The Effect of Collaboration with top-funded Scholars on Scientific Production

Authors:

Seyed Reza Mirnezami, srmirnezami@sharif.edu

Catherine Beaudry, catherine.beaudry@polymtl.ca

Leila Tahmooresnejad, leila-2.tahmooresnejad@polymtl.ca

Word count: 10598

Abstract

The theoretical model developed in this paper predicts that collaboration with top-funded scientists positively affects the number of scientific publications of an individual scientist. Having combined data on funding and publication of Quebec scientists, this paper empirically tests the theory predictions. The paper examines numerous definitions of top-funded scientists as those in the top 10%, or top 5% in terms of total funding, funding from the public sector, and funding from the private sector. The results show that collaborating with such top-funded scientists has a positive effect on a scientist's number of publications, hence confirming our theoretical predictions.

Keywords: Scientist Performance; Collaboration; Publication Determinants; Research Funding; Top-Funded Scientists

1 Introduction

Scientists' academic performance has been extensively discussed in and out of the literature and many of its determinants are currently known as potential motives for publishing papers in peer reviewed journals. Among others, age, gender, private and public funding, field and context are the most important factors. Funding definitely plays a major role in shaping scientific production and has been extensively investigated in the literature. A number of scholars show the positive effect of funding on academic production (Crespi and Geuna, 2008; Pavitt, 2000, 2001; Salter and Martin, 2001). Others find evidence on the effect of different types and sources of funding, some of which indicate that the effect of public funding and private funding are both positive (Berman, 1990; Gulbrandsen and Smeby, 2005). Other studies found that funding from the private sector has a detrimental effect on scientific production (Goldfarb, 2008; Kleinman and Vallas, 2001).

In addition to funding, scientific collaboration and academic networking can also positively affect scientific production. Using data on collaborative research conducted in Canada, Godin and Gingras (2000) indicated that not only is there a non-negative correlation between research quality and its collaborative nature, but that new research opportunities and scientific production are developed during research collaboration. Melin (1996) also referred to the positive effect of co-authorship on scientific production at the local, national, and international levels. Research collaboration is also a positive determinant of patenting in addition to scientific publication. Azoulay et al. (2007) provided some evidence that having a co-author who patented before “increases the likelihood of a patent application” (pp. 601) and also that scientists in universities that have good patenting records generally patent more. In another study, Zhou (2003) also identified the importance of having a creative co-worker in leading to increased creativity for a researcher.

The networking effect can also be investigated from the standpoint of peripheral support in universities and research institutions. For instance, Crane (1965) highlighted the effect of university prestige on research production and Johnes (1988) indicated that the number of staff in a department and the number of co-authors of published articles can significantly explain the research production of UK economic departments.

Combining the known effect of networking and scientific collaboration on the one hand and a scientist’s funding capacity on the other hand, provides the perfect setting to investigate the effect of collaborators’ funding on scientific production. Collaboration with top-funded scientists can be an opportunity for accumulating valuable experience and tacit knowledge, resulting in higher and better scientific production. This investigation is of great importance when considering the number of programmes targeted at attracting talent by setting up prestigious chairs, fellowships and grants in academia. For instance, the Canada Research Chair (CRC) programme aims to help Canadian universities for ‘‘attracting and retaining leading researchers’’ from other countries through expanding fruitful academic collaboration. Among others, there are quite similar programmes to absorb talents in Germany (Humboldt Research Fellowship), Switzerland (Mobility fellowships), France (Chateaubriand Fellowship), European Union (Marie Sklodowska-Curie Fellowships) and Japan (Monbukagakusho Scholarship).

This study thus tests whether collaboration with top-funded scientists shifts up a scientist's performance. The remainder of paper is organised as follows: A brief literature review and the theoretical framework are presented in section 2; the data and the econometric model are described in section 3; analysis of econometric results is the focus of section 4; and finally section 5 concludes.

2 Theoretical model and empirical framework

This paper investigates the effect of collaboration with top-funded scientists from both theoretical and empirical points of view. The theoretical model proposed in this paper tries to quantify the power and quality of a scientist's collaboration in terms of production of publishable knowledge. The transformation of collaboration into scientific production is explained by a utility function of researchers which is linear and risk neutral: A researcher compares the utility gained by publishing with the research cost, salaries and stipends paid to research assistants and graduate students. The ensuing empirical analysis then tests the theoretical model.

A brief review of the empirical literature is first preseted to show its compatibility with our theoretical model. Inspired by Jensen and Meckling (1976) and by Durnev and Kim (2005), who developed concepts to model the incentives of firms' controller, the following paragraphs model the incentives of a researcher to publish a paper. This model aims to explain the cost-benefit process of a researcher to publish a paper.

A number of assumptions are required prior to developing the model:

1- A scientific collaboration between two researchers affects the research results. The power and quality of a scientist's collaboration in terms of production of publishable knowledge is represented by a constant 'α' that is set between 0 and 1. In other words, 'α' is the probability of acceptance of a submitted paper (when focusing on the number of papers) or the level of its impact on future research in the field (when focusing on the quality of papers).

2- The maximum contribution of all possible external factors of publication is represented by $\bar{\pi}$. In other words, it is the 'maximum external potential of knowledge production' (consequences of the factors affecting scientific production such as research budget or research support from the university at organizational level).

3- The utility function of researchers is linear and risk neutral. A researcher compares the utility gained by publishing with the research cost and salaries paid to research assistants. We suppose that every unit of scientific outcomes, which is published, decreases the external potential of the knowledge production. In addition, such external potential should be summed up with personal willingness/ability of a scientist to publish a paper. This sum shows the aggregate possibility of scientific publications. In our utility function, this willingness is normalized to 1. The j^{th} unit of external and personal scientific potential is multiplied by α, the factor of collaboration effectiveness to calculate the gained utility from the j^{th} article. Then it is compared with the constant cost of publication (c) to determine whether it is worth publishing or not.

If the j^{th} article is being published, the following inequality should be valid, indicating that the benefit of a publication should be more than its cost. The left hand-side of the inequality represents the utility/benefit/quality of one unit of scientific outcome or knowledge when it becomes a paper. The right hand-side of the inequality represents the cost of publication.

$$\alpha\left(1+\pi(j)\right)>c$$
$$where\ \pi(j)=\bar{\pi}-j$$

4- Our last assumption states that if there is no support for a scientist $\left(\pi(j)=0\right)$, publishing is not possible ($c > \alpha$). It means that a scientist cannot publish only based on personal characteristics and without getting support from the academic community.

The next step consists in finding the optimum number of papers $\left(j^*\right)$. The utility gained from the j^{th} publication is decreasing in the number of articles. Therefore, j increases until the inequality becomes equal.

$$j^*=\frac{\alpha\left(1+\bar{\pi}\right)-c}{\alpha}$$

If j^* is divided by the 'maximum external potential of knowledge production' $(\bar{\pi})$, we can calculate the percentage of external scientific potential which has been published as papers p^*. This p^* is a standardized level of publication regardless of the amount of external support and funding. It is a more appropriate tool for our analysis than j^* because j^* cannot be compared among the different types of external support that the scientists receive.

$$p^* = \frac{\alpha(1+\bar{\pi})-c}{\alpha\bar{\pi}} = 1+\frac{\alpha-c}{\alpha\bar{\pi}} = 1+\frac{1}{\bar{\pi}}-\frac{c}{\alpha\bar{\pi}}$$

It is possible to find the sensibility of p^* to factors such as $\bar{\pi}$, c, and α to extract the theoretical hypotheses that will be tested empirically further in the paper. The above model suggests the following statements:

1- Greater institutional and peripheral support results in more publications. This statement results directly from our theoretical model:

$$\frac{\partial p^*}{\partial \bar{\pi}} = -\frac{\frac{\alpha-c}{\alpha}}{\bar{\pi}^2} > 0$$

2- Researchers with higher costs of publication are less likely to publish:

$$\frac{\partial p^*}{\partial c} = -\frac{1}{\alpha\bar{\pi}} < 0$$

3- Higher ‘power and quality of scientific collaboration’ (higher α) improves publication performance. Higher α in our model refers to collaboration with top-funded scientists.

$$\frac{\partial p^*}{\partial \alpha} = -\frac{c}{\alpha^2\bar{\pi}} > 0$$

This study then aims to empirically investigate the effect of collaboration with top-funded scientists on the number of scientific publications, which is related to the first and third predictions of our theoretical model. Acknowledging the proximity between funding and publishing, we consider that top-funded scientists are probably also star scientists.

There are numerous definitions of star scientists in the literature. For instance, Higgins et al. (2011) considered the recipients of a Nobel Prize as star scientists and showed that if they are affiliated to a biotechnology company they act as a signal of research quality to absorb resources and investment. The authors justify this by reasons such as Noble prize winners transfer tacit knowledge or bring valuable network to the company. In another study, Meyer (2006) proposed a different definition of star scientist (one who is active in both patenting and publication) and showed that inventor-authors' publications are over-proportionally high with a high number of citations, suggesting that the combination of publication and patenting activities is a good proxy for being a star scientist.

