# What determines researchers' scientific impact?

# A case study of Quebec researchers

Seyed Reza Mirnezami
Catherine Beaudry
Polytechnique Montreal, P.O. Box. 6079, Montreal, QC, H3C 3A7, Canada

Vincent Larivière
University of Montreal, P.O. Box. 6128, Montreal, QC, H3C 3J7 Canada

## Keywords:

Bibliometrics, Science Policy, Citations, Scientific Publications, Scientific Impact

## Abstract

Using a data set integrating information about researchers' funding and publication in the province of Quebec (Canada), this paper intends to identify the main determinants of citation counts as one measure of research impact. Using two-stage least square regressions to control for endogeneity, the results confirm the significant and positive relationship between the number of articles and citation counts. Our results also show that scientists with more articles in higher impact factor journals generally receive more citations and so do scientists who publish with a larger team of authors. Hence the greater visibility provided by a more prolific scientific production, better journals, and more co-authors, all contribute to increasing the perceived impact of articles. The paper also shows that male and female receive the same number of citations, all else being equal. In most of domains, the amount of funding does not have a significant effect on the citation counts. These results suggest that the most important

determinants of researchers' citations are the journals in which they publish, as well the collaborative nature of their research.

## 1 Introduction

Research impact and its determinants are an important topic in science policy, as governments and public authorities seek to maximize the benefit of public spending on knowledge production and science advancement. The main condition for having a strong impact of public research on industry or making a long-lasting effect on basic research is to generate research of high impact. Over the last decades, a vast body of literature has used citations to assess the scientific impact of scholarly research (Adam 2002; Brown and Gardner 1985; Kostoff 2002; Narin 1976). Although the classical interpretation of citations is that citations are building blocks in the construction of knowledge (Moed, 2005), many other factors have been shown to influence citation counts.

Assuming that citations can be a proxy for research impact (Kostoff 1998; Moed 2006; Phelan 1999), it becomes important from a science policy perspective, to better understand the various factors that affect researchers' citation counts. The literature (see next second) has highlighted several of the factors. The size of research teams as an indicator of collaboration can influence research impact (Johnes 1988; Melin 1996). Similarly, previous findings shows that citation numbers depends on the domain, the prestige of the journal, and the social network of authors (Bornmann et al. 2008). Gender is also highlighted as a factor affecting the number of citations (Aksnes et al. 2011). The amount of research funding can be another determinant as it provides new opportunities for research and new sets of scientific findings (Harman 2000; Pavitt 2000, 2001), which could probably become an academic heritage.

We investigate the determinants of citations by analysing information about researchers funding and publication in the province of Quebec[1]. The first contribution of this paper is to analyze whether the various factors found in the literature are also positively associated, in our dataset, with changes in citation counts. Because those factors may be related with each other, the second contribution consists in collectively considering all of the mentioned factors in one model to test if their specific effects remain statistically significant. More specifically, our general research question is as follows: do the mentioned factors (number of articles, amount of funding, size of research team, gender, and research filed) significantly affect citation counts? This comprehensive consideration of different factors in one study allows for the testing of co-existing effect of different determinants and, as such, consists of an original contribution to the literature on citations.

The remainder of the paper goes as follows: Section 2 presents the conceptual framework and literature review; Section 3 explains the data set used in the paper and describes the research methodology; Section 4 analyses the regression results; and finally, Section 5 discusses the results and concludes.

## 2 Literature Review and Conceptual framework

. Moed (2005) argues that citation is formal and based on pre-defined evaluation procedures, open without any restriction, and enlightening rather than formulaic. A few authors (for instance Cole and Cole (1971); Bornmann et al. (2008); Norris and Oppenheim (2003)) more or less verify the correlation between the number of citations and research impact. However, important considerations in using citation counts must be highlighted.

[1] The dataset is presented in section 3.

A natural coherence between the number of citations and scientific impact generally implies that scientists select their references on the basis of the "quality" of the papers they cite, but this is not always the case. Authors sometimes cite papers to review the opposite view in the literature or to provide a general literature review (Amsterdamska and Leydesdorff 1989; Per O Seglen 1997b). In another study, Moed et al. (1985) note that citations refer to impact on the scientific community and it does not completely reflect research "quality". The authors argue that any publication should thus have a minimum "quality" to impact other research but other factors like visibility of journals and the extent to which researchers provide a public service are two other important determinants for citing a particular paper that do not necessarily have a strong correlation with research impact. Later, Moed (2009) argues that other citation-based indicators (e.g. journal impact factors, Hirsch indices, and normalized indicators of citation impact) can be substitute indicators for measuring the scientific impact of research.

