 lists five articles assessing gender effect on career advancements. Four out of five studies show a negative impact of gender on the chances of experiencing a career advancement (Corley *et al.* [2003]; Long *et al.* [1993]; Pezzoni *et al.* [2012]; Sabatier *et al.* [2006]; Wolfinger *et al.* [2008]).

Difficulties in promotion for women have an indirect impact on productivity by reducing the available resources for research, their prestige and their influence. At the same time, lower productivity decreases the chances to be promoted to higher ranks. This bidirectional causal relation between promotion and productivity raises an endogeneity problem. Several articles on gender gap have identified career status as an important determinant of productivity but without taking into account the endogeneity issue (*eg.* Fox and Faver [1985]; Sax *et al.* [2002]).

Scientists' Unmeasured Personal Characteristics

Although past studies are characterized by information richness in terms of variables and controls, they suffer from a substantial lack of longitudinal data (with some notable exceptions, such as Long [1992], [1990]; Long *et al.* [1993]; Turner and Mairesse [2005]). As a result none of the cross-sectional studies allow accounting for the unobserved (time-invariant) characteristics of

2. In the empirical analysis reported in section 3 we observe only a selected sample of 360 high quality journals in physics. Thus, from now on we will adopt a more appropriate terminology by defining non-publication periods as low-quality-publishing time spells, given that we do not observe papers published in lower quality journals in physics.

Table 2. *A selection of literature contributions on the effects of gender, motherhood and marital status on scientific career*

Article / book	Year(s) of analysis	Cross-sectional / longitudinal data (sample size)	Type of analysis: promotion / productivity	Gender (only in case of W-M comparison)	Motherhood	Marital status	Discipline (Country)	Controls list
(Long *et al.*, 1993)	1950-1967	Longitudinal (1006)	Promotion	Negative effect	Negative effect (although not significant) for lower ranks	Positive effect for lower ranks	Biochemistry (US)	Location; Past Productivity; Personal Characteristics
(Corley *et al.*, 2003)	1996	Cross-section (412)	Promotion	No effect	NA	NA	Biotechnology; biochemistry; bioengineering; microelectronics (US)	Seniority; Grant; Productivity
(Sabatier *et al.*, 2006)	1990-2002	Longitudinal (583)	Promotion	Negative effect	NA	NA	Biology (France)	Productivity; Personal Characteristics; fund raising activities; administrative activities; geographical mobility
(Wolfinger *et al.*, 2008)	1981-1995	Longitudinal (more than 30000)	Promotion	Negative effect	Negative effect of having young children for lower ranks	Negative for lower ranks	Humanities; social sciences; hard sciences; engineering; mathematics; physics (US)	Ethnicity; Personal Characteristics; Discipline; PhD Performance
(Pezzoni *et al.*, 2012)	2000-2005	Cross-section (1282)	Promotion	Negative effect	NA	NA	Physics (FR,IT)	Research Collaboration; Productivity; Structural Factors; Discipline;

the scientist, such as intrinsic ability, capacity and effort. Kyvik [1990] and Long and Fox [1995] highlights the need for longitudinal data in gender related research in order to overcome this limitation.

Another problem of cross-sectional datasets is the lag between the time when the research activity is conducted and the publication of its results on journals. Several studies are based on surveys asking for a self-declaration of publication scores, usually during the 2 or 3 years before the questionnaire. Unfortunately, the productivity measured in the questionnaire could be the results of previous research efforts. Then, the productivity of a scientist may not be affected by her (time-variant) characteristics at the time of the survey. For instance, the effect of having a child may take more time to be visible, especially in disciplines where the time span between research and publication of the results is not the one assumed by the authors. Past studies simply ignore the problem of estimating a reasonable publication time-lag of the research work. When studying scientists belonging to different disciplines often the same presumed time-lag is applied.

Work Environment

We found in our literature review at least four factors related to the work environment that may affect differently women and men: the discipline characteristics, the specificities of the studied countries, the specificities of the institutions and the propensity to collaboration of the individuals.

Discipline specificities may affect productivity for several reasons directly or indirectly related to the scientist's gender. To what concern reasons directly related to gender, the deficit model may be at play in historically men-centered disciplines. Women may face more difficulties in becoming part of the scientific community, in publishing on good journals or entering prestigious institutions, due to discrimination –Sonnert and Holton [1995] and Zuckerman *et al.* [1991] provide several examples. If the difference model is at play, women may behave differently by self-selecting for not entering disciplines socially considered as a male realm.

To what concern reasons indirectly related to gender, Stack [2004] shows, in a brief literature review, that studies in hard disciplines are more likely to find a negative relations between productivity and having children. In disciplines characterized by wider presence of women (like social sciences) the effect measured is weaker. The explanation, according to Stack, is related to the well-developed women research network in social sciences that may, for instance, compensate the potential costs in terms of time and effort due to children care. The large presence of women results in an internal disciplinary regulation that does not discriminate scientists with child caring duties. The usual career paths in the discipline do not penalize the individuals who take prolonged stops in research activity due to motherhood, allowing for a career model that does not coincide with the *male clockwork*. Table 1 and 2 show a large prevalence of studies in hard disciplines with few contributions of studies analyzing gender productivity gap in humanities and social sciences.

As well as the coverage of the disciplines, also the geographical distribution of the studies is strongly biased. The large part of the papers in literature are based on the evaluation of gender productivity gap in the US research system. In particular the "*Survey of Doctorate Recipients*" has been often used as source of data (Fox and Stephan [2001]; Levin and Stephan [1998]; Stack [2004]. Results from US-centric studies in terms of productivity and careers can be hardly generalized to the French Universities and CNRS, analyzed in section 3. French research system is more centralized and French institutions are characterized by a substantial lack of autonomy both in terms of funding and in terms of employing scientists (Lissoni *et al.* [2011]). CNRS conducts the large part of the national research, leaving to universities the educational tasks. Academics and CNRS scientists are considered civil-servants and their contracts are centrally regulated by the Ministry of Education. US universities, differently from French universities, freely bargain contracts with academics and are explicitly stratified in teaching and research universities. In such a different institutional scenario, gender impact may be affected by recruiting procedures and by the attractiveness of the private labor market.

Countries and organizations may also implement different policies in favor of women, aiming at limiting the gender productivity gap. Strong gender policies may mitigate the effects of family duties on scientific productivity. Analysis of the effects of the policies in favor of women are still a largely unexplored field in economics and sociology of science. In the second part of the paper we compare the gender effect accounting for the specificities of two different organizations: French Universities and CNRS. Although the comparison does not tell us about the policies in favor of women in place in the two organizations, it tells us what is the extent of gender productivity gap in the two different organizations.

Collaboration of scientists, specialization and large research teams have become in recent years the most common paradigm for scientific research, especially in hard sciences (Beaver and Rosen [1978]; Katz and Martin [1997]; Price [1963]). Reasons for collaborating are many: scientific, technical (material resources, big research tools, specialized skills and knowledge) or simply practical (related to the costs of communication and travelling). Collaboration has been included in the list of possible gender related determinants of productivity. Fox and Mohapatra [2007] show that, in hard disciplines, collaboration enhances productivity, especially when collaboration is outside the scientist's department. Research productivity usually benefits from the flow of new external knowledge and skills. Fox and Mohapatra assess the gender composition of the team in relation to the gender of the scientist, finding statistical evidence that men's productivity increases when the research team is composed by a majority of male graduate students, while teams characterized by an intensive presence of women (or mixed) do not show similar effects. Kyvik and Teigen [1996] find that research collaboration has a positive effect for women, and less for men. Fox and Faver [1985] measure the connections of the scientist to the scientific community with the appointment to an editorial board of a scientific journal. They find that the appointment enhances significantly scientific productivity both of women and men. Zuckerman *et al.* [1991], on the basis of interviews, point out that the publication record upsurge, sometimes observed for eminent female scientists in late career, might be related to the opportunities for collaborative research that are more frequent in later stages of the career.

Empirical results suggest that there is strict relation between gender and collaboration, although the approaches available in literature are heterogeneous and not directly comparable. Lower intensity of collaboration for women could be one of the causes of the productivity puzzle. However, the reason why women should collaborate less is not clear. Apart from the possible discrimination effect another cause of lower collaboration attitude may be identified in the mobility limitation due to family obligations (Marwell *et al.* [1979]; Shauman and Xie [1996]). Having children and family care duties, for instance, may prevent them from joining the best research groups possibly located elsewhere. In the second part of this paper we include research network specificities and characteristics of work environment as controls.

A PANEL DATA ECONOMETRIC ANALYSIS

Data, Sample and Descriptive Statistics

We base our analysis on a longitudinal database including all the publications of all physicists, active in academic year 2004/2005, in CNRS and French universities, respectively 159 (15.9%) and 839 (84.1%) female and male CNRS physicists, and 410 (22.6%) and 1403 (77.4%) female and male university physicists.[3] We measure their publication productivity as the weighted sum of the articles they published over three year periods or triplets, taking as weights the five years impact factors of the journals in which these articles are published. If we label productivity, number of articles and average impact factor by respectively *Prod*, *Art* and *IF*, we simply have in logarithms *log(Prod) = log(Art) + log(IF)*. As we explained below, we will not count in our productivity measure the articles published in journals with very low visibility, defined as journal with a five year impact factor of less than 0.5.

We have little but important personal information on individual university and CNRS physicists: mainly their dates of birth and hence their age, their gender, their ranks in 2004/2005 and seniority in these ranks. The two main ranks in French universities are "*Professeur"* (*i.e.*, professor, or for short PR) and "*Maître de Conférences"* (*i.e.*, lecturer or MCF), while their close counterparts in terms of careers in CNRS are "*Directeur de Recherche"* (*i.e.*, research director or DR) and "*Chargé de Recherche"* (*i.e.* research associate or CR). We know when university physicists who are professors (PR) in 2004/2005 have been promoted from lecturers (MCF), and when those who are lecturers in 2004/2005 were recruited. Similarly we know when CNRS physicists who are research directors (DR) in 2004/2005 have been promoted from associate researchers

3. We cover all fields of physics, with the exception of *astronomy physics* and *nuclear physics*, and we use here the word physicist to designate non-*astro* and non-*nuclear* physicists. The sub-field classification is different in French universities and CNRS. For universities we thus consider section 28: Condensed matter and materials and section 30: Diluted matter and optics, and exclude section 29: Elementary constituents. For CNRS we keep section 2: Physical theories: methods, models and applications, section 4: Atoms and molecules-Lasers and optics- Hot plasmas, section 5: Condensed matter physics: structure and dynamics, and section 6: Condensed matter physics: structures and electronic properties, and we excluded section 3: Particles, nuclei, interactions: from the laboratory to the cosmos. Although physicists usually publish in their own subfield and tend to collaborate with colleagues in their sub-field, they also publish in other sub-fields and frequently collaborate with colleagues from other sub-fields.

(CR) and when CNRS associate researchers in 2004/2005 were recruited. Table 3 shows that women, especially older women, are underrepresented in higher ranks in both organizations. The proportions of female physicists of more than forty in 2004/2005 who are DR or PR are 47% at CNRS and 23% in the universities, while these proportions for male physicists are of 67% and 51% respectively.