The common element to different definitions of star scientists reside in the order of magnitude of the attributes considered. A star scientist should have more than an ordinary or average effect. In this paper, we propose to also consider top-funded scientists as stars. Numerous articles indeed provide evidence to the effect that funding impacts scientific production, as a costly activity, mainly through supporting research teams or providing scientific infrastructure (e.g. Beise and Stahl (1999), Pavitt (2001), and Pavitt (2000)). In a review article of the effect of funding, Salter and Martin (2001) suggested the following six types of contribution of publicly funded research: new knowledge, advance instrumentation and methodologies, skills to conduct basic research, expansion of research networks, dealing with complex problems, and establishing spin-off companies. Considering the close relationship between funding and scientific production, it is reasonable enough to also consider top-funded scientist, i.e. those that have the means to publish more, as stars. Because of the substantial effects of 'funding' and 'collaboration' on scientific production from both qualitative and quantitative points of view, this paper argues that the combination of the two concepts may propagate their beneficial effects on scientific production.

The literature also provides evidence to the effect that research collaboration is an important determinant of scientific production. Scientific collaboration and being an active member of a scientific network positively affect scientific production (Azoulay et al., 2007; Chwe, 2000; Godin and Gingras, 2000; Hicks, 1995; Johnes, 1988; Melin, 1996). Scientific collaboration can also be analysed from a social network point of view. Chwe (2000) analysed social networks using a game theoretic model and identified three levels of hierarchy as the minimal sufficient network structure for efficient collaboration. These hierarchical stages are 'initial adopters', 'followers', and then 'late adopters'. The author highlighted that "a communication network helps coordination in exactly two ways: by informing each stage about earlier stages, and by creating common knowledge within each stage" (pp. 1). The implication of such structure in the scientific community would thus be a network in which scientists collaborate and produce new outputs, supposing that they know past information. In such capacity, several studies highlight the impact of collaboration and co-inventorship network characteristics on the quality of inventions (Bagchi-Sen, 2007; Beaudry and Schiffauerova, 2011; Bercovitz and Feldman, 2011; Darby and Zucker, 1996; Rothaermel and Hess, 2007; Zucker and Darby, 1996, 1997, 2001; Zucker et al., 2002).

Combining these two lines of reasoning as determinants of scientific production, we may expect that a 'greater than usual' impact, one that amplifies the effect of two independent contributors, funding and collaboration, be associated with the concept of star scientist. Both the effects of funding and collaboration on scientific production have been well investigated in literature. To the best of our knowledge, however, there is not a comprehensive study on the effect of collaborators' funding on scientific production. Based on the known correlation between funding and scientific production, we consider funding as a proxy of being star. In other words, being a top-funded scientist is an instrument showing the intrinsic skills and abilities of scientist. This research tests whether collaborating with top-funded scientists has an effect on the scientific production measured by the number of articles. Hence our hypothesis reads as follows:

Hypothesis:

Collaboration with top-funded scientists positively affects scientific production.

Other determinants of scientific production have been investigated in the literature, and they should therefore be considered in our empirical analysis as control variables. For instance, gender is known as a significant determinant of scientific production in the literature. Some researchers argue that differences in research production between men and women come from administrative positions of researchers and their marital status. Xie and Shauman (1998) indicated that "women scientists publish fewer papers than men because women are less likely than men to have the personal characteristics, structural positions, and facilitating resources that are conducive to publication" (pp. 863). Xie and Shauman (1998) nonetheless concluded that gender differences in research production has declined over time, while at the same time the population of female scientists has proportionally increased - this decline is also observed in Abramo et al. (2009). Fox (2005) argued that the effect of gender is complex in a way that it is not possible to simply separate the effect of married and single researchers and he refers to the career of the spouse and to the family composition as two important factors of such complexity. For example, women with preschool children show higher production than women without children and women with school-age children. Nakhaie (2002) also investigated the gender effect in different durations of time and argued that Canadian female professors publish less than their male colleagues do, both in a lifetime period and during a shorter period, but that such effect is higher for the former.

Age is another independent determinant of scientific production examined in the literature. There is evidence that the effect of age varies between different disciplines (Kyvik, 1990). The authors indicated that the production level in the social sciences is independent of age. In the humanities, however, publishing activity declines during the age-period of 55-59, but it is followed by a new peak in the group of 60 years old and over. There is a different story in the medical sciences in which the production falls when researchers reach 55 years and older. In the natural sciences, production continuously decreases with aging because new scientific tools, methods, and equipment are continuously introduced and older researchers may have problems becoming familiar with them (Kyvik, 1990).

The inverted U-shaped effect of age (Quadratic or second-order effect of age) receives significant support in the literature. In a comprehensive study, Kyvik and Olsen (2008) tried to justify the hypotheses in the literature explaining the age effect on academic production and categorized them in six groups. All of the following hypotheses have been locally verified based on the data set used: (a) The utility maximizing hypothesis implies that academic staff conduct less research as they age because the expected utility of time spent on research diminishes; (b) The seniority burden hypothesis refers to the increasing administrative load as a career advances, which decrease the focus on scientific matter; (c) The cumulative disadvantage hypothesis suggests that scientists who do not win research awards will gradually lose their incentives for further research; (d) The age decrement hypothesis proposes that older scientists mostly conduct research with a lower intellectual and physical level than that of their younger colleagues; (e)

The obsolescence hypothesis implies that the younger scientists use novel tools, techniques, and methodologies for research more easily than their older colleagues; (f) The intellectual deadlock hypothesis suggests that older scientists have less tendency to "reorient their research towards new scientific or social problems" (Kyvik and Olsen, 2008, pp. 442).

In terms of funding, Manjarrés-Henríquez et al. (2009) found that industrial R&D contracts and private funding are effective if the professor is in need of money. In the other words, "R&D contracts with industry and academic research activities have synergistic effects on scientific production, but only when R&D contracts account for a small percentage of a researcher's total funding, otherwise, there are decreasing marginal returns to scientific output" (pp. 799). Feldman and Graddy-Reed (2013) also showed some evidence about the positive effect of philanthropic funding coming from not-for-profit (NFP) organisations.

There are many examples for positive effect of public funding on scientific production. One famous justification for developing public support of science is that it is a public good (Partha and David, 1994). From another standpoint, Callon (1994) argued that science is not a public good because of its intrinsic and natural properties, but a public good due to the fact that it is a source of diversity and flexibility. The fall in proportion of public funding relative to the other sources of funds has some influence on the university research trend. Geuna (2001) argued that the fading role of public funding can result in over-use of resources, focus on short-term research endeavour, conflict in incentive structures, and "exacerbation of the impact of cumulative and self-reinforcement phenomena present in the process of scientific production" (pp. 626). In other words, the abovementioned disadvantages can become positive when public funding increases. In terms of private funding, there are some articles questioning the positive and significant effect of private funding on scientific production, implying that private funding has either a neutral or a detrimental effect (Hottenrott and Thorwarth, 2011; Kleinman and Vallas, 2001).

Other university-specific effects also influence scientists' production. Crane (1965) indicated that scientists in major universities are more likely to be productive than scientists in minor universities. The main justification for such a phenomenon comes from the fact that more prestigious universities are better able to select the best students, are richer and hence can recruit better researchers, and are more apt at providing research procurement. Kyvik (1990), Blackburn et al. (1978), and Landry et al. (2007) also mentioned the variation of publication counts in different fields.

3 Data and methodology

Data and variables

In order to validate the hypothesis, a data set based on the integration of Quebec scientists' funding and publications was built. In terms of publications, Thompson Reuters Web of Science provides information on scientific articles (date of publication, journal name, authors, coauthors, citations, and authors' affiliations). In addition, the Quebec University Research Information System (*Système d'information sur la recherche universitaire* or SIRU) of the Ministry of Education, Leisure and Sports was used to extract the grants and contracts, including yearly amounts, source, type, or other funding information of all Quebec university scientists during the period 2000-2012. These two databases were merged based on a scientist's unique identifier that was perfectly disambiguated by years of work at the Observatoire des sciences et des technologies (OST) in Montreal.

The dependent variable in our regression equations is an indicator of scientist's production and is measured by the natural logarithm of the yearly number of published articles [*ln(nbArticle)*].

The pertinent variables to validate our hypothesis identifies whether amongst the collaborators of an individual, there are co-authors who are in the top 10% and top 5% most funded individuals, distinguishing total funding, public sector funding, and private sector funding. To generate the variables measuring collaboration with top-funded scientists, a set of dummy variables has been generated to identify whether a scientist has collaborated with a top funded scientist: *ColT90*, and *ColT95* are the dummy variables that are equal to 1 if any of the coauthors in that year is amongst the top funded scientists (top 10%, and top 5% of total funding respectively). The funding amount used to generate these dummies is the sum of operational funding and funding for purchasing instruments. The variables of *ColPub90* and *ColPub95* are similarly built but only for public funding of coauthors and the variables of *ColPriv90* and *ColPriv95* have the same method of generation but for private funding of coauthors.