Along these lines, Kostoff (1998) investigates the theory of citation and suggests that every citation results from the combination of two main reasons: the real component of intellectual heritage and random components of self-interest. Although there is a random component, the author argues that the random effect disappears in the aggregation of citation counts and therefore the number of citations is a good indicator of the "quality" of research. Phelan (1999) provides the same justification.

Assuming that citation counts are good proxies for the scientific impact of research, determinants of citation should be investigated. Some studies highlight a trade-off between "quality" and quantity of research (Broad 1981; Butler 2002; Hayes 1983), which suggests that "high-quality" work takes time and, hence, cannot be produced at the same speed as less-quality work. In contrast, some studies argue that scientists who publish more articles may be more visible to other members of scientific community and such scientists may thus receive more citations—an indication of cumulative effects in science (Merton 1968). There is some evidence in the literature indicating that a higher number of articles results in an increased number of citations in general (Feist 1997; Hayati and Ebrahimy 2009). Following this line of research, article counts can thus be considered as a proxy for the visibility of researchers (Aaltojärvi et al. 2008; Bar-Ilan et al. 2012).

However, the visibility of researchers does not solely depend on the number of articles they published and can stem from other factors. Ale Ebrahim et al. (2014) show that publication marketing tools and strategies significantly increase the article visibility and citation impact. Fowler and Aksnes (2007) indicate that self-citation improves the visibility of authors' prior works. It should be noted that the self-citation is not necessarily a negative point as it acts as a signal to readers about the author's prior work and background information (Sammarco 2008).

In addition, publication in more prestigious journals (sometimes with higher impact factors) may provide a higher visibility for articles and hence gain more citations (Stegmann and Grohmann 2001). Calderini and Franzoni (2004) and P.O. Seglen (1997a) argue that the impact factor of a journal may be one candidate to measure research quality or prestige. Although it is not a perfect justification to use journal impact factor as a proxy of research quality—as the distribution of citations within a journal is skewed and, hence, the impact factor of a journal is based on a few highly-cited items—previous research has shown that impact factors of journals does influence the citation counts of papers (Larivière and Gingras 2010). Following this literature, we test here whether the number of articles of a researcher in a given year or the journals' impact factor can influence his/her number of citations. With this evidence in mind, we propose our first hypothesis:

**Hypothesis 1**: *A higher visibility, measured by the number of articles and the journal impact factors, can positively affect the number of citations received in the future.*

Conducting collaborative research can have a positive effect on research quality. Using data on collaborative research conducted in Canada, Godin and Gingras (2000) indicate that not only is there a positive correlation between citations and its collaborative nature, but also new research opportunities are developed during research collaboration.

Scientific collaboration, however, depends on academic fields. Abramo et al. (2009) show that in interdisciplinary studies there are more collaborations than strictly disciplinary research, and also that collaboration with foreign organizations are more common in basic fields than in applied ones. The number of individuals in a research team can be a proxy for the extent of networking as it shows the ability of researchers in collective scientific actions. More specifically, the number of staff in a department and the number of co-authors of articles at the local, national, and international levels can significantly explain research productivity of scientists (Johnes 1988; Melin 1996). However, Katz and Martin (1997) argue that co-authorship is just a partial indicator of collaboration. They also mention that the concept of 'research collaboration' is hard to define: First, it is a social relationship and it is not possible to identify the beginning and the end; second, determinants of scientific collaboration depend on institutions, domains, countries, and time. Our second hypothesis relates to research team size and goes as follows:

**Hypothesis 2**: *The number of individuals in the list of authors significantly affects the future number of citations received by each scientist.*

Funding and the effect of financial resources on generating research outputs have been investigated by a great number of scholars over the years (Arundel and Geuna 2004; Harman

2000; Pavitt 2000, 2001). In most analyses, raising funds has been found to have a positive effect on scientific production. In terms of funding from the private sector, Gulbrandsen and Smeby (2005) find that industry funding is correlated with "new and interesting research topics and is prerequisite to accomplishing expensive and interesting projects" (pp. 947). The authors also indicate that professors with industrial funding collaborate more with other researchers, both from academia and industry.