Table 3. *Number and proportion of female and male CNRS and university physicists in two age groups and low and high rank for the last triplet 2003-2005*

PHYSICISTS	WOMEN-CNRS	MEN-CNRS	CNRS	WOMEN-UNIV	MEN-UNIV	UNIV
Less than 40 years						
Low Ranks (CR/MCF)	34 *100%*	182 *94%*	216 *94%*	136 *99%*	323 *93%*	459 *95%*
High Ranks (DR/PR)	0 *0%*	11 *6%*	11 *6%*	2 *1%*	26 *7%*	28 *5%*
Sub Total	34 *100%*	193 *100%*	227 *100%*	138 *100%*	349 *100%*	487 *100%*
40 years and more						
Low Ranks (CR/MCF)	66 *53%*	213 *33%*	279 *36%*	208 *76%*	509 *48%*	717 *54%*
High Ranks (DR/PR)	59 *47%*	433 *67%*	492 *64%*	64 *24%*	545 *52%*	609 *46%*
Sub Total	125 *100%*	646 *100%*	*771* *100%*	272 *100%*	1,054 *100%*	*1,326* *100%*
All						
Low Ranks (CR/MCF)	100 *63%*	395 *47%*	495 *50%*	344 *84%*	832 *59%*	1176 *65%*
High Ranks (DR/PR)	59 *37%*	444 *53%*	503 *50%*	66 *16%*	571 *41%*	637 *35%*
Total	159 *100%*	839 *100%*	998 *100%*	410 *100%*	1,403 *100%*	1,813 *100%*

We gathered the information on scientific publications from the Institute for Scientific Information (ISI) Web of Science, by matching the last names and first names of the physicists with the last names and initial of first names (as coded by ISI) for the authors of all articles published in physics journals going back to 1975. We thus obtained respectively some 39,500 and 35,000 distinct scientific articles for CNRS and the universities for a set 368 journals. These journals include all the main journals classified as being mainly in physics by ISI. We also required have more than a minimum "quality" or "visibility" threshold, defined by an average five year impact factor of at least 0.5. In other words we excluded all the physics journals receiving in average per article less than

0.5 citations over five years.[4] We have also replicated our econometric analysis using two different quality thresholds, one high and the other very high, defined by a five years average impact factor of respectively 5 and 10. As documented in Appendix B, we basically find very similar qualitative results than the ones we obtain and present here.

In spite of matching the names of our lists of CNRS and French university physicists to those of authors of papers in physics journals only, we could not fully avoid the problem of homonymy. Keeping homonyms could be a source of biases for an analysis like ours, if for some reason that would generate a spurious correlation between lower or higher productivity and the spelling of names. We dealt with homonymy (as we did in Lissoni *et al.* [2011]) in two ways: first by removing the few physicists that were homonyms in the two lists of CNRS and university physicists, then by removing those physicists that appeared to have a productivity record far too high for being real and thus were most certainly revealing an homonymy with other physicists not in our CNRS and university lists.[5]

Based on our publication data we constructed two unbalanced panels, respectively for CNRS and university physicists, in which the oldest scientists (say born before 1950 and more than 55 in 2005) are present from 1975 to 2005 and the younger ones from their year of first recorded publication to 2005. The time lag between the submission and publication dates of articles being variable, we thought appropriate (as in Long [1992]), to measure their publication productivity over three years "or triplet." This is long enough to smooth the time profile of productivity satisfactorily and to define significant spells with no publications, nonetheless short enough to keep a large number of observations and preserve a reasonable time dimension for our two panels. We had, however, to exclude from them the triplet 1975–1978 for which the publication data appeared sparse and not too reliable. We also decided to set apart for all physicists the triplets when we first see them, in order to control for individual initial conditions in our econometric analysis by using their productivity as measured in these first triplets.[6] As shown Table A1 of Appendix A, our unbalanced panels are thus made at most of the eight triplets from (1982–1984) to (2003–2005) and at least of the two triplets (2000–2002) and (2003–2005).

Table 4 reports the average publication productivity per triplet of CNRS and French university physicists separately for women and men, as well median, standard deviations and corresponding number of observations. In the upper panel, these statistics are computed for all triplets, that is including the

4. There are about 100 physics journals in the Web of science with a five years average impact less than 0.5, corresponding respectively to some 6,200 and 4,400 distinct articles for CNRS and the universities (*i.e.* 15% and 13% of the ones we consider). Note that Nature and Science are not included in our set of physics journals. Their scope is much larger than physics, and a paper in physics published in them is usually developed in details in one or more papers in a physics journal.

5. In addition to the some 400 CNRS and 450 university nuclear physicists in 2004-2005, we had to remove about 20 homonyms in our lists of CNRS and university physicists. We had also to discard about five CNRS and five university physicists, because their computed productivity appeared much too large to be real (with 30 papers or more per year for at least 10 years!).

6. In practice we had to consider in our analysis that physicists' careers started from the first triplets when we observe they have published before we know they were recruited, or alternatively that they started from the triplets when they were recruited, if we do not observe they have published before.

non-publishing triplets. These are the triplets with truly zero articles or with only articles in journals with a very low five years impact factor of less than 0.5, which we deemed better and simpler to treat as zero articles in our analysis. The middle panel of Table 4 reports the same simple descriptive statistics of productivity per triplet, but excluding the subset of non-publishing triplets, while the lower panel shows also these statistics for productivity measured in logarithms and not in absolute levels. Taking logarithms excludes the non-publishing triplets as well, but has the advantage of greatly normalizing the statistical distribution of productivity, which is in general much preferable in econometric analyses.

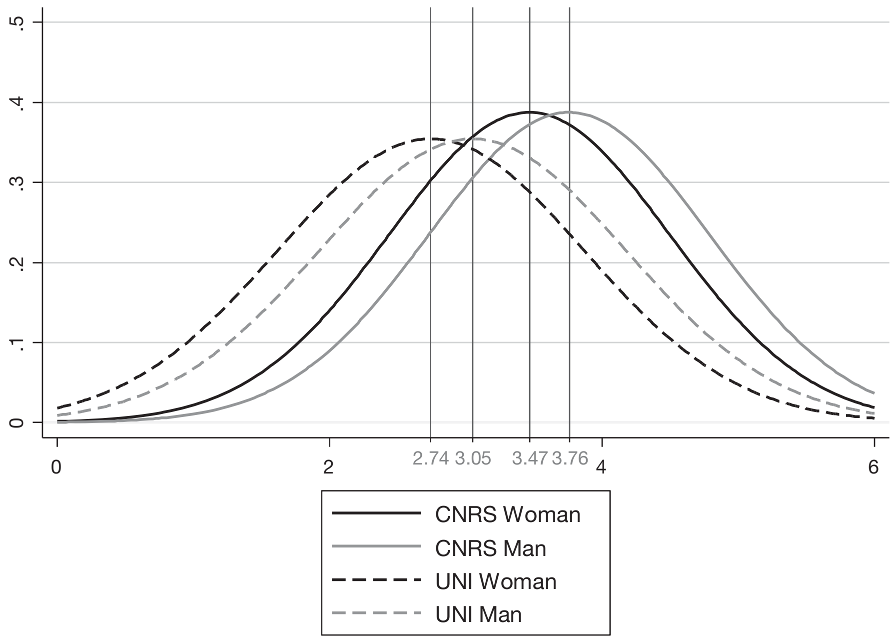
Table 4. *Descriptive statistics on average publication productivity for female and male CNRS and university physicists, including and excluding non-publishing triplets*

PHYSICISTS	WOMEN-CNRS	MEN-CNRS	W/M (or W-M in logs)	WOMEN-UNIV	MEN-UNIV	W/M (or W-M in logs)
Including non-publishing triplets						
Mean	38.58	59.96	0.65	13.14	27.15	0.49
Median	27.08	43.64	0.62	0	12.8	–
td dev.	45.63	60.83		45.38	41.05	
Obs.	815	4,702		1,848	7,408	
Excluding non-publishing triplets						
Mean	48.52	65.99	0.74	30.2	37.13	0.81
Median	38.48	49.15	0.78	16.72	22.74	0.74
td dev.	46.23	60.61		64.97	43.98	
Obs.	648	4,272		804	5415	
In logaritms (excluding non-publishing triplets)						
Mean	3.47	3.76	– 0.29	2.74	3.05	– 0.31
Median	3.65	3.89	– 0.24	2.82	3.12	– 0.30
td dev.	1	1.03		1.11	1.13	
Obs.	648	4,272		804	5,415	

Looking first at the upper panel, we observe that the average publication productivity per triplet for female CNRS physicists is equal to 38.6 impact factor weighted number of articles, or 3.9 in terms of equivalent number of articles per triplet in journals with impact factor 10. It is lower by about one third (35%) than the average productivity for their male CNRS colleagues of 6.0 equivalent articles. The average productivity gender gap is even much wider in French universities, being of about half (50%), with an average productivity of 1.3 equivalent articles for female physicists as against 2.7 for their male colleagues. The median of productivity for female university physicists is actually zero, implying that in average they publish in less than 50% of triplets. The evidence given in the middle and lower panels, which filter out the non-publishing triplets, shows gender productivity gaps that are not as striking but still very large. They are in absolute levels of about 25% (down from 35%) in CNRS and 20% (down from 50%) in universities. In logarithms, the log-differences between female and male CNRS and university physicists are of the same order of magnitude, about – 0.30.[7]

Beyond the average evidence, it is important to stress and keep in mind that the distributions of individual productivity are greatly dispersed, as shown in table 4 by the wide standard deviations in absolute levels and logarithms. The implication is that a large proportion of female physicists are actually more productive than their male colleagues, although in average they appear very significantly less productive. As illustrated by Graph 1, publication productivity in CNRS is higher for 45% of women than the average productivity of men, while it is lower for 34% of men than the average productivity of women. The comparable figures for the universities are close: 42% and 38% respectively.[8]

Graph 1. *Distribution of observed log-productivity for female and male CNRS and university physicists**



* The observed log-productivity distributions of Graph 1 suppose for simplicity an approximately normal log-distribution.

7. Note that for both CNRS and the universities the average productivity is much higher than median productivity in the upper panel, as well as in the middle panel, showing that the distribution of productivity in absolute levels is highly skewed to the right. This is not the case in the lower panel where the average is close enough to the median, showing that the log-distribution of productivity is not too dissymmetric, and confirming that log normality can be viewed as a good enough first approximation.

8. Although the log-productivity distributions as drawn in Graph 1 assume for simplicity a log-normal distribution, the percentages given in the text of women (men) more (less) productive than their average man (woman) colleague are the percentages computed on the actual observations.

Table 5. *Proportion of non-publishing triplets, average publication productivity, and total number of observations for female and male CNRS and university physicists, in two age groups and low and high rank for the last triplet 2003–2005*

PHYSICISTS		WOMEN-CNRS	MEN-CNRS	W-M	WOMEN-UNIV	MEN-UNIV	W-M
Less than 40 years							
Low Ranks (CR/MCF)	% of non-pub-triplets	7%	3%	4%	42%	13%	29%
	Average productivity (logs)	3.60	3.83	– 0.23	2.73	3.20	– 0.47
	Obs.	*85*	*473*		*355*	*847*	
High Ranks (DR/PR)	% of non-pub-triplets	–	3%	–	0%	8%	– 8%
	Average productivity (logs)	–	4.33	–	3.26	3.80	– 0.54
	Obs.	*0*	*40*		*7*	*77*	
Sub Total	% of non-pub-triplets	7%	3%	4%	41%	12%	29%
	Average productivity (logs)	3.60	3.87	– 0.27	2.75	3.26	– 0.51
	Obs.	*85*	*513*		*362*	*924*	
40 years and more							
Low Ranks (CR/MCF)	% of non-pub-triplets	28%	15%	13%	64%	45%	19%
	Average productivity (logs)	3.14	3.37	– 0.23	2.63	2.57	0.06
	Obs.	*352*	*1,238*		*1,113*	*2,881*	
High Ranks (DR/PR)	% of non-pub-triplets	17%	8%	9%	50%	16%	34%
	Average productivity (logs)	3.70	3.89	– 0.19	2.98	2.25	0.73
	Obs.	*378*	*2,951*		*373*	*3,603*	
Sub Total	% of non-pub-triplets	22%	10%	12%	60%	29%	31%
	Average productivity (logs)	3.45	3.74	– 0.29	2.74	3.01	– 0.27
	Obs.	*730*	*4,189*		*1,486*	*6,484*	
All							
Low Ranks (CR/MCF)	% of non-pub-triplets	24%	12%	12%	58%	37%	21%
	Average productivity (logs)	3.25	3.51	– 0.26	2.66	2.77	– 0.11
	Obs.	*437*	*1711*		*1,468*	*3,728*	
High Ranks (DR/PR)	% of non-pub-triplets	17%	8%	9%	49%	16%	33%
	Average productivity (logs)	3.70	3.89	– 0.19	2.98	3.26	– 0.28
	Obs.	*378*	*2,991*		*380*	*3,680*	
Total	% of non-pub-triplets	20%	9%	11%	56%	27%	29%
	Average productivity (logs)	3.47	3.76	– 0.29	2.74	3.05	– 0.31
	Obs.	*815*	*4,702*		*1,848*	*7,408*	

The reductions in the average gender productivity gap, when non-publishing triplets are not counted, are much more pronounced for universities than for CNRS, as well as more pronounced for women than for men in both organizations. Such discrepancies are to a large extent likely to reflect different engagements in activities external to research. These are typically teaching activities to which French university academics are normally devoting half of their working time, contrary to their CNRS colleagues who can dedicate themselves fully to research. But they are also non-research and non-teaching activities that can differ with gender, as well as with age and rank status, such as parenthood, child care and typically stronger engagements of women in family and household activities; or increasing responsibilities for senior professors and research directors. Table 5 complements Table 4 by giving the average log productivity, the percentage of non-publishing triplets, and the numbers of observations, for female and male CNRS and university physicists that are in low ranks (CR or MCF) or in high ranks (DR or PR), when less or more than forty years old in the last triplet 2003–2005.