The database also includes other determinants of scientific production. Research funding can be awarded from different sources: the public sector, the private sector, or an organization with social and political missions, i.e. not-for-profit organisation (NFP). In addition, research funds can serve two purposes: it is directly used for research cost and researchers' salary (operational cost – O) or it indirectly helps research teams in buying instrument or logistic expenditure of laboratories (infrastructure costs – I). With these two mentioned categories and the three types of funding sources, it is possible to generate six research funding variables for each researcher [*ln(PublicfundingO)*, *ln(PublicfundingI)*, *ln(PrivatefundingO)*, *ln(PrivatefundingI)*, *ln(NFPfundingO)*, and *ln(NFPfundingI)*]. Although we have information for these six variables, this research only focuses on the effect of the operational budget because funding for the purpose of research tools and instruments does not have a regular pattern, i.e. it depends on the research needs, field, and handiness of updated research instruments. Hence infrastructure grants will not be used in this paper. To make the individual funding more informative, the amount of funds which is awarded to a team of researchers should be divided by the number of researchers in the team. In addition, the funding variables are measured in three-year averages to smooth out large variations in yearly funding.

To complete the dataset, we add the yearly average number of authors in the papers of each scientist [*nbAuthor*]. In addition, individual socio-demographic characteristics regarding age and gender of scientists are also added to the data [*Age*, *dFemale*]. Appendix 1 reviews the names and descriptions of the variables in the data set and appendix 2 summarizes the variables' descriptive statistics.

Methodology and econometrics model

To measure the effect of "collaboration with top-funded scientists" on a scientist's performance (the number of articles), a regression equation is fitted to the available data. We have access to information over a number of years for each researcher, which allow the construction of a panel database. In addition to the dummy variables of collaboration with top-funded scientists as the main independent right-hand-side (RHS) variable, the left-hand-side (LHS) variable of the regression should also include a number of control variables that affect the number of articles.

The main RHS comprises of the dummy variables indicating whether the scientist has any collaboration with top-funded scientists [*ColT90, ColT95, ColPub90, ColPub95, ColPriv90, ColPriv95*], as well as funding for research operational costs [*ln(PublicfundingO), ln(PrivatefundingO), ln(NFPfundingO)*].

One may argue that research collaboration increases scientific production because of the simple fact that writing an article as a single author may take double the time of writing an article with a second author. To control for this issue and for the obvious impact it may have on scientific production, we have added the lagged average number of co-authors [*ln(nbAuthor)*] to the RHS variables.

In addition, we control for a number of fixed effects to account for any impact that our explanatory variables do not cover. Control dummy variables include gender [*dFemale*], age [*Age*], research field, universities, and year. For example, McGill University and University of Montreal (UdeM) produce more scientific publications (figure 1). The small universities are grouped according to their active disciplines and other institutional similarities. The University of Quebec (excluding the three entities listed in the next sentence) and Bishop University are in the same group. The second group includes "École de technologie supérieure" (ETS), "Université du Québec à Montréal" (UQAM), and "Institut national de la recherche scientifique" (INRS), all specific constituents of the Université du Québec network. Figure 2 shows that the fields of Science, Engineering, Medical Science, and Health Science are more productive than others. We also add year dummy variables to account for year-specific characteristics of the research system as exemplified by the evolution of article counts over time by university group (figure 1), research field (figure 2), and individual scientist (figure 3). The significant time trend and differences between different universities/research fields justify the existence of these dummy variables in the model.

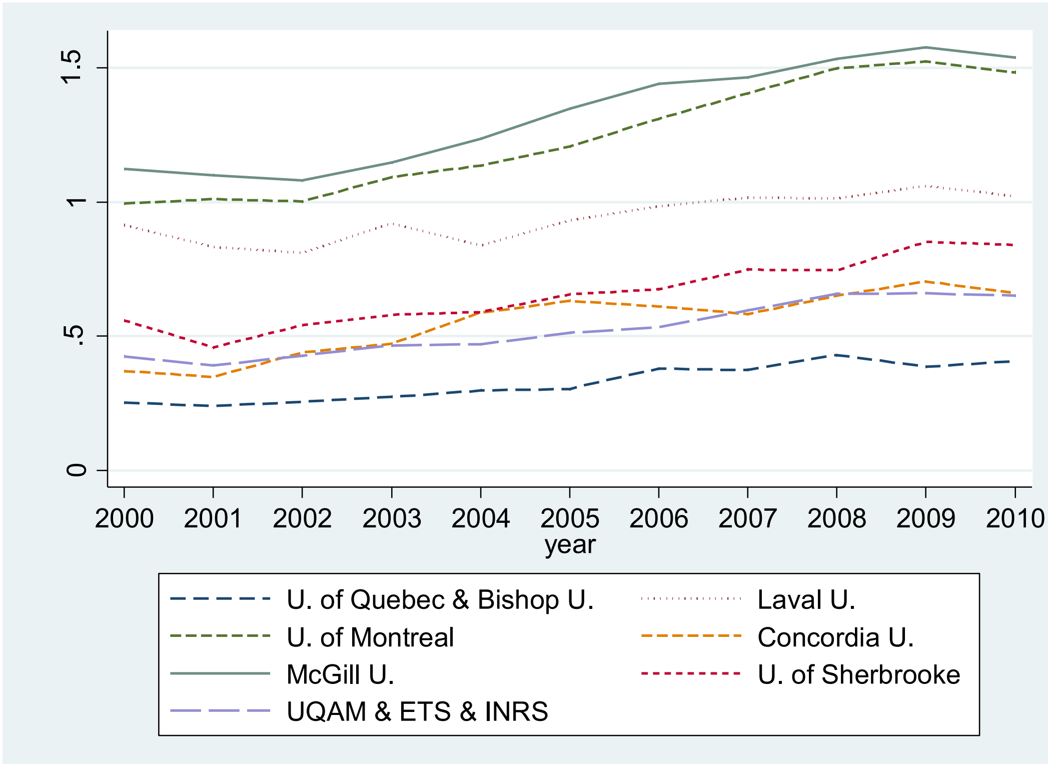

Figure 1 – Average number of articles published by the scientists of each university

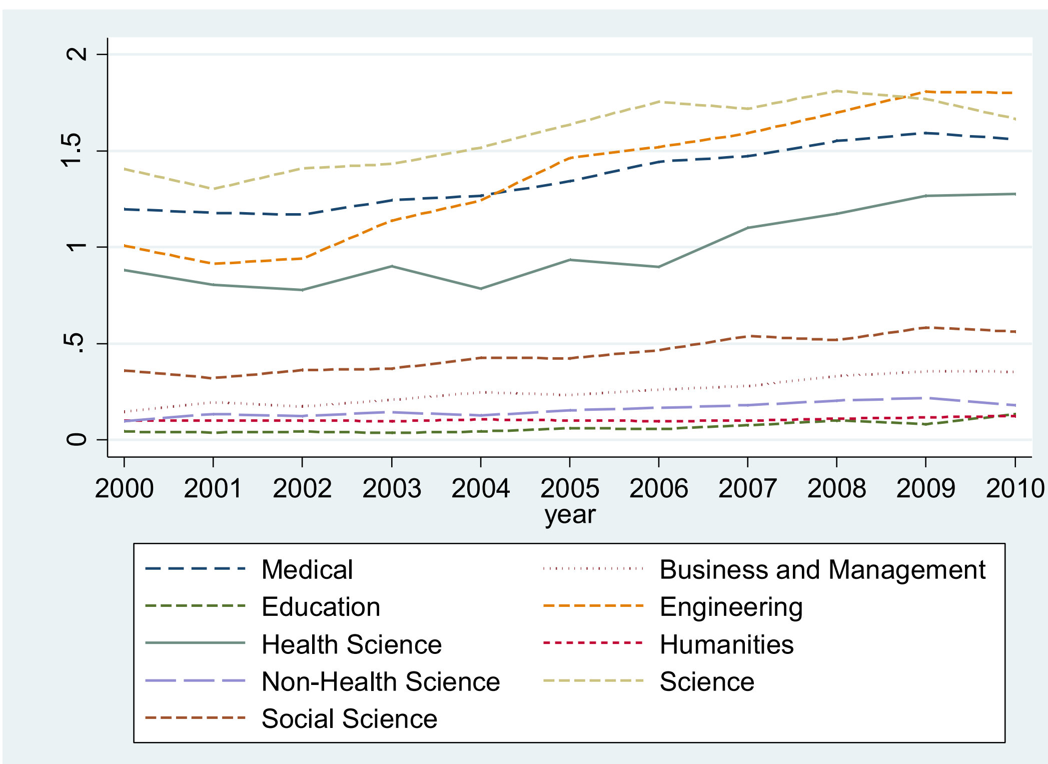

Figure 2 – Average number of articles published by the scientists in different fields