Public funding can also have numerous effects on knowledge production. Beise and Stahl (1999) suggest the following six types of contribution publicly funded research may have: source of new useful knowledge, new instrumentation and methodologies, skills developed by those involved in carrying out basic research, expansion of national and international networks, dealing with complex problems[2], and creation of spin-off companies. However, it is not possible to claim the existence of a uniform and consistent argument about the effect of public funding; Appleyard (1996) indicates that public grants may have different effects on scientific production in two different countries or two different organisations.

To formulate a hypothesis on the effect of funding, it can be inferred that funding is a necessary condition for scientific research production but that it is not a sufficient condition for generating high quality research. Research funding acts as a tool for researchers, and should be combined with researchers' skills and expertise to generate high quality results. From another perspective, it is possible to argue that funding is generally enough to contribute to the advancement of knowledge, but that it does not guarantee that this knowledge will be of high impact. . Moreover, in laboratory-based fields, the role of funding may be more important in being able to perform

[2] The paper argues that solving complex problems provides great benefit for the firms and organizations facing such problems.

research at all. The source of funding may also be important in affecting the scientific impact of research, as it shapes the project's goal and application. Hence our third hypothesis reads as follows:

**Hypothesis 3**: *Research funding from public, private for profit, and private not for profit sources may have a positive effect on research scientific impact.*

The regression analysis described below will test whether having more financial support from the public, private and not-for-profit (NFP) sectors for the purposes paying research functional cost expenditures may increase the number of citations.

## 3 Data and Methodology

The data set used in this article integrates information about researchers funding and publication in the province of Quebec. We have access to Thomson Reuters Web of Science database on scientific articles (2000-2012), which includes information about date of publication, journal name, authors, affiliations, and number of citation each article receives. Funding information of scientists comes from the Quebec University Research Information System (*Système d'information sur la recherche universitaire* or SIRU) of the Ministry of Education, Leisure and Sports (MELS). This database reports funding information including grants and contracts of all Quebec academics, on a yearly basis during the period 1985-2012. Age and gender of scientists was obtained from the MELS internal database. After integration of the required variables from two sources, the data set description and summary statistics are reviewed in appendices 1 and 2.

Our dependent variable, the number of citations, varies from one discipline to another for two main reasons: (1) the number of papers and the amount of knowledge production is discipline dependent, and (2) citations may have different meaning in different disciplines. Two field classification are used in this paper. A first classification is used to categorize researchers into 9 broad domains, based on the U.S. classification of instructional programs[3]. The second classification is based on the field classification use by the U.S. National Science Foundation it is Science and Engineering Indicators series. This latter field and subfield classification is used to perform the normalization of citations. For instance, the citation patterns of papers in Economics is different from those of Political Science, despite the fact that both of them are subfields of Social Sciences. To control for such disciplinary-specific behaviour, we therefore calculate the relative number of citations [ln(*nbCitation*)], measured by the number of citations received so far by an article[4] divided by the average citation rate of the papers published in the same year in the same NSF speciality. This measure is then transformed by taking its natural logarithm to normalize the variables and satisfy the necessary conditions for regression equations in this paper[5].

We use variables which count the yearly number of articles [ln(*nbArticle*)], the yearly average number of authors in the papers of scientist [ln(*nbAuthor*)], and the five-year average of journal impact factor in which scientists publish [ln(*Impactfactor*)]. We also add dummy variables indicating the affiliated university of scientists. The variables [ln(*nbArticle*)] and

[3] The domains are Medical Science, Health Science, Business and Management, Social Science, Education, Humanities, Non-Health Professional, and Engineering, Science.

[4] More recent articles will have cumulated citations over a smaller number of years. To account for this fact, we add year dummy variables to the regression analysis.

[5] There are four main assumptions in regression analysis regarding the data: (1) the variables should have normal distributions; (2) the linear relationship should exist between dependent and independent variables; (3) the variables should be reported without error; and (4) the variance of errors is the same across all levels of the IV (homoscedasticity). With the same justification, the following variables in this paper are also transformed by logarithm function: funding, number of articles, number of authors and journal impact factor. For the variables that include a 0 value, we take the natural logarithm of 1 + the value of the variable.