Table 5 shows that, conditional on ranks, both the frequency of non-publishing spells and log-productivity increase with age for men and women in CNRS and universities, though to different degrees. It also appears that, conditional on age, the non-publishing spell frequency decreases from low to high ranks, but log-productivity increases, again for men and women and both organizations and to a different degree. However, since promotion to high rank increases with age and there are relatively few DR and PR younger than forty, trying to disentangle the effects of rank and age on male and female publication productivity and the gender gap is not straightforward. To assess these effects separately we need to rely on an econometric framework with not only a productivity equation but with two other equations, one to account for promotion and the other for non-publishing spells. Being of interest per se, these two equations enable us to correct for the endogeneity of promotion and the selectivity of publishing spells in the productivity equation.

Econometric Framework: Specification and Estimation

Gender is only one of the various determinants of scientific productivity that can be considered. To assess with reasonable confidence the impacts of these determinants, we have to specify a suitable econometric model. We have also to worry about specification errors that can affect this model, since we very well know that the model is not “the true model” (assuming it exists) but an incomplete and approximate one in many respects. There are various types of specification errors that are more or less likely and that can potentially result in more or less severe biases in the estimated parameters of main interest, the gender productivity gap in our case. Even if we cannot control or correct for such biases, the model can be still of great interest if we are aware of them when assessing the magnitude of the estimated parameters of interest and interpreting their meaning.

Our model consists basically of a standard linear regression equation relating publication log-productivity, as defined above, to gender and to other important determinants we can measure in practice. We will focus on gender as well as age, insofar as we estimate significant interactions between them. The productivity regression is associated in our model with two other “probit-type” regressions accounting respectively for promotion to high rank and for non-publishing spells. Going beyond what can be gathered from simple descriptive statistics, we will find that these two equations are key for a better understanding of the meaning of the observed gender productivity gap and a proper evaluation of what we may regard as the “true” gender productivity gap.

The Promotion Probit Equation

Career advancements and scientific productivity are strongly related. Most productive scientists have more chances of being promoted from a lower to a higher rank, and when promoted they have greater opportunities of collaboration and better access to resources helping them to be more productive. This two-way causality is the source of endogeneity biases when including rank as an explanatory variable in the productivity regression, where rank is measured as a binary indicator, noted *Rank*, equal to one for the higher rank (DR or PR) and to zero for the lower rank (CR and MCF). We shall see that these biases affect not only the estimate of the rank coefficient but also the coefficients of other variables, in particular gender and age. We correct here for this problem by estimating the productivity regression by two-stage least squares. In practice as we explain below, this amounts to estimate the productivity regression by ordinary least squares and including, instead of the rank binary indicator, the probability of promotion, noted *Prob(promotion)*, as predicted in the promotion equation.

Our promotion equation is a simple probit equation of the rank binary indicator on two groups of variables and on time dummies. The first group is that of gender, age and age squared, with and without the interactions. Gender is coded by a binary indicator noted *Woman* (equal to one for a woman and zero for a man) and age noted *Age* is measured in years (and centered to 40 years and divided by 10 for an easy reading of the estimates). A second group is that of productivity in the previous triplet: precisely the number of articles (noted *L.Log(Art)* in logarithms), the corresponding average five year impact factor of the journals in which these articles have been published (noted *L.Log(IF)* in logarithms), and a binary indicator of no publication (equal to one or zero in case of a non-publishing or publishing spell, and noted *L.No publication*). Finally, the time dummies take care of unobserved general factors. Among them, it proxies for the numbers of positions opened each triplet for the promotion to DR at CNRS and to PR in universities, which depend largely from exogenous budgetary decisions taken at the research and higher education ministerial level in France.[9]

The Publishing Spell Selection probit Equation

Our publishing spell selection equation is a probit equation similar to the one for promotion, with a binary indicator of publishing triplets as a dependent variable, noted *Publishing*, and equal to one for a publishing triplet and zero otherwise. It also includes two groups of explanatory variables and time dummies. The first group for age and gender is exactly the same as for the promotion equation. The second, analogous to the one of past productivity, measures the persistence of publication activity in the past three triplets by means of four binary dummies. These four dummies (noted *Persistence 111*, *Persistence 110/101/011*, *Persistence 100/010/001* and *Persistence 000*) respectively indicate that a CNRS or university physicist has published at least one article in each of the three consecutive triplets, or in two, or in one, or in none of them. In the equation, these four dummies are lagged by one triplet, covering the time span from *t-1* to *t-3*. Finally, the time dummies control for general unobserved factors such as increasing emulation and pressure to publish in CNRS and universities.

Although similar in their specification, and, as we shall see, with implications on our estimates of the productivity effects of gender and age, which are going

9. In a related work we document the impact on recruitment and promotion in French universities due to a massive budgetary increase, and investigate its longer term consequences on publication productivity (Lissoni *et al.* [2011]).

in the same direction, the promotion equation and the publishing spell selection equation work very differently. In the case of promotion, we observe rank and promotion, which we know are potentially a major source of endogeneity in the productivity equation. We can thus specify jointly the productivity and promotion equations to identify and estimate them as a system of two simultaneous equations. Assuming that the lagged explanatory variables of promotion are predetermined with respect to current productivity, which seems reasonable, we can separately estimate the promotion equation in a first step. Then, as already indicated, we can estimate consistently the productivity equation in a second step, by substituting the first step predicted probability of promotion to the rank indicator.

In the case of publishing spells, we know that they can reflect a variety of factors for many of which we have no available measures. The best we can do is to specify a selection equation as a function of those factors we measure, and to estimate it jointly with the productivity equation as a two equation Tobit-type model. To estimate consistently this model we simply relied on Heckman's two-step method, where the probit equation is estimated in the first step, and the productivity equation estimated in the second step, including as an additional explanatory variable the first step inverse Mill's ratio, that is the predicted probability of non-publishing noted *Prob(non-publishing)*.

The Productivity Regression

Our productivity equation is a basic linear regression of log-productivity, noted *log(Prod)*, on four groups of explanatory variables and time dummies. The first group is that of gender and age just as in the promotion and selection equations. The second concerns the "initial productivity" of individual scientists in the first triplet when we observe them and which we kept apart in the construction of our panel data samples. We thought better to characterize initial productivity by separating its quantity and quality components: the number of articles published in the first triplet noted *log(First Art)*, the corresponding average five years journal impact factor noted *log(First IF),* and a binary indicator noted *No first publication* (equal to one for a non-publishing triplet and zero otherwise).

Initial productivity variables can capture, to a significant extent, the positive (or negative) effects of early scientific success (or lack of success) throughout the process of cumulative advantage (or disadvantage) at play in scientific productivity (Merton [1968]; Allison *et al.* [1982]; David [1994]). They proxy largely for permanent individual unobservable effects, or "fixed" effects, and are associated with a number of underlying variables such as individual capability, motivation and talent. They have the great advantage of taking out relatively little of the variability of the dependent regression variable, while fixed individual effects usually wipe out very much of that variability, with the harmful consequence of increasing strongly potential biases due to measurement errors in variables (Mairesse [1990]; Griliches and Mairesse [1998]).

The third group of variables consists of the predicted probabilities of promotion and non-publishing spells *Prob(promotion)* and *Prob(non-publishing)*. As already explained, they are included in the productivity equation to correct for the endogeneity of the rank and the selectivity of publishing spells.

The fourth group is the one we refer to as collaboration variables. It is the largest in terms of number of variables, since we can measure them on the basis of the publication data we collected, and they are very important determinants of scientific productivity. The importance of collaboration has been rising rapidly in the last decades in most disciplines, particularly in physics. Collaboration can take many forms, from face to face interactions with laboratory colleagues and discussions with participants in conferences, to internet exchanges among members of invisible colleges (Price and Beaver [1966]; David [1998]; Mairesse and Turner [2006]; Stephan *et al.*, this volume). It reflects an increasing specialization and division of work, as well as the "death of distance" largely permitted by the ICT revolution. Our collaboration variables thus characterize important aspects of the working environment of individual researchers, be it local, institutional, national or international.

Based on the information we can retrieve from the publication data, we defined in total eighteen collaboration variables: four quantitative in logarithms, four stand-alone binary indicators and two sets of respectively six and four binary indicators non-overlapping and adding up to one (which implies that one in each set must be taken as reference for the others in the productivity regression). All these collaboration variables are also lagged by one triplet to largely circumvent inverse causality effects from productivity to collaboration. They appear with their abbreviated labels in Table 8 showing the productivity regression estimates. Simple descriptive statistics for them are also given in Table A2 of Appendix A.

The quantitative variables are the following. The first one *L.log(number of authors harmonic average)* is precisely the logarithm of the harmonic average of the number of authors of the articles published by the individual physicist in the previous triplet (*i.e.* counting the physicist too). The medians of this harmonic average for the CNRS and university samples are respectively of 3.7 and 4.1 authors per article. Using the harmonic average allows us to test whether the coefficient of this variable estimated in the productivity regression is close to one, which would be the implication of measuring the scientist productivity on the basis of what is called publication fractional counts, that is down-weighting articles in proportion to their number of authors. As we will see, this test is strongly rejected, confirming that the frequent practice of fractional counts does not make much sense to assess and analyze publication productivity at the individual scientist level.

The next two other quantitative variables *log*(*Art. national coauthors*) and *log*(*IF national coauthors*) measure the average log-productivity, in terms of both quantity and quality per triplet, of CNRS and university coauthors of our individual CNRS or university physicists.[10] The last quantitative variable *log*(*National coauthors of coauthors*) measures the average numbers per triplet of the CNRS and university coauthors of the CNRS and university coauthors themselves of our individual physicists.[11] There are two cases when these three coauthors variables

10. We can compute these variables only for the CNRS and university physicists, that is nearly all physicists working in France, because being in our samples we know their publications. To compute them for the very many international coauthors will imply that we can collect their publications, which would be a huge and difficult task. We can, however, more easily characterize the extent of international collaboration of the CNRS and French university physicists on the basis of the foreign affiliations mentioned in their publication.

11. To avoid some form of double counting of the articles of individual physicists and those of their coauthors, we have been careful to exclude from *Art. national coauthors* the articles coauthored with the physicists themselves. We also applied such exclusion to *IF national coauthors* and *National coauthors of coauthors*.

cannot be computed: if a physicist has no CNRS or university co-authors or if the coauthors themselves have no publications with CNRS or university physicists other than this physicist. To take care of these two cases, we have thus to include in the productivity regression two stand-alone binary indicators, noted respectively *No national coauthors* and *Coauthor national no publication*. Because the coauthors variables are lagged in the productivity equation, they cannot also be computed in the case when a physicist, although publishing in the current triplet, did not in the previous triplet. To take care of this third case, we have to include in the productivity regression the lagged stand-alone binary indicator *L.No publication*, already defined to be included in the promotion equation.

We also consider a fourth stand-alone binary indicator noted *Reference author*, equal to one if a physicist in a triplet is listed as the reference author in at least one article published in this triplet. This generally indicates that he or she has had a relatively important role among the coauthors in producing the article, and it can be a signal of a very dynamic and productive researcher. We observe that the average proportion of physicists listed as reference author in at least one article per publishing triplet is quite high: respectively 69 % and 53% for CNRS and universities.