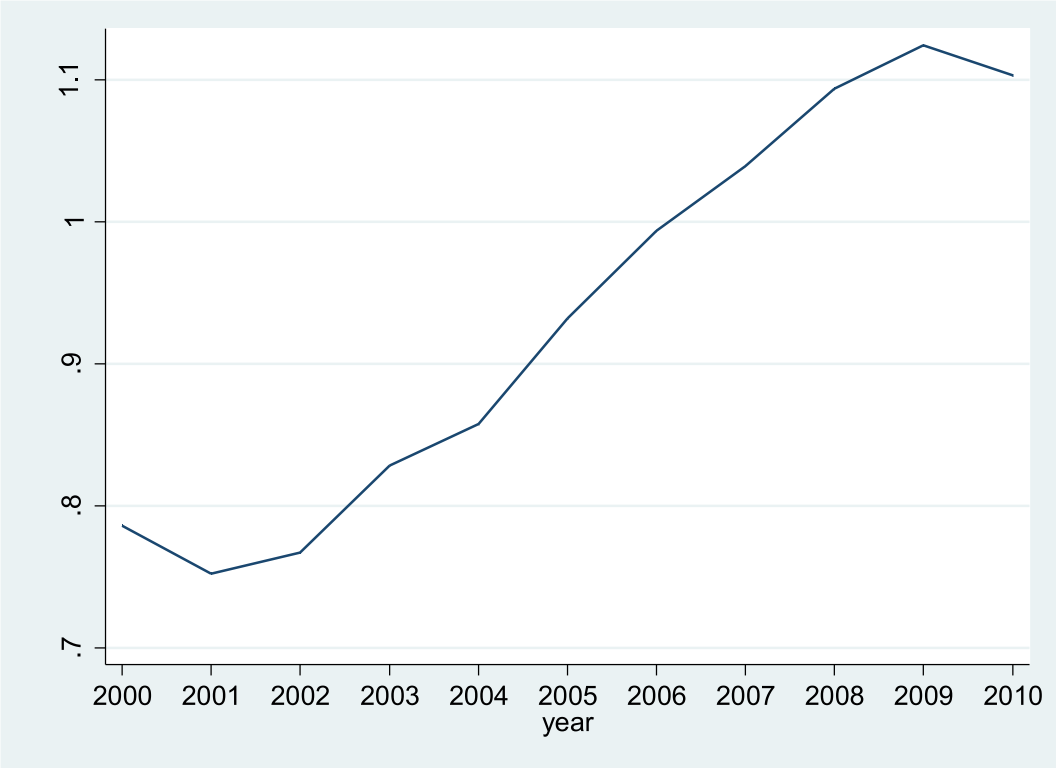

Figure 3 - Average number of articles per scientist in each year

The possible reason behind the observed yearly differences is that research capacity fluctuates on a yearly basis depending on the economy and research policy, and such fluctuations may in turn affect research output. University dummies and research field dummies play the same role as academic publication norms and standards, and as such research settings and related motivations are partially university and field dependent. The "Year 2000", "McGill University", and the "Medical science" field are selected as reference points and are thus omitted dummy variables. Considering the mentioned explanatory variables, the resulting model is given by:

$$\ln\left(nbArticle_{it}\right) = f\begin{pmatrix} \ln\left(PublicFundingO_{it}\right), \ln\left(PrivateFundingO_{it}\right), \ln\left(NFPFundingO_{it}\right), \\ ln\left(nbAuthor_{it}\right), ColTX_{it}, dFemale_{i}, Age_{it}, Age_{it}^{2}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

where *T* (in *ColTX*) is calculated either on the total amount of funding of the top-funded scientist, the total amount of public funding or the total amount of private funding, *X* (in *ColTX*) represents the top 10% or 5% most funded scientists, and the *D*s account for field, university and year dummy variables.

It is important to note that because the two variables [ln(*PublicfundingO*)] and [ln(*nbArticle*)] are determined by each other, which is the source of endogeneity, the ordinary least square (OLS) models are biased. The main reason for this potential endogeneity is that the scientists are assessed in their demands for public funding based on their CV and past effectiveness while at the same time, publication and research quality depend on the funding capability of researchers.

Using instrumental variables (IV) is a commonly suggested method to address endogeneity problems. There are two requirements for using IVs: (1) the instruments must be correlated with the endogenous variable, and (2) the instruments should not be correlated with the error term in the main regression equation. In other words, the instruments cannot suffer from the same endogeneity problem. If there is more than one instrument for the endogenous variable, it is necessary to perform a two-stage regression, in which the first stage estimates the endogenous variable (or instrumented variable) on a list of instrumental variables. Such estimation removes the error term of the first stage and keeps the estimated amount for the second stage. Neglecting the error term of the endogenous variable and inserting the estimated amount in the main regression equation is one way to remove the correlation of the RHS variables with the error term and to deal with the endogeneity problem.

A few variables act as instruments for the amount of public funding. The number of scientists in a university [*nbScientistUni*] can explain the allocation of money amongst scientists. We anticipate that a university with a higher number of scientists may be able to benefit from cost sharing of research expenditures, hence reducing the need for larger amounts of individual funding. The rank of a scientist in the field is another variable that can be used to explain research funds allocation. There is an established tradition in the literature to compare past research performance of scientists to allocate funds (Ho et al., 2006; Liefner, 2003).

The rank of previous funding can be another choice to predicting the future amount of funding. There is an echo effect for the amount of funding, which means that highly funded scientists are better able to get new sources of research money. The logic behind this argument is that decent research funds have effective networking capacity (Winter et al., 2006) to create different opportunities to get funds in a country such as Canada (Salazar and Holbrook, 2007). It should also be noted that the rank of funding is an ordinal variable and as such does not have information about amounts (the amount of research fund may indicate the capacity of knowledge production while the rank of funding does not provide such information). Therefore, it is an informative signal to estimate the amount of funding but it is not the funding itself. The rank of a scientist in the field in terms of three-year average funding for the purpose of operational costs and direct research expenditure [*PubORank*] is the second instrument used to predict the amount of funding. The third instrument is the total funding of each research cluster in each university [*totPublicfundingOcluster*].

In the first stage, the amount of public funding [ln(*PublicfundingO*)] is estimated by the instruments, aggregate amount of funding, the rank of scientist in terms of funding, and the number of university, and the variables of the second stage regression. To avoid simultaneity problems, public funding is not contemporaneous to the instruments; hence one-year lags of the instruments are used in the first-stage regression. Considering the mentioned explanatory variables, the resulting model is given by:

$$1^{st}\ stage: \ln\left(PublicFundingO_{it}\right) = g\left(PubORank_{it-1}, \ln(totPublicfundingOcluster_{it-1}), \ln\left(nbScientistUni_{it-1}\right)\right)$$

$$2^{nd}\ stage: \ln\left(nbArticle_{it}\right) = f\begin{pmatrix} \ln\left(PublicFundingO_{it}\right), \ln\left(PrivateFundingO_{it}\right), \ln\left(NFPFundingO_{it}\right), \\ \ln\left(nbAuthor_{it}\right), ColTX_{it}, dFemale_{i}, Age_{it}, Age_{it}^{2}, D_{Field}, D_{University}, D_{Year} \end{pmatrix}$$

In a well-specified model, the variables of the RHS (including the instrumental variables) should not be highly correlated with each other. A low correlation refers to a good level of independence and explanatory power of RHS variables. The correlation matrix is reported in appendix 3 and shows that the correlation coefficients are acceptable for our estimating regression equations.

4 Results and discussion

Because a number of our independent variables are individual fixed effects (gender, university affiliation for vast majority of scientists), we estimate random effect 2SLS regressions for panel data (with the *xtivreg* command in Stata). The second stage of the two-stage panel regressions for collaboration with top 5% and 10% (in terms of public funding, private funding, and funding from not-for-profit sector) are presented in tables 1 to 5. The first stage regressions are reported in appendices which show that instruments significantly explain the public funding endogenous variable.

Our results indicate that the coefficients of the amounts of funding from the public, private, and not-for-profit sectors are all significant and positive in all regressions: the more funded the scientists are, the more articles they publish. Having sufficient funds is a necessary condition for researchers to buy instruments and hire assistants in order to follow new ideas and conduct research at the frontier of knowledge. However, the scales of effect are different in funding sources. The difference can be explained by characteristics of each funding, already discussed in the literature. Geuna and Nesta (2003) claimed that increased industrial funding will force researchers to shift to more applied research, neglecting their normative responsibilities for knowledge development. Similarly, Partha and David (1994) argued that institutions and social norms of open science are functionally maximizing the long term growth of scientific knowledge, but are not powerful to be socially optimum in terms of producing economic rents and commercial outputs from the existing stock of scientific knowledge. Similarly, Kleinman and Vallas (2001) addressed a paradox called "the industrialization of the academy and the collegialization of industrial research" (pp. 451). In other words, scientists and engineers in private industry get a higher level of autonomy and control while scientists working in university settings are faced with the opposite case. In their empirical analysis, they address this paradox by arguing that norms, codes and practices in industry are penetrating the academy and vice versa. However, such infiltrating is asymmetric in favour of industrial norms.

Unlike private funding, public funding has different role. Pavitt (2001) referred to the importance of public support for scientific infrastructure development and highlights its role in the effectiveness of public grants in the US. In another study, Pavitt (2000) argues that equipment and networks are necessary for the development and implementation of research, and is thus costly. Providing required funding for conducting research via different science policies has an important role to push scientists toward the production of more effective and efficient research. Crespi and Geuna (2008) found that there is a delay for such policies to be effective. One well accepted justification for developing public support of science is that it is public good (Partha and David, 1994). Callon (1994) further argued that science is not a public good because of its intrinsic and natural properties, but a public good due to the fact that it is a source of diversity and flexibility.