[ln(*Impactfactor*)] are used for testing hypothesis 1 and the variable of [ln(*nbAuthor*)] is used to test hypothesis 2.

The next set of variables addresses specific funding information on the amount of money raised by each scientist. We separate the sources of research funding into public sector, private sector, or non-for-profit organization (NFP) with social and political mission. From another standpoint, research funding can have two purposes: it can be directly used for operation cost (O) such as research cost and researchers' salary (mainly student stipends and technical staff salaries) or it indirectly help research team buying instruments or larger infrastructure (I). In our framework, it is possible to have six variables for research funding for each researcher[6] [(1) ln(*PublicfundingO*); (2) ln(*PrivatefundingO*); (3) ln(*NFPfundingO*); (4) ln(*PublicfundingI*); (5) ln(*PrivatefundingI*); and (6) ln(*NFPfundingI*)][7]. The mentioned funding variables are used to test hypothesis 3.

Gender is also available in dataset [*dFemale*]. A great number of scholars have examined the gender effect on the number of publications and citations. Some articles show that women's publications are slightly less cited than that of men (Aksnes et al. 2011; Sugimoto et al. 2013). Long (1990) explains that women's opportunities for collaboration are significantly less than those of men's because women are more likely to have greater family responsibility than men. There are also some studies showing that women generally publish fewer articles than men do (Hesli and Lee 2011; Leahey 2006). This may result in less visibility and hence fewer citations

[6] In the cases that the funding belongs to more than one researcher, the total amount of funding is divided by the number of researchers to make number more informative and reliable about its effect on scientific publication.

[7] In practice, we do not use *PrivatefundingI* and *NFPfundingI* in our analysis because of there too rare occurrence in our database. Funding for instruments or large infrastructure has a long-term effect that we are not able to capture with a maximum number of 12 years in the sample including the necessary lag structure. As will be explained in the methodology section ln(*PublicfundingI*) is used as an instrument for ln(*PublicfundingO*).

for women. However, Long (1992) finds that there is no difference between men and women in terms of the number of citations or level of contribution in a team's work. All being considered, the effect of gender on the number of citations should be tested and controlled in our model.

There are two possible levels to investigate number of citations: the author level or the article level. As we measure the effect of funding on relative citations, we are obliged to consider the author as the observation unit, rather than the article, as it is impossible to attribute a specific amount of funding to a particular article (even with using information contained in the acknowledgements). All variables are therefore aggregated (summed) at the scientist-year level (the database base being built as a panel).

Our regression analysis aims to understand the citation count determinants, hence we estimate the contribution to the relative number of citations (left hand side variable) of the right hand side or explanatory variables reviewed above. It is important to note that the two variables of [ln(*PublicfundingO*)] and [ln(*nbCitation*)] are determined by each other simultaneously, and hence a potential source of endogeneity, which biases the ordinary least square (OLS) regressions. The main reason for this potential endogeneity is that the scientists are assessed for public funding based on their CV and past effectiveness while at the same time, publication and research quality depends on the funding capability of researchers.

Using instrumental variables (IV) is a standard technique suggested in literature to deal with such an endogeneity issue. Instruments must be correlated with the endogenous variable, and should not be correlated with the error term in the main regression equation, which implies that the instruments should not 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 (named here as 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.

In the first stage, the amount of public funding [ln(*PublicfundingO*)] is estimated by the following variables: (1) the number of articles is an important factor that measures the past performance of scientists [ln(*nbArticleAvg3*)] as it is the main component of one's CV; (2) infrastructure related public funding is a proxy indicating how much a scientist is equipped to conduct research in the frontier of knowledge [ln(*PublicfundingO*)][8]; (3) age and its square, which generally measures the a scientist' research experience [*Age, Age*$^2$]. There is evidence in literature about the non-linear effect of age (Bernier et al. 1975; Diamond 1986; Kyvik and Olsen 2008). It should be also noted that in first-stage regression, public funding is not affected by the number of articles in the same year but from the previous years. Hence a one-year lag of the three-year average of the number of articles is used as an instrument in the first stage regression. Using the same rationale, a one-year lag is also applied for the effect of infrastructure funding. The significance of the coefficients from the first stage regression (appendix 4) shows that these instruments are appropriate choices (they are significant).