To characterize the importance of international collaboration, we constructed a set of six non overlapping binary indicators according to a classification of CNRS and university physicists coauthors and triplets in six distinct categories, based on the more or less precise affiliation information recorded in the published articles. The first indicator noted *Single author* is that of a physicist publishing in a triplet without any co-author; this is the case of only 1% of the observations for both CNRS and the universities. The second one noted *Coauthors of unknown type* is that of a physicist publishing in a triplet with coauthors for whom we have no affiliation information or one too vague to be useful; this the case of 8% and 15% of the observations for CNRS and the university respectively. The third and fourth indicators noted *Only national coauthors* and *National/ Nuclear coauthors* characterize the triplets in which a physicist has only national coauthors, with or without colleagues classified as nuclear physicists. The fifth and sixth indicators labeled *National/International coauthors* and *National/Nuclear/International coauthors* characterize the triplets in which a physicist has also at least one international coauthor. As we have explained, we have excluded from our samples nuclear physicists because their field stands apart in many respects from the other fields of physics, in particular because they tend to publish very much and with very many coauthors to a point that the notion of individual author itself disappears. Nonetheless, we observe a fair amount of collaboration between the physicists in our samples and those classified in nuclear physics, and we thought appropriate to set apart this specific type of cross-field collaboration. The average proportion of observations for physicists with only national coauthors compared to that with international coauthors is of 24% versus 66% (of which 3% and 12% involving nuclear physicists) for CNRS physicists and of 38% versus 46% (of which 2% and 5% in volving nuclear physicists) for university physicists.

Furthermore, we thought important to characterize the persistence over time of the collaboration between CNRS and university physicists. Although the former are mainly committed to research activities and the latter are devoting half time to research and half time to teaching activities, they collaborate frequently with each other. They do in particular in UMR *(Unité Mixte de Recherche)*, which are special laboratories or research centers, where academics and CNRS scientists can work on common research projects. The denomination and frontiers of UMR changing frequently and the affiliation recorded in publications being rarely accurate enough, we measure the existence of collaboration between physicists of the two organizations on the basis of the lists of their names. Precisely we take as evidence of a more or less extended collaboration the coauthorship of at least one article in one, two or three triplets over three consecutive triplets. We thus constructed a set of four alternative binary indicators noted: *Coll. UN/CNRS 000, Coll. UN/CNRS 100/010/001, Coll. UN/CNRS 011/101/110* and *Coll. UN/CNRS 111*, respectively signaling collaboration between CNRS and university physicists in none of the three triplets *t, t–1* and *t–2*, in one of the three, two of the three, and all three. We find that CNRS and university physicists coauthor at least one publication in all three consecutive triplets for respectively 29% and 22% of the CNRS and university sample observations, while they do in at least one triplet out of three for respectively 61% and 48% of them.

Finally, likewise as in the promotion and publishing spell equations, time dummies are included in the productivity equation to control for general unobserved factors such as increasing trends in scientific collaboration and publications.

Econometric Findings

Before highlighting the findings related to the productivity gender gap on which we focus in this paper, it is appropriate to have a first look at our overall results and comment rapidly on them.

A First Look

The estimates for CNRS and the universities are given in Tables 6 and 7 for the promotion and publishing spell selection probit equations and in Table 8 for the productivity regression. In these three tables we present in the first two columns the CNRS and university estimates for the specification without interactions between gender and age and in the two last columns the ones for the specification with such interactions.

The coefficient estimates of the promotion probit equation confirm our expectations. Productivity in the previous triplet, in terms of number of published articles and average journal impact factor, has a positive and very significant impact on the probability of being promoted from research associate (CR) to research director (DR) in the CNRS or from lecturer (MC) to professor (PR) in the universities. Having no publications in the previous triplet has a negative impact, large and statistically significant for universities, but not for CNRS. We observe also, as awaited, that the probability of promotion, conditional on past productivity, varies with age according to an inverted u-shaped curve, indicating that this probability is lower for the younger physicists as well as for the very senior physicists that are not already promoted.[12] Finally, we find that, conditional on past productivity and age, female physicists have significantly lower probabilities of promotion than their male colleagues. If we consider the case of a representative physicist, with an average productivity (six articles per triplet and an average journal impact factor of 5) and aged forty in the triplet 2003-05, the chances of being promoted are 6.3% lower for a woman than for a man in CNRS, while they are 16.3% lower in universities.

Table 6. *Promotion probit equation for CNRS and university physicists, with and without age*gender interactions*

Rank indicator	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (= 1)	– 0.29***	– 0.46***	– 0.26***	– 0.46***
(Age-40)/10	1.78***	1.44***	1.78***	1.44***
((Age-40)/10)^2	– 0.47***	– 0.50***	– 0.46***	– 0.49***
(Age-40)/10 * Woman			– 0.00	0.03
((Age-40)/10)^2 * Woman			– 0.03	– 0.02
Lagged productivity				
L.log(Art)-log(6)	0.25***	0.46***	0.25***	0.46***
L.log(avg IF)-log(5)	0.23***	0.12***	0.23***	0.12***
L.No Publications	– 0.090	– 1.00***	– 0.09	–1.00***
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	– 0.78***	– 0.23***	– 0.79***	– 0.23***
Observations	5,517	9,256	5,517	9,256
Pseudo R2	0.38	0.28	0.38	0.28

The coefficient estimates of the publishing spell selection probit equation also confirm our expectations. The probabilities of publishing are significantly higher for the CNRS and university physicists who are more persistent in publishing in the three previous triplets than for those who are less persistent or not publishing. A representative woman physicist, aged forty in the triplet 2003-05 and publishing in all three preceding triplets, has a publishing probability respectively lower by 8.2% and 17.7% in CNRS and universities than a representative man colleague of the same age and publishing profile. We can also remark, as we shall see more clearly for productivity, that the estimated coefficient of the interaction term (age*woman) is positive, while the estimated coefficient of age is negative, which implies a reduction with age of the publishing gap of women relative to men.[13]

12. The estimated maximum probability of promotion is on the high side, respectively about 58 and 54 years for CNRS and university physicists.

13. The estimated coefficients for age and the interaction term (age*woman) are respectively negative and positive and about the same size (0.10) for both CNRS and universities; however it is only statistically different from zero at 10% level of confidence for universities and not for CNRS.

Table 7. *Publishing triplet selection probit equation for* CNRS *and university, with and without age*gender interactions*

Publishing Indicator	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (= 1)	– 0.21***	– 0.42***	– 0.21**	– 0.46***
(Age-40)/10	– 0.10*	– 0.08***	– 0.11**	– 0.10***
((Age-40)/10)^2	0.01	– 0.03	0.01	– 0.02
(Age-40)/10 * Woman			0.09	0.10*
((Age-40)/10)^2 * Woman			– 0.03	0.01
Productivity persistence				
L.Persistence 111	2.72***	2.24***	2.72***	2.25***
L.Persistence 110/101/011	1.73***	1.33***	1.73***	1.34***
L.Persistence 100/010/001	0.89***	0.79***	0.89***	0.81***
Reference. L.Persistence 000 (ref.)	–	–	–	–
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	1.92***	1.33***	1.92***	1.34***
Observations	5,517	9,256	5,517	9,256
Pseudo R2	0.40	0.33	0.40	0.33

The productivity equation, as detailed in the preceding section, includes four different groups of explanatory variables. All four groups have substantial and statistically significant impacts on scientific productivity. The three initial productivity variables have long lasting effects. In particular having no publications in the first triplet is a predictor of a much smaller scientific productivity in the future for CNRS physicists and even more so for university physicists. The control for the endogeneity of rank by including the predicted promotion probability has also a very large and significant impact on productivity, while the control for publishing selectivity by including the non-publishing triplet predicted probability has a more modest but still significant impact.

Among the set of collaboration variables, it is noteworthy that the average number of coauthors has no impact whatsoever on productivity, a result which implies at drastic rejection of the test that fractional counts of articles (*i.e.,* inversely weighting them by the number of their authors) would be a better measure of publication productivity at the individual scientist level. The evidence is that what actually matters for productivity is the nature and quality of collaboration for CNRS physicists as well as university physicists, with few differences only.

Physicists in both organizations that are coauthoring articles with national colleagues, who are productive and publish in particular in high impact factor journals, are themselves sizably more productive. CNRS and university physicists who happen to be reference authors appear markedly more productive. The two indicators that physicists have at least one international coauthor are for both organizations a strong signal that they are more productive. Collaborations between physicists in the two organizations, and in particular persistent

Table 8. *Productivity equation for* CNRS *and university physicists, with and without age*gender interactions**

Productivity: log(Prod)	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (= 1)	– 0.00	0.11***	0.05	0.25***
(Age-40)/10	– 0.69***	– 0.62***	– 0.70***	– 1.09***
((Age-40)/10)^2			– 0.05***	0.25***
(Age-40)/10 * Woman			0.24***	0.43***
((Age-40)/10)^2 * Woman			– 0.11**	– 0.14***
Initial productivity				
No first publications	– 0.51***	– 1.03***	– 0.55***	– 0.96***
log(first Art) –log(6)	0.21***	0.20***	0.21***	0.18***
log(avg.first IF) – log(5)	0.18***	0.27***	0.18***	0.25***
Promotion and non-publishing spells				
Prob(promotion)	1.98***	2.24***	2.07***	3.07***
Prob(*non-publishing triplets*)	– 0.34**	– 0.37***	– 0.33**	– 0.33***
Collaboration				
L.log(numberof authors harmonic average)	0.04	0.02	0.05	0.01
L.log(Articles national coauthors)-log(2)	0.06***	0.02	0.06***	0.01
L.log(IF national coauthors)-log(3)	0.18***	0.17***	0.18***	0.16***
L.log(National coauthors of coauthors)-log(3)	0.06*	0.03	0.06*	0.04
L.No national coauthor	0.21**	0.39***	0.22**	0.38***
L.National coauthor no publication	0.27***	0.18**	0.27***	0.20**
L.No publication	0.04	0.41***	0.02	0.52***
L.Reference author	0.29***	0.22***	0.29***	0.17***
L.Single author	– 0.36***	0.03	– 0.37***	0.097
L.Multi-authored but unknown type	0.04	– 0.08	0.03	– 0.06
L.National (ref.)	–	–	–	–
L.National / Nuclear	0.22**	0.09	0.21**	0.07
L.National / International	0.33***	0.17***	0.32***	0.14***
L.National / Nuclear/International	0.57***	0.32***	0.56***	0.26***
Ref. L.Coll. FR/CNRS 000	–	–	–	–
L.Coll. UN/CNRS 100/010/001	0.07*	0.04	0.05	0.08**
L.Coll. UN/CNRS 011/101/110	0.07	0.06	0.05	0.11**
L.Coll. UN/CNRS 111	0.05	0.13***	0.05	0.11***
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	3.22***	3.14***	3.28***	2.66***
Observations	4,920	6,219	4,920	6,219
Rho	– 0.36	– 0.34	– 0.33**	– 0.33***
Pseudo R2	0.307	0.349	0.312	0.370

* Estimated coefficients, based on OLS corrected for promotion endogeneity and non-publishing triplet selectivity.

collaborations over three triplets that suggest that they probably work in the same team in a common laboratory, boost modestly but significantly the productivity of the university physicists, not that of their CNRS colleagues.[14]

Beyond the diversity of our measures and in spite of their weaknesses, our evidence on collaboration points mainly to the endogenous process of mutual cooptation among the most productive scientists, revealing by the same token the importance of the quality of the working environment and research network.

Overall each the three groups of factors: collaboration, probabilities of promotion and publishing spell, and initial productivity, account significantly, even if more or less substantially, for differences in scientific productivity. Taken together, we find also that they basically invalidate the gender productivity puzzle in case of CNRS physicists and even reverse it for university physicists. These striking results are worthwhile documenting and trying to understand in details, which we do now.