In addition to the funding, our results show that the collaboration with top funded scientists is another significant determinant that fosters scientific production. There are different reasons for this effect. According to the papers assessing the effect of funding on scientific production (Beise and Stahl, 1999; Salter and Martin, 2001), well-funded scientists have a better performance. Collaboration with these scientists increases the chance of getting valuable experience and learning new skills through collaborative and collective research activities. The scientists who collaborated with well-funded scientists are more likely to publish a greater number of papers.

Economies of scale are the second possible reason behind our results. For instance, with a greater number of students that produce more articles, the per-unit fixed cost of shared resources (for example the cost of a laboratory technician) is lower, which then leads to a decreased marginal cost of publication. The same can be said of instruments and infrastructure. The third explanation stems from having access to modern instrumentation that more funds can provide. Neumann and Finaly-Neumann (1990) argued that research publication is directly influenced by support such as research instruments. Scientists who collaborate with top-funded scientists have access to better instruments and this would result in more scientific outputs.

Furthermore, the possible fourth reason is related to the fact that large amounts of funding are generally given to a team of researchers. It is important to note that SIRU accounts for interuniversity transfers. Hence a scientist who transfers funds to a colleague in another university will have his funds reduced by the transferred amount. The recipient will thus have an increased amount of funds as a consequence. It is however not possible to track funds from the same grants across universities as the reporting is not standardized across universities, i.e. we cannot reconstruct the teams a posteriori.

Regarding past networking, we can also argue that having higher a number of co-authors in the past significantly increases scientific production, hence supporting (Johnes, 1988; Melin, 1996). Turning to the other independent variables in the regression models, we conclude that female scientists have a significantly lower scientific production in all regressions. Long (1990) explained that women's opportunities for collaboration are significantly less than those of men's because women have young children. Kyvik and Teigen (1996) found that childcare and lack of research collaboration are the two factors which negatively affect the scientific publishing of women. In another study, Leahey (2006) argued that the reason of low production of women is that women specialize less than men, which is an important factor for research production. Beaudry and Larivière (2016) however showed that per dollar invested, women are generally equally productive as their male colleagues.

The age of a scientist also exhibits a significant inverted U-shaped effect with a peak at age 51. In other words, scientists have increasing scientific production until they reach the age of 51 but their performance drops afterward. This statement is compatible with earlier works that explain the age effect (Bernier et al., 1975; Diamond, 1986; Levin and Stephan, 1991). Our model also verifies the fixed effect of university and research discipline in addition to the year-specific effect on scientific production.

Tables 2 to 5 present the regressions where "collaboration with top funded scientists" is interacted with "other determinants of scientific production". Table 2 shows that having had a greater number of co-authors in the past increases scientific production, a phenomenon that is amplified if the researcher has collaborated with top-funded scientists. Table 3 indicates that although female scientists are less productive than their male colleagues in general, collaborating with top-funded scientists partially alleviates this negative effect, but only when they work with the top 5% scientists in terms of public funding. Table 4 and 5 also show the interactive effect of private funding and non-for-profit funding when a scientist has research collaboration with top-funded scientist. For private funding, there are only three significant interactive effects (collaboration with top-funded scientists at 10% level). The interesting point is that if a scientist has research collaboration with another scientist who is top-funded in terms of private funding, the effect of own private funding becomes negative. It would therefore seem that the duplication of private funding does not yield additional output, whereas there is a positive/reinforcing effect of public and of NFP funding. This is not surprising since the goal of private funding is rarely to produce scientific articles. The other two interactive effects are positive. For non-for-profit funding, the story is a bit different and interactive variables have positive and significant effect except for [*ColT95*ln(NFPfundingO)*] and [*ColPub95*ln(NFPfundingO)*], which are non-significant. We expect that there is not enough critical mass in the top 5% NFP-funded scientists to allow for measuring a significant effect.

Table 1 – Second stage regression results

Dependent Var: nbArticle	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$dFemale_i$	-0.0824 ***	-0.0827 ***	-0.0839 ***	-0.0836 ***	-0.0841 ***	-0.0841 ***
	0.0116	0.0115	0.0115	0.0115	0.0116	0.0117
Age_{it}	0.0328 ***	0.0331 ***	0.0318 ***	0.0325 ***	0.0339 ***	0.0343 ***
	0.0034	0.0034	0.0034	0.0034	0.0035	0.0035
Age_{it}^2	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\ln(nbAuthor_{it})$	0.1645 ***	0.1695 ***	0.1676 ***	0.1739 ***	0.1742 ***	0.1774 ***
	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066
$\ln(PublicfundingO_{it})$	0.0445 ***	0.0467 ***	0.0431 ***	0.0459 ***	0.0500 ***	0.0503 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
$\ln(PrivatefundingO_{it})$	0.0059 ***	0.0059 ***	0.0065 ***	0.0066 ***	0.0025 ***	0.0043 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$\ln(NFPfundingO_{it})$	0.0045 ***	0.0048 ***	0.0052 ***	0.0053 ***	0.0054 ***	0.0056 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$ColT90_{it}$	0.1802 ***					
	0.0058					
$ColT95_{it}$		0.1819 ***				
		0.0062				
$ColPub90_{it}$			0.1760 ***			
			0.0056			
$ColPub95_{it}$				0.1723 ***		
				0.0061		
$ColPriv90_{it}$					0.1725 ***	
					0.0060	
$ColPriv95_{it}$						0.1641 ***
						0.0067
Constant	-0.2738 ***	-0.2828 ***	-0.2380 ***	-0.2662 ***	-0.3384 ***	-0.3387 ***
	0.0896	0.0896	0.0894	0.0894	0.0901	0.0903
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5389	5268	5428	5258	5104	4871
sigma	0.4820	0.4814	0.4809	0.4799	0.4836	0.4846
rho	0.4078	0.4034	0.4052	0.3982	0.4080	0.4065
R^2 ***within groups***	0.0504	0.0462	0.0508	0.0443	0.0442	0.0384
R^2 ***overall***	0.2667	0.2632	0.2688	0.2642	0.2569	0.2520
R^2 ***between groups***	0.4423	0.4397	0.4454	0.4439	0.4323	0.4283

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

Table 2 – Second stage regression results with considering joint effect of research collaboration and number of articles

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
dFemale$_i$	-0.0821 ***	-0.0826 ***	-0.0837 ***	-0.0835 ***	-0.0838 ***	-0.0841 ***
	0.0116	0.0116	0.0115	0.0115	0.0116	0.0117
Age$_{it}$	0.0327 ***	0.0332 ***	0.0317 ***	0.0325 ***	0.0337 ***	0.0343 ***
	0.0034	0.0034	0.0034	0.0034	0.0035	0.0035
Age$_{it}^2$	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ln(*nbAuthor$_{it}$*)	0.1523 ***	0.1645 ***	0.1513 ***	0.1647 ***	0.1607 ***	0.1769 ***
	0.0078	0.0073	0.0076	0.0072	0.0074	0.0071
ln(*PublicfundingO$_{it}$*)	0.0447 ***	0.0468 ***	0.0433 ***	0.0460 ***	0.0503 ***	0.0503 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
ln(*PrivatefundingO$_{it}$*)	0.0059 ***	0.0058 ***	0.0065 ***	0.0066 ***	0.0026 ***	0.0043 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ln(*NFPfundingO$_{it}$*)	0.0045 ***	0.0048 ***	0.0052 ***	0.0053 ***	0.0054 ***	0.0056 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ColT90$_{it}$	0.1312 ***					
	0.0175					
ColT95$_{it}$		0.1525 ***				
		0.0199				
ColPub90$_{it}$			0.1080 ***			
			0.0171			
ColPub95$_{it}$				0.1154 ***		
				0.0194		
ColPriv90$_{it}$					0.0976 ***	
					0.0194	
ColPriv95$_{it}$						0.1602 ***
						0.0229
ColT90$_{it}$*ln(nbArticle$_{it}$)	0.0293 ***					
	0.0099					
ColT95$_{it}$*ln(nbArticle$_{it}$)		0.0167				
		0.0108				
ColPub90$_{it}$*ln(nbArticle$_{it}$)			0.0407 ***			
			0.0097			
ColPub95$_{it}$*ln(nbArticle$_{it}$)				0.0324 ***		
				0.0105		
ColPriv90$_{it}$*ln(nbArticle$_{it}$)					0.0441 ***	
					0.0109	
ColPriv95$_{it}$*ln(nbArticle$_{it}$)						0.0022
						0.0123
ColT90$_{it}$*dFemale$_i$						
Constant	-0.2564 ***	-0.2783 ***	-0.2142 ***	-0.2543 ***	-0.3188 ***	-0.3382 ***
	0.0898	0.0897	0.0896	0.0896	0.0902	0.0904
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5399	5262	5455	5264	5123	4869
sigma	0.4820	0.4818	0.4808	0.4801	0.4834	0.4847
rho	0.4083	0.4044	0.4053	0.3989	0.4082	0.4068
R^2 within groups	0.0507	0.0463	0.0514	0.0446	0.0448	0.0384
R^2 overall	0.2663	0.2630	0.2686	0.2641	0.2567	0.2519
R^2 between groups	0.4420	0.4393	0.4454	0.4436	0.4316	0.4282