Having the estimated amount of public funding from the first stage to tackle the endogeneity issue, the second stage estimates the relative number of citations received up to 10 years following the publication year [ln(*nbCitation*)] on funding [ln(*PublicfundingO*), ln(*PrivatefundingO*), and ln(*NFPfundingO*)], the number of articles [ln(*nbArticle*)], the average number of authors per paper [ln(*nbAuthor*)], and the five-year average of the journal impact

[8] Infrastructure funding is generally a measure of the capability of generating original knowledge but when it comes from private or non-for-profit (NFP) sources, it is not significant as instrument and it only plays a small role in the development of infrastructure.

factor [ln(*Impactfactor*)]. It should be noted that the funding variables are measured in three-year averages to smooth out large variations in yearly funding. The effect of gender is also tested.

Moreover, the interaction between the journal impact factor and the number of articles [ln(*nbArticle*) * ln(*Impactfactor*)] is added to investigate whether there is a moderating effect of the impact factor on the number of articles. The interactive variables may have additional and complementary explanatory power. For this paper, although the number of articles and journal impact factor may have an individual effect on the number of citations, their interaction can have significant collective effect.

In addition, we control for university fixed effects to account for any impact that our explanatory variables do not cover. For example, papers from McGill University and the University of Montreal (UdeM) receive more citations (figure 1[9]) than those of other universities in the province. We also add year dummy variables to account for year-specific characteristics of the research system as exemplified by the evolution of citations over time (figure 2). The significant time trend and differences between universities justify the existence of these dummy variables in the model. The possible reason behind yearly differences is that research fluctuates each year based on the economy and research policy and such fluctuation may affect the research quality. University dummy variables can have the same role as research setting and related motivations are partially university dependent. Considering the mentioned explanatory variables, the resulting model is given by:

[9] The small universities are grouped according to their institutional similarities. The University of Quebec 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).

$$1^{st}\ stage: \ln\left(PublicFundingO_{it}\right) = g\left(\ln\left(nbArticleAvg3_{it-1}\right), \ln\left(PublicFundingI_{it-1}\right), Age_{it}, Age_{it}^{2}\right)$$