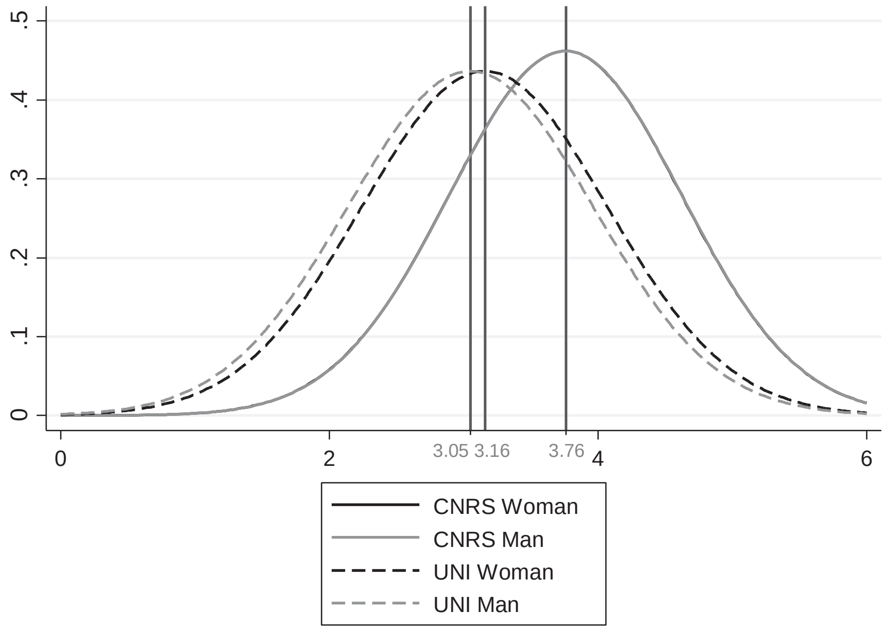
A Detailed Look at the Pieces of the Gender Productivity Puzzle

The results of Table 8 on the estimated gender effects on productivity with and without interactions with age can be simply illustrated by Graphs 2 and 3. Graph 2 can be compared and contrasted to Graph 1. While Graph 1 charts the observed log-productivity distributions, Graph 2 shows the corresponding distributions for the predicted log-productivity separately for "representative" female and male CNRS and university physicists. The predicted log-productivity is, based on the estimated productivity regression without age*gender interactions; and the representative male and female physicists which are forty years old in the last triplet 2003-2005 and are all supposed to have the average characteristics of male physicists in the same organization. We see that for CNRS the predicted distributions for men and women coincide fully with the same estimated average log-production of 3.76 (in absolute levels 4.3 equivalent articles per triplet in journals with impact factor 10). For universities thez differ slightly but to women advantage with an estimated average log- productivity of 3.16 versus 3.05 for men (2.6 versus 2.1 equivalent articles) and a statistically significant estimated gender gap of 0.11 (0.5 equivalent article).

Graph 3 illustrates the fact the gender and age interact significantly in accounting for productivity, by the change in the predicted average log-productivity with age, based on the estimated productivity regression and for representative physicists with the same average characteristics as for Graph 2. The main message of Graph 3 in fact concerns more what can be called the age productivity gap than the gender gap itself. Is shows that for CNRS the estimated log-productivity decreases strongly with age and similarly for men and women: from about 4.0 at age forty (5.5 equivalent articles per triplet in journals with impact factor 10) to about 2.2 at age sixty (only 0.9 equivalent articles). For universities

14. Two surprising results can also be noted: having no publication in a triplet (*L.No publication*) is a significant predictor of being productive in the next triplet for university physicists but not for their CNRS colleagues, while publishing in a triplet but without coauthors (*L.Single author*) is a significant predictor of lower productivity in the next triplet in the case of CNRS physicists but not university physicists. The first one may simply reflect the fact that non-publishing triplets, since they are more frequent for university physicists, tend to alternate frequently with publishing triplets. The second case is actually extremely rare and difficult to interpret *a priori*.

Graph 2. *Distribution of predicted log-productivity for representative* CNRS *and university representative women and men physicists in the last triplet 2003–2005**

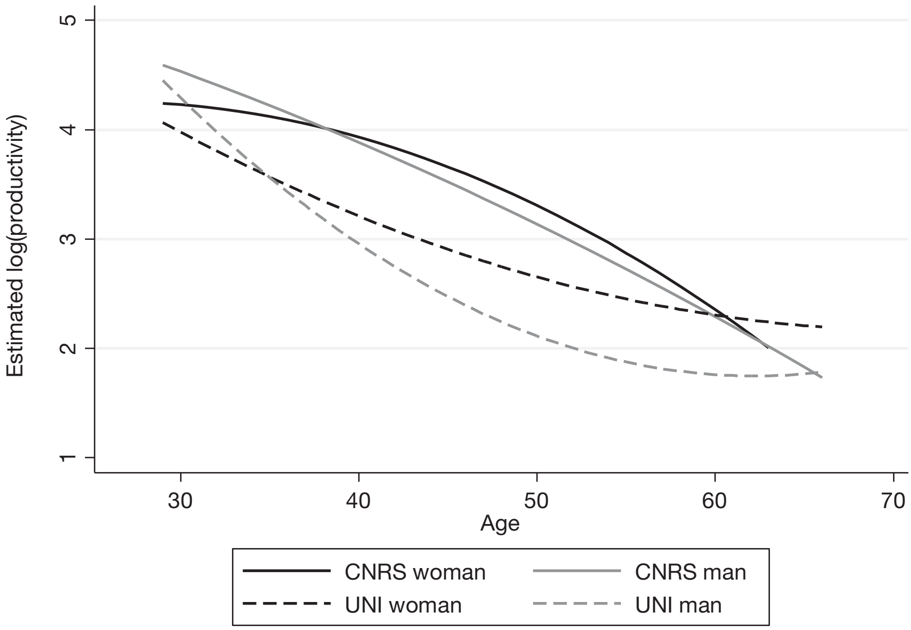


* The predicted log-productivity distributions of Graph 2 are based for simplicity on the estimates of the regressions without age*gender interactions of Table 8 and they suppose an approximately normal log-distribution as in Graph 1. Representative physicists are 40 years old in 2003-2005, and are supposed to have all the same average characteristics than men physicists in the same organization as concerns all other control variables: collaboration, probability of being promoted and experiencing non-publishing triplets, and initial productivity.

the estimated log-productivity decreases with age less rapidly for women than for men: from about 3.5 (3.3 equivalent articles) at age of thirty five, when predicted log-productivity is equal for men and women, to respectively about 1.8 and 2.2 for men and women (or 0.6 and 0.9 equivalent articles) at age sixty when it is higher for women.

In Table 9, complemented by Tables A3 and A4 of Appendix A, we try to answer the question of what are the respective contributions the four groups of factors in accounting for the gender productivity gap. This answer cannot be straightforward, since these contributions are interrelated and thus more or less dependent of the order in which we want considerer them. It is also complicated by the fact that the size of the estimated age*gender interactions and age and age squared impacts are not constant but also changing significantly with these set of factors.

In Table 9, we include the four groups of factors in the productivity regressions in the following order: gender, age, age*gender interactions and time dummies as starting or reference estimates, then sequentially collaboration variables, rank alone with no correction for endogeneity, rank with correction for endogeneity, publishing spell selectivity, and last initial productivity variables. While at the level of an individual scientist these different factors reflect external constraints that are very much binding and extremely difficult, or even impossible, to change, Graph 3. *Change with age of predicted log-productivity for CNRS and university representative women and men physicists in the last triplet 2003–2005**



* The predicted log-productivity distributions of Graph 3 are based on the estimates of the regressions with age*gender interactions of Table 8. Representative physicists are 40 years old in 2003–2005, and are supposed to have all the same average characteristics than men physicists in the same organization as concerns all other control variables: collaboration, probability of being promoted and experiencing non-publishing triplets, and initial productivity.

this is less true at a more collective level. The order of entering these factors in the regressions of Table 9 is chosen to supposedly correspond to an increasing degree of exogeneity and rigidity of such constraints they reflect at such at the collective level. It can be justified in the policy perspective of promoting gender parity and fostering productivity in scientific research in universities or public research organizations like CNRS. The quality of collaboration and gender equality in promotion both depend closely on the resources and openess of the working environment and can be stimulated by appropriate incentives. The issue of non-publishing spells is more difficult to address insofar as their frequent occurrence can reflect a variety of external activities and duties, such as teaching, family engagements, or administrative and management responsibilities. This is also the case for scientists' initial accomplishments, which can be considered as largely given, early on their careers, be they due to their capacities, favorable circumstances or both.

Accepting these arguments for adopting the sequential combination of the main groups of factors in the productivity regressions of Table 9, we find that collaboration variables and rank with endogeneity correction contribute nearly alone to the full disappearance of the average gender productivity gap for CNRS, and well as to its disappearance and inversion for universities. The contribution for rank endogeneity is particularly striking for universities. In contrast, the additional contributions of publishing spells selectivity and initial productivity appear quite minor. We also observe similar changes for both CNRS and

Table 9. *Estimates of gender and age coefficients in productivity regression including group of control variables sequentially**

CNRS: Log(Prod)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age and Gender							
Woman(=1)	– 0.29***	– 0.29***	– 0.17***	– 0.16***	0.00	0.01	0.05
(Age-40)/10		– 0.06**	– 0.03	– 0.08***	– 0.73***	– 0.72***	– 0.70***
((Age-40)/10)^2		– 0.06***	– 0.05**	– 0.05**	– 0.03*	– 0.03*	– 0.05***
(Age-40)/10 X Woman		0.09	0.07	0.08	0.22***	0.22***	0.24***
((Age-40)/10)^2 X Woman		– 0.05	– 0.04	– 0.04	– 0.12**	– 0.12**	– 0.11**
Constant	3.76***	3.87***	2.92***	2.88***	2.38***	2.47***	3.28***
Time dummies	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Collaboration	*no*	*no*	*yes*	*yes*	*yes*	*yes*	*yes*
Rank	*no*	*no*	*no*	*yes*	*yes*	*yes*	*yes*
Rank endogeneity	*no*	*no*	*no*	*no*	*yes*	*yes*	*yes*
Selectivity	*no*	*no*	*no*	*no*	*no*	*yes*	*yes*
Initial productivity	*no*	*no*	*no*	*no*	*no*	*no*	*yes*
Observations	4,920	4,920	4,920	4,920	4,920	4,920	4,920
R2	0.009	0.020	0.071	0.081	0.162	0.229	0.312

UNIVERSITY: Log(Prod)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age and Gender							
Woman(=1)	– 0.31***	– 0.31***	– 0.20***	– 0.13**	0.23***	0.28***	0.25***
(Age-40)/10		– 0.13***	– 0.04*	– 0.19***	– 1.21***	– 1.17***	– 1.09***
((Age-40)/10)^2		0.01	-0.00	0.03	0.28***	0.27***	0.25***
(Age-40)/10 X Woman		0.05	0.05	0.12**	0.47***	0.47***	0.43***
((Age-40)/10)^2 X Woman		– 0.06	– 0.05	– 0.06	– 0.17***	– 0.17***	– 0.14***
Constant	3.05***	3.17***	2.50***	2.39***	1.76***	1.92***	2.66***
Time dummies	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Collaboration	*no*	*no*	*yes*	*yes*	*yes*	*yes*	*yes*
Rank	*no*	*no*	*no*	*yes*	*yes*	*yes*	*yes*
Rank endogeneity	*no*	*no*	*no*	*no*	*yes*	*yes*	*yes*
Selectivity	*no*	*no*	*no*	*no*	*no*	*yes*	*yes*
Initial productivity	*no*	*no*	*no*	*no*	*no*	*no*	*yes*
Observations	6,219	6,219	6,219	6,219	6,219	6,219	6,219
R-squared	0.008	0.020	0.120	0.154	0.313	0.353	0.370

* Estimates of gender and age coefficients in productivity regression controlling sequentially for gender (column 1), then age, age gender interactions and time dummies (column 2), then collaboration (column 3), then rank (column 4), then rank endogeneity (column 5), then non-publishing triplet selectivity (column 6), and last initial productivity (column 7).

universities in the factor contributions to the estimated age*gender interactions and age and age squared impacts.