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

Table 3 – Second stage regression results with considering joint effect of research collaboration and scientist's gender

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
dFemale$_i$	-0.0915 ***	-0.0909 ***	-0.0817 ***	-0.0898 ***	-0.0904 ***	-0.0912 ***
	0.0123	0.0119	0.0122	0.0118	0.0121	0.0119
Age$_{it}$	0.0328 ***	0.0331 ***	0.0317 ***	0.0325 ***	0.0340 ***	0.0344 ***
	0.0034	0.0034	0.0034	0.0034	0.0035	0.0035
Age$_{it}^2$	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
ln(*nbAuthor$_{it}$*)	0.1648 ***	0.1697 ***	0.1679 ***	0.1741 ***	0.1742 ***	0.1773 ***
	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066
ln(*PublicfundingO$_{it}$*)	0.0446 ***	0.0468 ***	0.0431 ***	0.0460 ***	0.0501 ***	0.0503 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
ln(*PrivatefundingO$_{it}$*)	0.0059 ***	0.0059 ***	0.0065 ***	0.0066 ***	0.0025 ***	0.0043 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ln(*NFPfundingO$_{it}$*)	0.0045 ***	0.0048 ***	0.0052 ***	0.0053 ***	0.0054 ***	0.0056 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
ColT90$_{it}$	0.1734 ***					
	0.0065					
ColT95$_{it}$		0.1727 ***				
		0.0070				
ColPub90$_{it}$			0.1778 ***			
			0.0064			
ColPub95$_{it}$				0.1653 ***		
				0.0069		
ColPriv90$_{it}$					0.1668 ***	
					0.0067	
ColPriv95$_{it}$						0.1542 ***
						0.0075
ColT90$_{it}$*dFemale$_i$	0.0287 ***					
	0.0129					
ColT95$_{it}$*dFemale$_i$		0.0403 ***				
		0.0142				
ColPub90$_{it}$*dFemale$_i$			-0.0070			
			0.0127			
ColPub95$_{it}$*dFemale$_i$				0.0304 **		
				0.0141		
ColPriv90$_{it}$*dFemale$_i$					0.0257 *	
					0.0137	
ColPriv95$_{it}$*dFemale$_i$						0.0460 ***
						0.0157
Constant	-0.2735 ***	-0.2816 ***	-0.2372 ***	-0.2656 ***	-0.3394 ***	-0.3403 ***
	0.0895	0.0896	0.0894	0.0894	0.0901	0.0903
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5401	5285	5440	5274	5109	4884
sigma	0.4815	0.4809	0.4804	0.4793	0.4835	0.4844
rho	0.4070	0.4025	0.4039	0.3970	0.4079	0.4063
R^2 within groups	0.0506	0.0466	0.0507	0.0445	0.0444	0.0388
R^2 overall	0.2662	0.2626	0.2690	0.2638	0.2567	0.2516
R^2 between groups	0.4414	0.4386	0.4457	0.4430	0.4319	0.4276

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

Table 4 – Second stage regression results with considering joint effect of research collaboration and private funding

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
$dFemale_i$	-0.0824 ***	-0.0827 ***	-0.0841 ***	-0.0836 ***	-0.0844 ***	-0.0842 ***
	0.0115	0.0115	0.0114	0.0114	0.0116	0.0117
Age_{it}	0.0326 ***	0.0330 ***	0.0315 ***	0.0323 ***	0.0339 ***	0.0343 ***
	0.0034	0.0034	0.0034	0.0034	0.0035	0.0035
Age_{it}^2	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\ln(nbAuthor_{it})$	0.1655 ***	0.1699 ***	0.1695 ***	0.1746 ***	0.1734 ***	0.1771 ***
	0.0066	0.0066	0.0066	0.0066	0.0066	0.0067
$\ln(PublicfundingO_{it})$	0.0446 ***	0.0468 ***	0.0432 ***	0.0459 ***	0.0500 ***	0.0503 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
$\ln(PrivatefundingO_{it})$	0.0048 ***	0.0059 ***	0.0046 ***	0.0063 ***	0.0045 ***	0.0048 ***
	0.0009	0.0008	0.0009	0.0008	0.0009	0.0008
$\ln(NFPfundingO_{it})$	0.0045 ***	0.0048 ***	0.0052 ***	0.0053 ***	0.0054 ***	0.0056 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$ColT90_{it}$	0.1725 ***					
	0.0070					
$ColT95_{it}$		0.1819 ***				
		0.0078				
$ColPub90_{it}$			0.1618 ***			
			0.0069			
$ColPub95_{it}$				0.1685 ***		
				0.0076		
$ColPriv90_{it}$					0.1907 ***	
					0.0077	
$ColPriv95_{it}$						0.1740 ***
						0.0093
$ColT90_{it}*\ln(PrivatefundingO_{it})$	0.0022 **					
	0.0011					
$ColT95_{it}*\ln(PrivatefundingO_{it})$		0.0000				
		0.0011				
$ColPub90_{it}*\ln(PrivatefundingO_{it})$			0.0039 ***			
			0.0011			
$ColPub95_{it}*\ln(PrivatefundingO_{it})$				0.0010		
				0.0011		
$ColPriv90_{it}*\ln(PrivatefundingO_{it})$					-0.0043 ***	
					0.0012	
$ColPriv95_{it}*\ln(PrivatefundingO_{it})$						-0.0019
						0.0012
Constant	-0.2693 ***	-0.2811 ***	-0.2284 ***	-0.2624 ***	-0.3422 ***	-0.3397 ***
	0.0894	0.0895	0.0891	0.0893	0.0901	0.0903
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5428	5290	5506	5291	5118	4872
sigma	0.4805	0.4805	0.4783	0.4786	0.4834	0.4846
rho	0.4042	0.4010	0.3987	0.3948	0.4080	0.4066
R^2 *within groups*	0.0503	0.0462	0.0508	0.0442	0.0447	0.0385
R^2 *overall*	0.2674	0.2633	0.2704	0.2646	0.2567	0.2519
R^2 *between groups*	0.4435	0.4398	0.4478	0.4444	0.4317	0.4281

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

Table 5 – Second stage regression results with considering joint effect of research collaboration and NFP funding

Dependent Var: nbArticle	(1)	(2)	(3)	(4)	(5)	(6)
$dFemale_i$	-0.0823 ***	-0.0827 ***	-0.0839 ***	-0.0836 ***	-0.0843 ***	-0.0843 ***
	0.0115	0.0115	0.0114	0.0114	0.0116	0.0116
Age_{it}	0.0325 ***	0.0329 ***	0.0315 ***	0.0322 ***	0.0337 ***	0.0342 ***
	0.0034	0.0034	0.0034	0.0034	0.0034	0.0035
Age_{it}^2	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***	-0.0003 ***
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\ln(nbAuthor_{it})$	0.1655 ***	0.1702 ***	0.1691 ***	0.1749 ***	0.1753 ***	0.1783 ***
	0.0066	0.0066	0.0066	0.0066	0.0066	0.0066
$\ln(PublicfundingO_{it})$	0.0446 ***	0.0468 ***	0.0432 ***	0.0460 ***	0.0502 ***	0.0504 ***
	0.0020	0.0020	0.0020	0.0020	0.0020	0.0020
$\ln(PrivatefundingO_{it})$	0.0059 ***	0.0059 ***	0.0065 ***	0.0066 ***	0.0024 ***	0.0043 ***
	0.0007	0.0007	0.0007	0.0007	0.0007	0.0007
$\ln(NFPfundingO_{it})$	0.0035 ***	0.0045 ***	0.0039 ***	0.0048 ***	0.0038 ***	0.0048 ***
	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007
$ColT90_{it}$	0.1720 ***					
	0.0072					
$ColT95_{it}$		0.1769 ***				
		0.0081				
$ColPub90_{it}$			0.1645 ***			
			0.0071			
$ColPub95_{it}$				0.1655 ***		
				0.0080		
$ColPriv90_{it}$					0.1546 ***	
					0.0076	
$ColPriv95_{it}$						0.1480 ***
						0.0089
$ColT90_{it}*\ln(NFPfundingO_{it})$	0.0020 **					
	0.0011					
$ColT95_{it}*\ln(NFPfundingO_{it})$		0.0011				
		0.0011				
$ColPub90_{it}*\ln(NFPfundingO_{it})$			0.0029 ***			
			0.0010			
$ColPub95_{it}*\ln(NFPfundingO_{it})$				0.0015		
				0.0011		
$ColPriv90_{it}*\ln(NFPfundingO_{it})$					0.0041 ***	
					0.0011	
$ColPriv95_{it}*\ln(NFPfundingO_{it})$						0.0033 ***
						0.0012
Constant	-0.2680 ***	-0.2793 ***	-0.2308 ***	-0.2617 ***	-0.3325 ***	-0.3359 ***
	0.0894	0.0895	0.0891	0.0892	0.0899	0.0902
Number of observations	32299	32299	32299	32299	32299	32299
Number of scientists	5602	5602	5602	5602	5602	5602
$\chi 2$	5428	5295	5492	5302	5156	4906
sigma	0.4805	0.4804	0.4786	0.4782	0.4820	0.4836
rho	0.4042	0.4007	0.3994	0.3939	0.4043	0.4039
R^2 within groups	0.0503	0.0461	0.0507	0.0442	0.0443	0.0384
R^2 overall	0.2673	0.2636	0.2698	0.2647	0.2580	0.2527
R^2 between groups	0.4436	0.4403	0.4472	0.4448	0.4341	0.4296

Notes: *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively; Year dummies, field dummies, and university dummies are significant. The minimum year activity, average year activity, and maximum year activity are 1, 6.13, and 12 respectively.