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

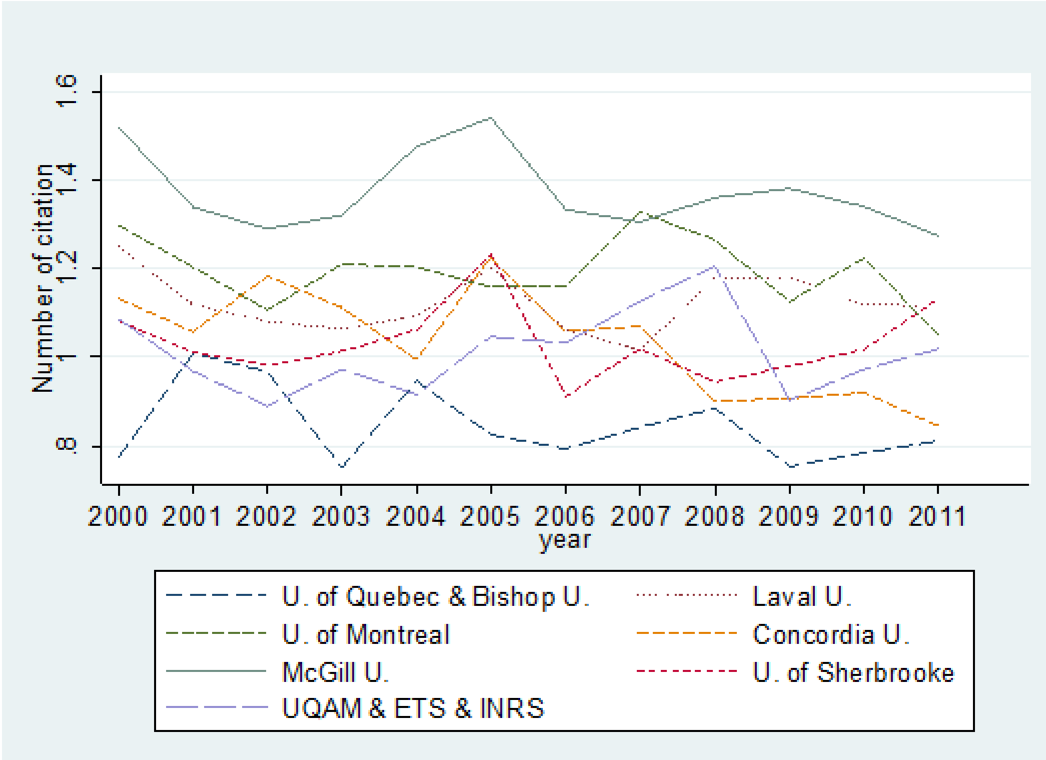


**Figure 1 – Discipline-normalized citation rates of papers from Quebec universities**


**Figure 2 – Average of Discipline-normalized citation rates of Quebec papers, by year**

Our data set is built as a panel, for which the ‘xtivreg’ Stata command is appropriate to estimate two-stage least square models (2SLS)[10]. The regression analysis is conducted for each domain separately. This results in a better understanding of determinants specific to different domains. Just focusing on the regression analysis for entire data set may generate some holistic arguments without any domain-specific interpretation. For instance, funding is not similar across domains, nor is the average number of authors per article. As the sample size per discipline precludes estimating the regressions for each discipline, the compromise used in this paper is to run regressions on groups of disciplines. Appendix 1 describes the variables, while summary statistics are presented in appendix 2 and the correlation table is shown in appendix 3.

## 4 Regression analysis

Because a number of our independent variables are individual fixed effects (for instance gender or university affiliation), we therefore prefer to estimate random effect 2SLS regressions, which are reported in table 1 (first stage regression results are reported in appendix 4). However, to check for the robustness of regression results, the data will be treated for both cross section and panel data. In addition to the two-stage regression (*xtivreg* command in Stata), a simple ordinary least square (OLS) without endogeneity is also tested for both cross section and panel data (*reg* command and *xtreg* command in Stata). The OLS and Panel OLS analysis are reported in appendices 5 and 6.

[10] Because we are using relative citations and not direct counts of citations, Poisson or negative binomial regressions are not appropriate.

**Table 1 - regression result (The second stage of two-stage least square regression) - *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively – Year dummies and university dummies are significant[11]**