In Tables A3 and A4, we propose a more pragmatic approach to appraise the respective contributions of the four groups of factors in accounting for the gender productivity gap. In Table A3, the main factors are included alone, one at a time, in the productivity regressions in addition to gender, age, their interactions and time dummies. In Table A4, the main factors are excluded in turn, one at a time, from our central productivity regressions of Table 8 (and last columns in Table 9) that include all groups of factors. The evidence provided by these tables is on the whole fairly clear-cut, with the exception of publishing spells. It underscores the predominant effect of rank endogeneity in accounting for the productivity gap. It shows also that the striking decrease of the overall impact of age and age squared on CNRS and university productivity is mainly attributable to rank endogeneity. It confirms fully the important effect of collaboration and largely the minor effect of initial conditions, with the exception of a significant reduction in the gender gap for CNRS when they are included first. However, when we include publishing spell selectivity alone, we find that its effect is actually quite important for both organizations, but when we exclude it alone from the list of main factors, we find that it remains important in the case of universities, but not CNRS. Further analysis shows that publishing spell selectivity is intricately related with collaboration for CNRS and with rank endogeneity for universities. Its effect appears quite sizable when it comes into play before collaboration for CNRS and before rank endogeneity for universities.

We also performed a number of robustness checks and sensitivity analyses to strengthen our findings with regard to the gender productivity gap. Two issues mainly arise. One is whether our results are preserved if we use a more flexible specification of the productivity equation allowing the impacts of main factors to differ for male and female physicists. We have allowed for such flexibility of the productivity equation by including interactions between gender and the main factors, and found that the gender with age and age squared interactions are basically the only ones relevant and statistically significant.

The second issue is whether our results remain valid if we focus the analysis on high quality and visibility publications. As mentioned when explaining the construction of our samples, we have replicated the present analysis twice, considering only articles published in journals with respectively high (more than 5) and very high (more than 10) five years impact factors. We obtain very similar results for the gender productivity gap and orders of magnitude of the estimated impacts of the main variables, wich are documented in details in Tables B1 to B10 of Appendix B.

CONCLUSION

We have developed in this paper an econometric framework considering together a scientific productivity function and two other equations, one for the promotion of scientists to high rank and one for the occurrence of non-publishing spells. This framework is needed to account for the interrelated differences between female and male scientists on publication productivity, rank status and non-publishing spells, that we observe on two panel data samples constructed for CNRS and French university physicists. The econometric results we have found are interesting in a number of respects, but the most striking findings concerns the so-called gender productivity puzzle on which we have chosen to focus our analysis here. While the observed average publication productivity gap of women physicists relative to their men colleagues is about one third in both CNRS and universities, women appear as productive as men in CNRS and even significantly more in universities, if we control for other determinants of productivity, in particular unequal chances of promotion and frequent non-publishing spells.

In view of such findings, we felt compelled to explore the empirical economic and sociological literature on the gender productivity puzzle, which we briefly review in the first part of our paper. This literature comprises careful and excellent studies documenting the productivity gap and investigating a number of its possible determinants and sources. It offers interesting results, but often heterogeneous and unrelated and sometimes contradictory, from which no clear-cut evidence emerges on the main sources of the gender productivity gap. Although analytical and methodological issues are rarely discussed, we found considerations confirming our main conclusions: the necessity of having longitudinal data bases and the need of a general framework of analysis, the importance of taking into account career advancements, the occurrence of low publishing spells, collaboration and work environment, initial conditions, and cumulative advantage.

Many directions of work can be followed to generalize and improve our study. We have tried to keep our analysis relatively simple, but it could also benefit from relying on more elaborate model specifications and estimation methods. Theoretical understanding as the one developed in several contributions of this volume is needed and could be extremely beneficial. Replicating our study for different disciplines and in other countries would be of course also important. But far more critical is undoubtedly the active pursuit of the data efforts that are necessary to develop richer analyses and permit specific investigations.

We find for example that scientific productivity declines with age more rapidly for male than for female scientists and thus the estimated gender gap also decreases with age. These results suggest rather strongly that the causes may be found in activities external to research and likely to vary with age, such as parenthood and family engagement, or administrative and management responsibilities. But to know better and to assess to what extent this is the case we need to have information on if and when scientists had children, and if and when they have been involved in more or less demanding responsibilities. Similarly if we are interested in assessing precisely why the publication productivity of scientists in public research organization like CNRS and universities are quite different, we must be informed on actual teaching load, on quality of recruitment, career incentives, research resources and work environment. As a general rule in empirical investigation like ours, the existence, availability and elaboration of the data are indeed at the heart of the matter.

REFERENCES

ALLISON, P.D., LONG, J.S., KRAUZE, T.K. [1982]. "Cumulative advantage and inequality in science," *American Sociological Review*, 47 (5): 615–625.

ALLISON, P.D., STEWART, J.A. [1974]. "Productivity differences among scientists: Evidence for accumulative advantage," *American Sociological Review*, 39 (4): 596–606.

BEAVER, D. DE B., ROSEN, R. [1978]. "Studies in scientific collaboration," *Scientometrics*, 1 (1): 65–84.

COLE, J.R., ZUCKERMAN, H. [1984]. "The productivity puzzle: Persistence and change in patterns of publication of men and women scientists." In MAEHR, P. and STEINKAMP, M. W. (eds). *Advances in motivation and achievement*. Greenwich (CT): JAI Press, 217–256.

CORLEY, E., BOZEMAN, B., GAUGHAN, M., [2003]. "Evaluating the impacts of grants on women scientists' careers: the curriculum vitae as a tool for research assessment." In SHAPIRA, P., KUHLMANN, S. (eds). *Learning from science and technology policy evaluation: Experiences from the United States and Europe*. Cheltenham (UK), Northhampton (Mass.): Edward Elgar Publishing, 293.

CRANE, D. [1972]. *Invisible colleges: Diffusion of knowledge in scientific communities*, Chicago: University of Chicago Press,

DAVID, P.A. [1994]. "Positive feedbacks and research productivity in science: reopening another black box." In GRANDSTRAND, O. (ed.). *Economics and Technology*. Elsevier: Amsterdam. 65–89.

DAVID, P.A. [1998]. "Communication Norms and the Collective Cognitive Performance of "Invisible Colleges." In NAVARETTI, G. B. DASGUPTA, P., MÄLER, K.G and SINISCALCO, D. (eds), *Creation and Transfer of Knowledge*. Springer Berlin Heidelberg, 115–163.

DASGUPTA, P., DAVID, P.A. [1994]. "Toward a new economics of science," *Research Policy*, 23 (5): 487–521.

FINKEL, S.K., OLSWANG, S.G. [1996]. "Child Rearing as a Career Impediment to Women Assistant Professors," *Review of Higher Education*, 19 (2): 123–139.

FOX, M.F. [1981]. "Sex, salary, and achievement: Reward-dualism in academia," *Sociology of Education*, 54 (2): 71–84.

FOX, M.F. [1983]. "Publication productivity among scientists: A critical review," *Social Studies of Science*, 13 (2): 285–305.

FOX, M.F. [2005]. "Gender, Family Characteristics, and Publication Productivity among Scientists," *Social Studies of Science*, 35 (1): 131–150.

FOX, M.F., FAVER, C.A. [1985]. "Men, women, and publication productivity: Patterns among social work academics," *The Sociological Quarterly*, 26 (4): 537–549.

FOX, M.F., MOHAPATRA, S. [2007]. "Social-organizational characteristics of work and publication productivity among academic scientists in doctoral-granting departments," *The Journal of Higher Education*, 78 (5): 542–571.

FOX, M.F., STEPHAN, P.E. [2001]. "Careers of young scientists: Preferences, prospects and realities by gender and field," *Social Studies of Science*, 31 (1): 109–122.

GRILICHES, Z., MAIRESSE, J. [1998]. "Production Functions: The Search for Identification." In STRÖM, S. (ed.). *Econometrics and Economic Theory in the 20th Century: The Ragnar Frish Centennial Symposium*. Cambridge University Press, 169-203.

KATZ, J.S., MARTIN, B.R. [1997]. "What is research collaboration?," *Research policy*, 26 (1): 1–18.

KYVIK, S. [1990]. "Motherhood and scientific productivity," *Social Studies of Science*, 20 (1): 149–160.

KYVIK, S., TEIGEN, M. [1996]. "Child care, research collaboration, and gender differences in scientific productivity," *Science, Technology & Human Values*, 21 (1): 54–71.

LEVIN, S.G., STEPHAN, P.E. [1991]. "Research productivity over the life cycle: Evidence for academic scientists," *American Economic Review*, 81 (1): 114–132.

LEVIN, S.G., STEPHAN, P.E. [1998]. "Gender Differences in the Rewards to Publishing in Academe: Science in the 1970s," *Sex Roles*, 38 (11-12): 1049–1064.

LISSONI, F., MAIRESSE, J., MONTOBBIO, F., PEZZONI, M. [2011]. "Scientific productivity and academic promotion: a study on French and Italian physicists," *Industrial and Corporate Change*, 20 (1): 253–294.

LONG, J.S. [1990]. "The origins of sex differences in science," *Social Forces*, 68 (4): 1297–1316.

LONG, J.S. [1992]. "Measures of sex differences in scientific productivity," *Social Forces*, 71 (1): 159–178.

LONG, J.S., ALLISON, P.D., MCGINNIS, R. [1993]. "Rank advancement in academic careers: Sex differences and the effects of productivity," *American Sociological Review*, 58 (5): 703–722.

LONG, J.S., FOX, M.F. [1995]. "Scientific careers: Universalism and particularism," *Annual Review of Sociology*, 21: 45–71.

MAIRESSE, J. [1990]. "Time-series and Cross-sectional Estimates on Panel Data : Why Are They Different and Why Should They Be Equal ?." In HARTOG, J. RIDDER, G. and THEEUWES, J. (eds), *Panel Data and Labor Market Studies*. Amsterdam, North-Holland, 81-95.

MAIRESSE, J., TURNER, L. [2006]. "Measurement and Explanation of the Intensity of Co-publication in Scientific Research: An Analysis at the Laboratory Level." In ANTONELLI C., FORAY, D., HALL, B. and STEINMUELLER, E. (eds), *New Frontiers in the Economics of Innovation and New Technology: Essays in honour of Paul David*. Edward Elgar Publishing, 255-295.

MARWELL, G., ROSENFELD, R., SPILERMAN, S. [1979]. "Geographic constraints on women's careers in academia," *Science*, 205 (4412): 1225–1231.

MERTON, R.K. [1968]. "The Matthew effect in science," *Science*, 159 (3810): 56–63.

MERTON, R.K. [1973]. *The sociology of science: Theoretical and empirical investigations*, Chicago: University of Chicago Press.

PEZZONI, M., STERZI, V., LISSONI, F. [2012]. "Career progress in centralized academic systems: Social capital and institutions in France and Italy," *Research Policy*, 41 (4): 704–719.

PRICE, D. DE S., [1963]. *Big science, little science*, New York: Columbia University Press, 119–119.

PRICE, D.J. DE S., BEAVER, D. [1966]. "Collaboration in an invisible college," *American psychologist*, 21 (11): 1011.

SABATIER, M., CARRERE, M., MANGEMATIN, V. [2006]. "Profiles of academic activities and careers: Does gender matter? An analysis based on French life scientist CVs," *The Journal of Technology Transfer*, 31 (3): 311–324.

SAX, L.J., HAGEDORN, L.S., ARREDONDO, M., DICRISI III, F.A. [2002]. "Faculty research productivity: Exploring the role of gender and family-related factors," *Research in Higher Education*, 43 (4): 423–446.

SHAUMAN, K.A., XIE, Y. [1996]. "Geographic mobility of scientists: sex differences and family constraints," *Demography*, 33 (4): 455–468.

SONNERT, G., HOLTON, G.J. [1995]. *Who succeeds in science? The gender dimension*, New Brunswick (N.J.): Rutgers University Press.

STACK, S. [2004]. "Gender, children and research productivity," *Research in higher education*, 45 (8): 891–920.

STEPHAN, P.E., LEVIN, S.G. [1992]. *Striking the mother lode in science. The importance of age, place, and time*, New York: Oxford University Press.

TOREN, N. [1991]. "The nexus between family and work roles of academic women in Israel: Reality and representation," *Sex roles*, 24: 651–667.

TURNER, L., MAIRESSE, J. [2005]. Individual Productivity Differences in Public Research: How important are non-individual determinants? An Econometric Study of French Physicists' publications and citations (1986-1997). Centre National de la Recherche Scientifique, *Mimeo*.