5 Conclusion

This article developed a theoretical model to show that collaboration with top-funded scientists has a positive effect on the number of articles published by a scientist. The empirical results from a two-stage panel regression model verify the results predicted by the theoretical model. Using a rich dataset on publication and funding of Quebec scientists, the paper shows that collaborating with top-funded scientists has indeed a positive effect on the production of a scientist.

The results also corroborate the evidence found in the literature regarding the positive effect of funding, the positive effect of past networking (measured by number of co-authors), the inverted U-shaped effect of age, and the fewer number of publications by women compared to that of men. In addition, the fixed effects in the models, university, research field, and year, all show significant effects. This implies that each university, each research field, or each year has an unmeasurable or hidden set of scientific, social, and financial characteristics affecting scientific production.

The main policy advice considering the significant and positive effect of collaboration with top-funded scientists on scientific production, consists in maintaining incentives (or giving a mandate) for top-funded scientists to continue their collaborations with lower-funded scientists, as the entire science system benefits from them. This kind of collaboration may positively affect academic production through different channels. First, it may result in a substantial transfer of tacit knowledge and yield more scientific publications. Second, it provides benefits such as economies of scale in knowledge production because well-funded scientists have generally larger teams of researchers and better research equipment. Third, the expansion of the research network is another viewed benefit of such collaboration. Fourth, collaboration with top-funded scientists also has a moderating effect on the relationship of other determinants with scientific production, which suggests an amplified positive impact if a researcher has a greater number of co-authors, has raised a larger amount of public funding, and is female. For the specific moderating effect with private finding, there are only few significant interactive effects. For non-for-profit funding, the coefficients of the interactive variables are positive and significant for most measures. Unfortunately, these interactive effects do not follow a consistent enough pattern and significance to be able to make a general statement.

As this study focuses on the province of Quebec in Canada, country specific policy recommendation are also derived. Funding programs in Canada such as industrial chairs funded by established firms, chairs appointed by federal granting councils (NSERC, and CIHR), and chairs supported by provincial granting bodies (Fonds de recherche du Québec) that aim to develop collaboration and capability building in academia, were reviewed in Mirnezami and Beaudry (2015). The main contribution of these 'chair' programs is the research grants they provide in addition to the title of 'chair holder'. Such funding can be counted as part of the normal research budget of researchers, thus appointees and chair holders can most probably be deemed top funded scientists. Hence such programs may have a double effect not only as an evidence for the subject of this study (collaboration with top funded scientist) but also on academic networking with industry for instance. To frame this interpretation and translate it into a recommendation, we argue that it is possible to define, and to some extent further enhance, the networking role of top researchers by prestigious nominations and to support them with greater amounts of funding. By the same token, this may increase the effect of collaboration with top funded scientists, i.e. by improving the efficiency of money spent in these programs. Other scientists may also have more willingness to conduct collaborative research with 'title' holders like research chairs, mainly because collaborating with them provides a greater visibility for teir research and facilitated access to talented students, researchers and teams.

The generality of the abovementioned interpretations and recommendations may be affected by number of limitations in this study. First, this study only covers Quebec scientists. Second, the database time coverage is limited and some variables like professional experience, quality of education, initial condition of researchers, and institutional support are not available for our study. Third, some data entries are missing in the original dataset and we do not have the funding information of all collaborators, only of Quebec collaborators, i.e. the impact or influence of Canadian outside of Quebec or international stars cannot be measured. In addition to addressing these shortcomings in future studies with more complete data entries and records, we propose suggestions for future work on this subject.

First, this paper defines the 'star' scientists based on the amount of funding, i.e. it identifies the top-funded scientists. As the funding of a scientist is determined by previous performance and intrinsic characteristics, it is an appropriate proxy for research quality. However, future research should consider and compare different definitions of 'star scientist' to provide a more comprehensive and robust interpretation. For instance, star scientists can be identified based on the number of patents, the mean normalised citation scores, the impact factor of the journals in which the researchers publish, and networking measure such as centrality or cliquishness indicators. The second suggestion for future studies is about different research methods. A deep investigation on the nature of research collaboration could help build a taxonomy to explain the benefit and knowledge spillover of different types of research collaboration with star scientists. Surveys and qualitative research would thus provide a complementary detailed understanding of the nature of the collaboration with star scientists. The final suggestion to investigate 'how' this effect happens and to identify channels for the effect of collaboration with top funded scientists on scientific production.

Appendix 1 - Variable description

Variable name	*Variable description*
$ColT90_{it}$ and $ColT95_{it}$	Dummy variables taking the value 1 if any of the coauthors of researcher *i* is amongst top funded scientists (top 10% and top 5% of total funding respectively)
$ColPub90_{it}$ and $ColPub95_{it}$	Dummy variables taking the value 1 if any of the coauthors of researcher *i* is amongst top funded scientists (top 10% and top 5% of public funding respectively)
$ColPriv90_{it}$, and $ColPriv95_{it}$	Dummy variables taking the value 1 if any of the coauthors of researcher *i* is amongst top funded scientists (top 10% and top 5% of private funding respectively)
ln(*$PublicfundingO_{it}$*)	Natural logarithm of the three-year average up to year *t* of public sector funding for the purpose of operational costs and direct expenditures of research of researcher *i*
ln(*$PrivatefundingO_{it}$*)	Natural logarithm of the three-year average up to year *t* of private sector funding for the purpose of operational costs and direct expenditures of research of researcher *i*
ln(*$NFPfundingO_{it}$*)	Natural logarithm of three-year average up to year *t* of funding from not-for-profit institutions (NFP) for the purpose of operational costs and direct expenditures of research of researcher *i*
ln(*$nbArticle_{it}$*)	Natural logarithm of the yearly number of articles published in year *t* by researcher *i*
ln(*$nbAuthor_{it}$*)	Natural logarithm of the three-year average of number of authors in the papers of researcher *i*
$PubORank_{it}$	Normalized rank of researcher *i* in the field in terms of three-year average of funding for the purpose of operational costs and direct expenditure of research
$nbScientistUni_{it}$	Number of scientists in the university and division of researcher *i*
ln($totPublicfundingOcluster_{it}$)	Natural logarithm of the three-year average up to year *t* of public sector funding for the purpose of operational costs and direct expenditures of research, which is aggregated over cluster of researcher *i*
Age_{it}	Age of researcher *i* in year *t*
$dFemale_i$	Dummy variable taking the value 1 if researcher *i* is a woman and 0 otherwise
$dULaval_i$, $dUMcGill_i$, …, $dUdeM_i$	Dummy variables indicating the university affiliation of researcher *i*
$dMedical_i$, $dHumanities_i$, …, $dScience_i$	Dummy variables indicating the field of researcher *i*
d2000, d2001, …., d2012	Dummy variables indicating the year

-

- Some variables are transformed by natural logarithm function to be normal variables and satisfy the necessary conditions for the right hand side variables of the regression equations.
- There are 9 divisions of Basic Medical Sciences, Business & Management, Education, Engineering, Health Sciences, Humanities, Non-Health Professional, Sciences, and Social Sciences
- Cluster is more detail categorization of researchers. There are 42 clusters of Agricultural & Food Sciences, Anthropology, Archaeology & Sociology, Biology & Botany, Business, Chemical Engineering, Chemistry, Civil Engineering, Computer & Information Science, Dentistry, Earth & Ocean Sciences, Economics, Education, Electrical & Computer Engineering, Fine & Performing Arts, Foreign Languages Literature, Linguistic, French/English, General Medicine, Geography, History, Kinesiology / Physical Education, Laboratory Medicine, Law & Legal Studies, Library & Information Sciences, Mathematics, Mechanical & Industrial Engineering, Media & Communication Studies, Medical Specialties, Nursing, Other Engineering, Other Health Sciences, Other Social Sciences & Humanities, Philosophy, Physics & Astronomy, Planning & Architecture, Political Science, Psychology, Public Health & Health Administration, Rehabilitation Therapy, Religious Studies & Vocations, Resource Management & Forestry, Social Work, and Surgical Specialties.

Appendix 2- Summary statistics, (No. observation: 32,299) - the variables are not summarized in logarithmic scale and they are raw numbers

	Mean	*Standard Deviation*	*Median*	*Minimum*	*Maximum*
nbArticle	3.3880	3.6327	2	1	85
PublicfundingO	113,648	200,875	63491.6	0	10,100,000
PrivatefundingO	21,609	96,627	0	0	5,508,461
NFPfundingO	21,197	126,788	0	0	8,404,625
dFemale	0.2348	0.4239	0	0	1
Age	50.482	9.209	50.000	16	92
nbAuthor	6.0821	34.6297	3.8556	0.3333	2063.646
PubORank	0.6704	0.2314	0.7090	0.0010	1
totPublicfundingOcluster	31,600,000	32,000,000	16,500,000	73615	121,000,000
nbScientistUni	452.84	446.59	300	5	1604
ColT90	0.4113	0.4921	0	0	1
ColT95	0.2766	0.4473	0	0	1
ColPub90	0.4113	0.4921	0	0	1
ColPub95	0.2712	0.4446	0	0	1
ColPriv90	0.3331	0.4713	0	0	1
ColPriv95	0.2165	0.4119	0	0	1

In some subfields of Physics, there are many scientists involved in one project and therefore, the maximum for the number of authors is high.

References

Abramo, G., D'Angelo, C.A., Caprasecca, A., 2009. Gender differences in research productivity: A bibliometric analysis of the Italian academic system. Scientometrics 79, 517-539.
Azoulay, P., Ding, W., Stuart, T., 2007. The determinants of faculty patenting behavior: Demographics or opportunities? Journal of Economic Behavior & Organization 63, 599-623.
Bagchi-Sen, S., 2007. Strategic considerations for innovation and commercialization in the US biotechnology sector. European Planning Studies 15, 753-766.
Beaudry, C., Schiffauerova, A., 2011. Impacts of collaboration and network indicators on patent quality: The case of Canadian nanotechnology innovation. European Management Journal 29, 362-376.
Beise, M., Stahl, H., 1999. Public research and industrial innovations in Germany. Research Policy 28, 397-422.
Bercovitz, J., Feldman, M., 2011. The mechanisms of collaboration in inventive teams: composition, social networks, and geography. Research Policy 40, 81-93.
Berman, E.M., 1990. The economic impact of industry-funded university R&D. Research Policy 19, 349-355.
Bernier, C., Gill, W., Hunt, R., 1975. Measures of excellence of engineering and science departments: a chemical engineering example. Chemical Engineering Education 9, 194-202.
Blackburn, R.T., Behymer, C.E., Hall, D.E., 1978. Research note: Correlates of faculty publications. Sociology of Education, 132-141.
Callon, M., 1994. Is Science a Public Good? Fifth Mullins Lecture, Virginia Polytechnic Institute, 23 March 1993. Science, Technology & Human Values 19, 395-424.
Chwe, M.S.Y., 2000. Communication and coordination in social networks. Review of economic studies 67, 1-16.
Crane, D., 1965. Scientists at major and minor universities: A study of productivity and recognition. American Sociological Review, 699-714.
Crespi, G.A., Geuna, A., 2008. An empirical study of scientific production: A cross country analysis, 1981–2002. Research Policy 37, 565-579.
Darby, M.R., Zucker, L.G., 1996. Star scientists, institutions, and the entry of Japanese biotechnology enterprises. National bureau of economic research.
Diamond, A.M., 1986. The life-cycle research productivity of mathematicians and scientists. Journal of Gerontology 41, 520.
Durnev, A., Kim, E., 2005. To steal or not to steal: Firm attributes, legal environment, and valuation. The Journal of Finance 60, 1461-1493.
Feldman, M.P., Graddy-Reed, A., 2013. Accelerating commercialization: a new model of strategic foundation funding. The Journal of Technology Transfer, 1-21.
Fox, M.F., 2005. Gender, family characteristics, and publication productivity among scientists. Social Studies of Science 35, 131-150.
Geuna, A., 2001. The changing rationale for European university research funding: are there negative unintended consequences? Journal of Economic Issues 35, 607-632.
Geuna, A., Nesta, L., 2003. University patenting and its effects on academic research. SEWPS Paper.
Godin, B., Gingras, Y., 2000. Impact of collaborative research on academic science. Science and Public Policy 27, 65-73.
Goldfarb, B., 2008. The effect of government contracting on academic research: Does the source of funding affect scientific output? Research Policy 37, 41-58.
Gulbrandsen, M., Smeby, J.C., 2005. Industry funding and university professors' research performance. Research Policy 34, 932-950.
Hicks, D., 1995. Published papers, tacit competencies and corporate management of the public/private character of knowledge. Industrial and corporate change 4, 401-424.

Higgins, M.J., Stephan, P.E., Thursby, J.G., 2011. Conveying quality and value in emerging industries: Star scientists and the role of signals in biotechnology. Research Policy 40, 605-617.
Ho, W., Dey, P.K., Higson, H.E., 2006. Multiple criteria decision-making techniques in higher education. International journal of educational management 20, 319-337.
Hottenrott, H., Thorwarth, S., 2011. Industry funding of university research and scientific productivity. Kyklos 64, 534-555.
Jensen, M.C., Meckling, W.H., 1976. Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of financial economics 3, 305-360.
Johnes, G., 1988. Determinants of research output in economics departments in British universities. Research Policy 17, 171-178.
Kleinman, D.L., Vallas, S.P., 2001. Science, capitalism, and the rise of the "knowledge worker": The changing structure of knowledge production in the United States. Theory and Society 30, 451-492.
Kyvik, S., 1990. Age and scientific productivity. Differences between fields of learning. Higher Education 19, 37-55.
Kyvik, S., Olsen, T.B., 2008. Does the aging of tenured academic staff affect the research performance of universities? Scientometrics 76, 439-455.
Kyvik, S., Teigen, M., 1996. Child care, research collaboration, and gender differences in scientific productivity. Science, Technology & Human Values 21, 54.
Leahey, E., 2006. Gender differences in productivity. Gender & Society 20, 754-780.
Levin, S.G., Stephan, P.E., 1991. Research productivity over the life cycle: evidence for academic scientists. The American Economic Review, 114-132.
Liefner, I., 2003. Funding, resource allocation, and performance in higher education systems. Higher Education 46, 469-489.
Long, J.S., 1990. The origins of sex differences in science. Social Forces 68, 1297-1316.
Manjarrés-Henríquez, L., Gutiérrez-Gracia, A., Carrión-García, A., Vega-Jurado, J., 2009. The effects of university–industry relationships and academic research on scientific performance: Synergy or substitution? Research in Higher Education 50, 795-811.
Melin, G., 1996. The networking university. Scientometrics 35, 15-31.
Meyer, M., 2006. Are patenting scientists the better scholars?: An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology. Research Policy 35, 1646-1662.
Mirnezami, S.R., Beaudry, C., 2015. The Effect of Having a Research Chair on Scientists' Productivity, The 15th International Society of Scientometrics and Informetrics Conference, Istanbul, Turkey.
Nakhaie, M.R., 2002. Gender Differences in Publication among University Professors in Canada*. Canadian Review of Sociology/Revue canadienne de sociologie 39, 151-179.
Neumann, Y., Finaly-Neumann, E., 1990. The support-stress paradigm and faculty research publication. The Journal of Higher Education, 565-580.
Partha, D., David, P.A., 1994. Toward a new economics of science. Research Policy 23, 487-521.
Pavitt, K., 2000. Why European Union funding of academic research should be increased: a radical proposal. Science and Public Policy 27, 455-460.
Pavitt, K., 2001. Public policies to support basic research: What can the rest of the world learn from US theory and practice?(And what they should not learn). Industrial and corporate change 10, 761-779.
Rothaermel, F.T., Hess, A.M., 2007. Building dynamic capabilities: Innovation driven by individual-, firm-, and network-level effects. Organization Science 18, 898-921.
Salazar, M., Holbrook, A., 2007. Canadian science, technology and innovation policy: the product of regional networking? Regional studies 41, 1129-1141.
Salter, A.J., Martin, B.R., 2001. The economic benefits of publicly funded basic research: a critical review. Research Policy 30, 509-532.
Winter, M., Smith, C., Morris, P., Cicmil, S., 2006. Directions for future research in project management: The main findings of a UK government-funded research network. International Journal of Project Management 24, 638-649.

Xie, Y., Shauman, K.A., 1998. Sex differences in research productivity: New evidence about an old puzzle. American Sociological Review, 847-870.
Zhou, J., 2003. When the presence of creative coworkers is related to creativity: role of supervisor close monitoring, developmental feedback, and creative personality. Journal of Applied Psychology 88, 413.
Zucker, L.G., Darby, M.R., 1996. Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. Proceedings of the National Academy of Sciences 93, 12709-12716.
Zucker, L.G., Darby, M.R., 1997. Individual Action and the Demand for Institutions Star Scientists and Institutional Transformation. American Behavioral Scientist 40, 502-513.
Zucker, L.G., Darby, M.R., 2001. Capturing technological opportunity via Japan's star scientists: Evidence from Japanese firms' biotech patents and products. The Journal of Technology Transfer 26, 37-58.
Zucker, L.G., Darby, M.R., Armstrong, J.S., 2002. Commercializing Knowledge: University Science, Knowledge Capture, and Firm Performance in Biotechnology. Management Science 48, 138-153.