WARD, K., WOLF-WENDEL, L. [2004]. "Academic motherhood: Managing complex roles in research universities," *Review of Higher Education*, 27 (2): 233–257.
WOLFINGER, N.H., MASON, M.A., GOULDEN, M. [2008]. "Problems in the pipeline: Gender, marriage, and fertility in the ivory tower," *Journal of Higher Education*, 79 (4): 388–405.
ZUCKERMAN, H., COLE, J.R., BRUER, J.T. [1991]. *The outer circle*, New York: Norton.

APPENDIX A

COMPLEMENTARY DESCRIPTIVE STATISTICS AND ESTIMATES FOR BASE LINE ECONOMETRIC ANALYSIS

Table A1. *Detailed configuration of CNRS and university unbalanced panel data study samples**

	CNRS			UNIVERSITY		
Number of:	Phys.	Obs. for all triplets	Obs. for publishing triplets only	Phys.	Obs. for all triplets	Obs. for publishing triplets only
INITIAL SAMPLE						
(1) All physicists active in 2004-2005	1,112	5,631	5,008	1,970	9,413	6,328
Of which						
Present for one triplet only*	10%	2%	2%	8%	2%	2%
STUDY SAMPLE						
(2) All physicists present for 2 triplets at least	998	5,517	4,920	1,813	9,256	6,219
Of which from:	*100%*	*100%*	*100%*	*100%*	*100%*	*100%*
1982-84 to 2003-05 (8)	30%	44%	46%	25%	39%	43%
1985-87 to 2003-05 (7)	13%	17%	16%	7%	9%	9%
1988-90 to 2003-05 (6)	9%	10%	10%	11%	13%	10%
1991-93 to 2003-05 (5)	11%	10%	9%	11%	11%	9%
1994-96 to 2003-05 (4)	11%	8%	8%	16%	13%	13%
1997-99 to 2003-05 (3)	11%	6%	6%	15%	9%	9%
2000-02 to 2003-05 (2)	14%	5%	5%	15%	6%	6%

* Total numbers of physicists and observations for all triplets and for publishing triplets only, and percentages for underlying balanced subsamples (from two to eight consecutive triplets).

Table A2. *Simple descriptive statistics (mean, median and standard deviation) for main variables included in* CNRS *and university productivity regressions*

	CNRS (obs. 4,920)			UNIVERSITY (obs. 6,219)		
	Mean	SD	Median	Mean	SD	Median
Dependent variable						
Log(Prod)	3.71	1.03	3.86	3.00	1.13	3.08
Log(Art)	1.77	0.87	1.79	1.32	0.89	1.38
Log(IF)	1.94	0.44	1.97	1.68	0.51	1.73
Age and Gender						
Age	42.20	8.64	41.00	42.40	8.86	41.00
woman (= 1)*	0.13	0.34	0.00	0.13	0.34	0.00
Initial productivity						
No first publications	0.02	0.15	0	0.49	0.21	0
log(first Art) –log(6)	– 2.16	0.83	– 2.20	– 2.41	0.93	– 2.48
log(avg. first IF) – log(5)	– 1.23	0.49	– 1.20	– 1.43	0.58	– 1.43
Promotion and non-publishing spells						
Prob(promotion)	0.31	0.30	0.21	0.34	0.29	0.29
Prob(non-publishing triplets)	0.12	0.18	0.08	0.35	0.30	024
Collaboration						
log(number of authors harmonic average)	1.28	0.49	1.31	1.40	0.43	1.40
log(Articles national coauthors) –log(2)	0.05	0.84	0.00	– 0.11	0.81	– 0.69
log(IF national coauthors) –log(3)	0.24	0.95	0.66	– 0.01	0.94	0.34
log(National coauthors of coauthors.) –log(3)	0.00	0.86	0.25	– 0.11	0.88	0.15
No national coauthor	0.23	0.42	0.00	0.25	0.43	0.00
National coauthor no publication	0.08	0.28	0.00	0.13	0.34	0.00
L. (No publication)	0.04	0.19	0	0.10	0.30	0
Reference author	0.69	0.46	1.00	0.53	0.50	1.00
1 Single author	0.01	0.12	0.00	0.01	0.09	0.00
2 Multi-authored but unknown type	0.08	0.27	0.00	0.15	0.36	0.00
3 National (ref.)	0.21	0.41	0.00	0.36	0.48	0.00
4 National / Nuclear	0.03	0.17	0.00	0.02	0.15	0.00
5 National / International	0.54	0.50	1.00	0.41	0.49	0.00
6 National / Nuclear/International	0.12	0.33	0.00	0.05	0.21	0.00
From (1) to (6)	*1.00*	–	–	*1.00*	–	–
7 Coll. UN/CNRS 000 (ref.)	0.39	0.49	0	0.52	0.5	1
8 Coll. UN/CNRS 100/010/001	0.18	0.39	0.00	0.16	0.37	0.00
9 Coll. UN/CNRS 110/101/011	0.14	0.34	0.00	0.10	0.30	0.00
10 Coll. UN/CNRS 111	0.29	0.45	0.00	0.22	0.42	0.00
From (7) to (10)	*1.00*	–	–	*1.00*	–	–

Table A3. *Gender productivity gap estimated including sequentially one group of control variables at a time only**

CNRS: Log(Prod)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age and Gender							
Woman(= 1)	– 0.29***	– 0.29***	– 0.17***	– 0.27***	0.02	– 0.19***	– 0.17***
(Age-40)/10		– 0.06**	– 0.03	– 0.16***	– 1.19***	– 0.00	– 0.08***
((Age-40)/10)^2		– 0.06***	– 0.05***	– 0.06***	– 0.04**	– 0.07***	– 0.08***
(Age-40)/10 * Woman		0.09	0.07	0.11*	0.33***	0.11	0.17***
((Age-40)/10)^2* Woman		– 0.05	– 0.04	– 0.06	– 0.18***	– 0.04	– 0.05
Constant	3.76***	3.87***	2.97***	3.77***	2.84***	4.01***	5.04***
Time dummies	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Collaboration	*no*	*no*	*yes*	*no*	*no*	*no*	*no*
Rank	*no*	*no*	*no*	*yes*	*yes*	*no*	*no*
Rank endogeneity	*no*	*no*	*no*	*no*	*yes*	*no*	*no*
Selectivity	*no*	*no*	*no*	*no*	*no*	*yes*	*no*
Initial Productivity	*no*	*no*	*no*	*no*	*no*	*no*	*yes*
Observations	4,920	4,920	4,920	4,920	4,920	4,920	4,920
R2	0.009	0.02	0.244	0.033	0.126	0.071	0.126

UNIVERSITY: Log(Prod)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age and Gender							
Woman(= 1)	– 0.31***	– 0.31***	– 0.19***	– 0.18***	0.35***	0.04	– 0.27***
(Age-40)/10		– 0.13***	– 0.04*	– 0.37***	– 1.58***	– 0.11	– 0.09***
((Age-40)/10)^2		0.01	– 0.00	0.06***	0.36***	– 0.00	– 0.02
(Age-40)/10 * Woman		0.04	0.05	0.16***	0.61***	0.07	0.04
((Age-40)/10)^2 * Woman		– 0.06	– 0.05	– 0.07	– 0.21***	– 0.06	– 0.00
Constant	3.05***	3.17***	2.60***	2.95***	1.70***	3.54***	4.85***
Time dummies	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Collaboration	*no*	*no*	*yes*	*no*	*no*	*no*	*no*
Rank	*no*	*no*	*no*	*yes*	*yes*	*no*	*no*
Rank endogeneity	*no*	*no*	*no*	*no*	*yes*	*no*	*no*
Selectivity	*no*	*no*	*no*	*no*	*no*	*yes*	*no*
Initial Productivity	*no*	*no*	*no*	*no*	*no*	*no*	*yes*
Observations	6,219	6,219	6,219	6,219	6,219	6,219	6,219
R2	0.008	0.02	0.24	0.073	0.307	0.12	0.17

* Estimates of gender and age coefficients in productivity regression with gender alone and no other variables (column 1), with gender, age, age gender interactions and time dummies only (column 2), and then including sequentially one group of control variables only: collaboration only (column 3), rank only (column 4), with rank endogeneity only (column 5), non-publishing triplet selectivity only (column 6), and initial productivity only (column 7).

Table A4. *Gender productivity gap estimated excluding sequentially one group of control variables at a time only**

CNRS: Log(Prod)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age and Gender							
Woman(= 1)	– 0.29***	0.05	0.13**	– 0.11**	– 0.11**	0.04	0.01
(Age-40)/10		– 0.70***	– 0.97***	– 0.08***	– 0.03	– 0.71***	– 0.72***
((Age-40)/10)^2		– 0.05***	– 0.07***	– 0.06***	– 0.06***	– 0.05***	– 0.03*
(Age-40)/10 * Woman		0.24***	0.33***	0.11*	0.10*	0.24***	0.22***
((Age-40)/10)^2* Woman		– 0.11**	– 0.15***	– 0.04	– 0.04	– 0.12**	– 0.12**
Constant	3.76***	3.28***	4.07***	3.75***	3.81***	3.25***	2.48***
Time dummies	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Collaboration	*no*	*yes*	*no*	*yes*	*yes*	*yes*	*yes*
Rank	*no*	*yes*	*yes*	*yes*	*no*	*yes*	*yes*
Rank endogeneity	*no*	*yes*	*yes*	*no*	*no*	*yes*	*yes*
Selectivity	*no*	*yes*	*yes*	*yes*	*yes*	*no*	*yes*
Initial Productivity	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*no*
Observations	4,920	4,920	4,920	4,920	4,920	4,920	4,920
R2	0.009	0.312	0.229	0.283	0.28	0.311	0.28

UNIVERSITY: Log(Prod)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age and Gender							
Woman(= 1)	– 0.31***	0.25***	0.30***	– 0.05	– 0.09*	0.19***	0.31***
(Age-40)/10		–1.09***	– 1.19***	– 0.12***	0.001	– 1.12***	– 1.19***
((Age-40)/10)^2		0.25***	0.28***	0.01	– 0.01	0.25***	0.28***
(Age-40)/10 * Woman		0.43***	0.49***	0.10*	0.043	0.43***	0.48***
((Age-40)/10)^2 * Woman		– 0.14***	– 0.15***	– 0.04	– 0.03	– 0.14***	– 0.17***
Constant	3.05***	2.66***	3.06***	3.70***	3.90***	2.57***	1.68***
Time dummies	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Collaboration	*no*	*yes*	*no*	*yes*	*yes*	*yes*	*yes*
Rank	*no*	*yes*	*yes*	*yes*	*no*	*yes*	*yes*
Rank endogeneity	*no*	*yes*	*yes*	*no*	*no*	*yes*	*yes*
Selectivity	*no*	*yes*	*yes*	*yes*	*yes*	*no*	*yes*
Initial Productivity	*no*	*yes*	*yes*	*yes*	*yes*	*yes*	*no*
Observations	6,219	6,219	6,219	6,219	6,219	6,219	6,219
R2	0.008	0.37	0.353	0.303	0.291	0.368	0.344

* Estimates of gender and age coefficients in productivity regression with no other variables (column 1), with all variables (column 2), and then excluding group of variables one at a time: collaboration only (column 3), rank endogeneity only (column 4), rank and rank endogeneity only (column 5), non-publishing triplet selectivity only (column 6), and initial productivity only (column 7).

APPENDIX B

MAIN RESULTS FOR TWO VARIANTS OF THE BASELINE ECONOMETRIC ANALYSIS, CONSIDERING ONLY PUBLICATIONS IN HIGH AND VERY HIGH QUALITY JOURNALS, WITH RESPECTIVELY FIVE YEAR AVERAGE IMPACT FACTOR IF >5 AND IF >10

Table B1. *Configurations of CNRS and university unbalanced panel data study samples for the base line econometric analysis and the two IF > 5 and IF > 10 variants**

	WOMEN-CNRS	MEN-CNRS	WOMEN-UNIV	MEN-UNIV
IF > 0.5 (baseline analysis)				
Physicists	159	839	410	1,403
Total number of observations	815	4,702	1,848	7,408
Publishing spells	648	4,272	804	5,415
IF > 5				
Physicists	134	779	207	1,104
Total number of observations	794	4,575	1,774	7,003
Publishing spells	566	3,858	573	4,010
IF > 10				
Physicists	109	696	130	773
Total number of observations	670	4,180	1,639	6,045
Publishing spells	334	2,777	278	2,117

* Total numbers of physicists and observations for all triplets and for publishing triplets with respectively IF > 0.5, IF > 5 and IF > 10. The baseline CNRS and university samples consider all the publications in the full set of 368 physic journals with IF > 0.5, while the IF > 5 and IF > 10 variant samples consider only the publications in the subsets of 113 and 27 physic journals with higher five year average impact factors respectively.

Table B2. *Descriptive statistics on average log publication productivity for woman and man CNRS and university physicists for the IF > 5 and IF > 10 variant samples**

(A) Only journals with an IF larger than 5 (IF > 5)	WOMEN-CNRS	MEN-CNRS	W-M	WOMEN-UNIV	MEN-UNIV	W-M
In logarithms (excluding lower quality publishing triplets)						
Mean	3.44	3.72	– 0.28	2.97	3.17	– 0.20
Median	3.53	3.78		2.91	3.15	
Std dev.	0.89	0.92		0.89	0.93	
Obs.	566	3,858		573	4,010	

(B) Only journals with an IF larger than 10 (IF>10)	WOMEN-CNRS	MEN-CNRS	W-M	WOMEN-UNIV	MEN-UNIV	W-M
In logarithms (excluding lower quality publishing triplets)						
Mean	3.43	3.62	– 0.19	3.08	3.25	– 0.17
Median	3.48	3.58		3.06	3.10	
Std dev.	0.73	0.81		0.78	0.77	
Obs.	334	2,777		278	2,117	

* Table in the same format as Table 4 in the text.

Table B3 (IF > 5). *Proportion of non-publishing triplets, average publication productivity, and total number of observations for female and male CNRS and university physicists in two age groups and low and high rank in the last triplet 2003-2005 for the IF > 5 variant samples**

	WOMEN-CNRS	MEN-CNRS	W-M	WOMEN-UNIV	MEN-UNIV	W-M
Low Ranks (CR/MCF)						
% of lower qual. pub. triplets	35%	21%	14%	70%	54%	16%
Average log-productivity	3.28	3.51	– 0.23	2.91	2.95	– 0.04
Observations	*419*	*1,631*		*1399*	*3,521*	
High Ranks (DR/PR)						
% of lower qual. pub. triplets	22%	13%	9%	58%	31%	27%
Average log-productivity	3.60	3.81	– 0.21	3.10	3.32	– 0.22
Observations	*375*	*2,944*		*375*	*3,482*	
Less than 40 years						
% of lower qual. pub. triplets	13%	8%	5%	56%	25%	31%
Average log- productivity	3.47	3.84	– 0.37	2.96	3.26	– 0.3
Observations	*78*	*487*		*334*	*856*	
40 years and more						
% of lower qual. pub. triplets	30%	17%	13%	70%	45%	25%
Average log-productivity	3.44	3.70	– 0.26	2.97	3.15	– 0.18
Observations	*716*	*4,088*		*1440*	*6,147*	
Total						
% of lower qual. pub. triplets	29%	16%	13%	68%	43%	25%
Average log-productivity	3.44	3.72	– 0.28	2.96	3.17	– 0.21
Observations	*794*	*4,575*		*1,774*	*7,003*	

* Table in the same format as Table 5 in the text.

Revue économique – vol. 66, N° 1, janvier 2015, p. 65-114

Table B4 (IF > 10). *Proportion of non-publishing triplets, average publication productivity, and total number of observations for female and male CNRS and university physicists in two age groups and low and high rank in the last triplet 2003-2005 for the IF > 10 variant samples**

	WOMEN-CNRS	MEN-CNRS	W-M	WOMEN-UNIV	MEN-UNIV	W-M
Low Ranks (CR/MCF)						
% of lower qual. pub. triplets	56%	41%	15%	84%	75%	9%
Average log-productivity	3.35	3.45	– 0.1	3.03	3.10	– 0.07
Observations	*379*	*1,506*		*1,332*	*3,192*	
High Ranks (DR/PR)						
% of lower qual. pub. triplets	43%	30%	13%	78%	54%	24%
Average log-productivity	3.51	3.70	– 0.19	3.19	3.35	– 0.16
Observations	*291*	*2,674*		*307*	*2,853*	
Less than 40 years						
% of lower qual. pub. triplets	29%	18%	11%	74%	46%	28%
Average log- productivity	3.39	3.61	– 0.22	3.12	3.25	– 0.13
Observations	*62*	*416*		*308*	*744*	
40 years and more						
% of lower qual. pub. triplets	52%	35%	17%	85%	68%	17%
Average log-productivity	3.64	3.67	– 0.03	2.97	3.26	– 0.29
Observations	*608*	*3,764*		*1,331*	*5,301*	
Total						
% of lower qual. pub. triplets	50%	34%	16%	83%	65%	18%
Average log-productivity	3.43	3.62	– 0.19	3.08	3.25	– 0.17
Observations	*670*	*4,180*		*1,639*	*6,045*	

* Table in the same format as Table 5 in the text.

Table B5 (IF > 5). *Promotion probit equation (estimated coefficients) for* CNRS *and university physicists, with and without age *gender interaction for the IF > 5 variant samples**

Rank indicator	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (= 1)	– 0.28***	– 0.53***	– 0.24**	– 0.50***
(Age-40)/10	1.81***	1.46***	1.82***	1.46***
((Age-40)/10)^2	– 0.48***	– 0.51***	– 0.48***	– 0.51***
(Age-40)/10 * Woman			– 0.02	– 0.06
((Age-40)/10)^2 * Woman			– 0.02	0.01
Lagged productivity				
L.log(Art)-log(6)	0.25***	0.40***	0.25***	0.40***
L.log(avg IF)-log(5)	0.24***	0.25***	0.25***	0.25***
L.No Publications	– 0.28***	– 0.95***	– 0.28***	– 0.95***
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	– 0.73***	– 0.12*	– 0.74***	– 0.13*
Observations	5,369	8,777	5,369	8,777
Pseudo R2	0.38	0.26	0.38	0.26

* Table in the same format as Table 6 in the text.

Table B6 (IF > 5). *Publishing triplet selectivity Probit equation (estimated coefficients) for* CNRS *and university, with and without age *gender interaction for the IF > 5 variant samples**

Publishing indicator	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (= 1)	– 0.18***	– 0.36***	– 0.23***	– 0.39***
(Age-40)/10	– 0.11***	– 0.09***	– 0.12***	– 0.12***
((Age-40)/10)^2	– 0.00	– 0.01	– 0.01	– 0.01
(Age-40)/10 * Woman			0.05	0.09
((Age-40)/10)^2 * Woman			0.04	0.01
Productivity persistence				
L.Persistence 111	2.43***	2.09***	2.43***	2.10***
L Persistence 110/101/011	1.61***	1.30***	1.61***	1.31***
L.Persistence 100/010/001	1.01***	0.80***	1.01***	0.81***
Reference L.Persistence 000 (ref.)	–	–	–	–
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	1.50***	1.01***	1.51***	1.02***
Observations	5,369	8,777	5,369	8,777
Pseudo R2	0.31	0.30	0.31	0.30

* Table in the same format as Table 7 in the text.

Table B7 (IF > 5). *Productivity equation for CNRS and university physicists, with and without age*gender interactions for the IF > 5 variant samples**

Productivity: Log(Prod)	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (= 1)	– 0.04	0.13***	0.01	0.32***
(Age-40)/10	– 0.51***	– 0.43***	– 0.54***	– 0.95***
((Age-40)/10)^2			– 0.03	0.25***
(Age-40)/10 * Woman			0.24***	0.39***
((Age-40)/10)^2* Woman			– 0.13***	– 0.15***
Initial Productivity				
No first publications	-0.58***	-1.08***	– 0.62***	– 0.99***
log(First Art) –log(6)	0.18***	0.24***	0.18***	0.21***
log(avg. first IF) – log(5)	0.16***	0.33***	0.16***	0.30***
Promotion and lower-quality publishing spells				
Prob(promotion)	1.48***	1.66***	1.56***	2.63***
Prob(lower-quality publishing triplets)	-0.23**	-0.27***	– 0.22**	– 0.24***
Collaboration	*yes*	*yes*	*yes*	*yes*
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	3.37***	3.63***	3.38***	3.02***
Observations	4,424	4,583	4,424	4,583
Pseudo R2	0.261	0.313	0.265	0.339

* Table in the same format as Table 8 in the text.

Table B8 (IF > 10). *Promotion probit equation (estimated coefficients) for CNRS and university physicists, with and without age *gender interaction for the IF >10 variant samples.*

Rank indicator	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (=1)	– 0.28***	– 0.63***	– 0.24**	– 0.63***
(Age-40)/10	1.82***	1.44***	1.82***	1.44***
((Age-40)/10)^2	– 0.47***	– 0.50***	– 0.47***	– 0.51***
(Age-40)/10 * Woman			– 0.02	– 0.02
((Age-40)/10)^2* Woman			– 0.03	0.01
Lagged productivity				
L.log(Art)-log(6)	0.24***	0.33***	0.24***	0.33***
L.log(avg IF)-log(5)	0.53***	0.50***	0.53***	0.50***
L.No Publications	0.027	– 0.64***	0.027	– 0.64***
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	– 1.01***	– 0.30**	– 1.02***	– 0.30**
Observations	4,850	7,684	4,850	7,684
Pseudo R2	0.36	0.23	0.36	0.23

Footnote: Table in the same format as Table 6 in the text.

Table B9 (IF > 10). *Publishing triplet selectivity Probit equation (estimated coefficients) for* CNRS *and university, with and without age *gender interaction for the IF >10 variant samples**

Publishing indicator	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (= 1)	– 0.17***	– 0.36***	– 0.23***	– 0.36***
(Age-40)/10	0.00	– 0.10***	0.00	– 0.09***
((Age-40)/10)^2	– 0.10***	– 0.00	– 0.10***	– 0.01
(Age-40)/10 * Woman			0.03	– 0.06
((Age-40)/10)^2 * Woman			0.06	0.02
Productivity persistence				
L.Persistence 111	1.96***	1.88***	1.96***	1.88***
L Persistence 110/101/011	1.28***	1.19***	1.28***	1.19***
L.Persistence 100/010/001	0.94***	0.84***	0.94***	0.84***
Reference L.Persistence 000 (ref.)	–	–	–	–
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	1.21***	0.75***	1.22***	0.76***
Observations	4,850	7,684	4,850	7,684
Pseudo R2	0.23	0.28	0.23	0.28

* Table in the same format as Table 7 in the text.

Table B10 (IF > 10). *Productivity equation for CNRS and university physicists, with and without age*gender interactions for the IF > 10 variant samples**

Productivity: Log(Prod)	CNRS	UNIVERSITY	CNRS	UNIVERSITY
Age and Gender				
Woman (=1)	– 0.00	0.10**	0.06	0.38***
(Age-40)/10	– 0.38***	– 0.24***	– 0.41***	– 0.71***
((Age-40)/10)^2			0.01	0.21***
(Age-40)/10 * Woman			0.17**	0.41***
((Age-40)/10)^2 * Woman			– 0.12**	– 0.20***
Initial Productivity				
No first publications	– 0.83***	– 0.97***	– 0.83***	– 0.88***
log(first Art) –log(6)	0.18***	0.23***	0.18***	0.21***
log(avg. first IF) – log(5)	0.18***	0.20***	0.18***	0.18***
Promotion and lower-quality publishing spells				
Prob(promotion)	1.09***	1.13***	1.14***	2.09***
Prob(lower-quality publishing triplets)	– 0.18**	– 0.18***	– 0.17**	– 0.20***
Collaboration	*yes*	*yes*	*yes*	*yes*
Time dummies	*yes*	*yes*	*yes*	*yes*
Constant	4.28***	3.83***	4.25***	3.29***
Observations	3,111	2,395	3,111	2,395
Pseudo R2	0.235	0.250	0.237	0.272

* Table in the same format as Table 8 in the text.