| Dependent variable: ln(*nbCitation*)$_{it}$ | Domain | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | A+B | C | D | C+D | E | F | G | H | I | H+I | All |
| ***dFemale***$_i$ | 0.0095 | 0.0033 | 0.0073 | 0.0396 | -0.0019 | 0.0055 | -0.1010 | -0.1614 ** | -0.0820 | 0.0109 | -0.0245 | -0.0140 | -0.0063 |
| | (0.0134) | (0.0231) | (0.0116) | (0.0331) | (0.0221) | (0.0188) | (0.1007) | (0.0722) | (0.0622) | (0.0298) | (0.0193) | (0.0161) | (0.0077) |
| **ln(*nbArticle*)**$_{it}$ | 0.0329 *** | 0.0985 *** | 0.0422 *** | 0.0985 ** | 0.0498 ** | 0.0464 ** | 0.0127 | -0.0129 | 0.0886 * | 0.0845 *** | 0.0604 *** | 0.0679 *** | 0.0570 *** |
| | (0.0107) | (0.0223) | (0.0097) | (0.0450) | (0.0234) | (0.0211) | (0.1274) | (0.1593) | (0.0520) | (0.0149) | (0.0119) | (0.0092) | (0.0059) |
| **ln(*Impactfactor*)**$_{it}$ | 0.2987 *** | 0.2522 *** | 0.2875 *** | 0.3815 *** | 0.2678 *** | 0.3026 *** | 0.2612 * | 0.4352 ** | 0.0715 | 0.2462 *** | 0.2301 *** | 0.2370 *** | 0.2458 *** |
| | (0.0207) | (0.0474) | (0.0189) | (0.0692) | (0.0367) | (0.0319) | (0.1493) | (0.1702) | (0.0891) | (0.0294) | (0.0242) | (0.0185) | (0.0105) |
| **ln(*nbArticle*)**$_{it}$***ln(*Impactfactor*)**$_{it}$ | 0.1058 *** | 0.0934 ** | 0.1062 *** | -0.0184 | 0.0254 | 0.0046 | 0.0947 | -0.1007 | 0.1765 * | 0.0330 | 0.0738 *** | 0.0592 *** | 0.0901 *** |
| | (0.0171) | (0.0431) | (0.0158) | (0.0784) | (0.0392) | (0.0345) | (0.1754) | (0.2154) | (0.0990) | (0.0264) | (0.0206) | (0.0160) | (0.0094) |
| **ln(*nbAuthor*)**$_{it}$ | 0.1945 *** | 0.0585 *** | 0.1785 *** | 0.0755 ** | 0.0959 *** | 0.0894 *** | 0.0612 | 0.0264 | 0.2066 *** | 0.1599 *** | 0.1296 *** | 0.1389 *** | 0.1357 *** |
| | (0.0090) | (0.0203) | (0.0083) | (0.0362) | (0.0168) | (0.0150) | (0.0844) | (0.0720) | (0.0435) | (0.0153) | (0.0089) | (0.0076) | (0.0046) |
| **ln(*PublicfundingO*)**$_{it}$ | -0.0044 | -0.0085 | -0.0062 | 0.0084 | 0.0270 ** | .02068 * | -0.0328 | 0.0462 | 0.0240 | 0.0109 | 0.0302 *** | 0.0237 *** | -0.0012 |
| | (0.0045) | (0.0133) | (0.0043) | (0.0193) | (0.0133) | (0.0113) | (0.0456) | (0.0647) | (0.0332) | (0.0128) | (0.0082) | (0.0069) | (0.0034) |
| ***ln(PrivatefundingO)***$_{it}$ | 0.0028 *** | -0.0002 | 0.0025 *** | -0.0073 | 0.0012 | -0.0015 | 0.0024 | -0.0338 | -0.0093 | -0.0016 | -0.0032 ** | -0.0031 *** | 0.0001 |
| | (0.0010) | (0.0028) | (0.0009) | (0.0050) | (0.0036) | (0.0029) | (0.0193) | (0.0222) | (0.0065) | (0.0021) | (0.0015) | (0.0012) | (0.0006) |
| **ln(*NFPfundingO*)**$_{it}$ | -0.0014 | 0.0010 | -0.0009 | 0.0057 | 0.0005 | 0.0025 | 0.0086 | -0.0135 | -0.0028 | 0.0011 | -0.0006 | 0.0002 | -0.0004 |
| | (0.0011) | (0.0024) | (0.0010) | (0.0044) | (0.0025) | (0.0022) | (0.0134) | (0.0155) | (0.0082) | (0.0019 | (0.0014) | (0.0011) | (0.0006) |
| ***Constant*** | 0.2381 *** | 0.3000 ** | 0.2526 *** | 0.0673 | 0.0294 | 0.0764 | 0.8766 *** | -0.0276 | -0.0610 | -0.0174 | -0.0405 | -0.0334 | 0.2489 *** |
| | (0.0416) | (0.1207) | (0.0398) | (0.1619) | (0.1095) | (0.0929) | (0.3144) | (0.4933) | (0.2976) | (0.1223) | (0.0763) | (0.0649) | (0.0309) |
| **Number of observations** | 9026 | 1761 | 10787 | 1071 | 2944 | 4015 | 253 | 369 | 547 | 4029 | 7261 | 11290 | 31563 |
| **Number of scientists** | 1270 | 301 | 1571 | 301 | 673 | 974 | 104 | 202 | 178 | 664 | 1100 | 1764 | 5387 |
| **$\chi 2$** | 3665 *** | 578 *** | 4188 *** | 462 *** | 811 *** | 1188 *** | 84 *** | 134 *** | 182 *** | 941 *** | 1722 *** | 2695 *** | 9448 *** |
| **Average year activity** | 7.11 | 5.85 | 6.87 | 3.56 | 4.37 | 4.12 | 2.43 | 1.83 | 3.07 | 6.07 | 6.60 | 6.40 | 5.86 |
| **$R^2$ within groups** | 0.25 | 0.22 | 0.24 | 0.20 | 0.14 | 0.16 | 0.27 | 0.24 | 0.20 | 0.15 | 0.13 | 0.14 | 0.18 |
| **$R^2$ overall** | 0.32 | 0.27 | 0.31 | 0.31 | 0.24 | 0.26 | 0.29 | 0.28 | 0.26 | 0.23 | 0.23 | 0.23 | 0.27 |
| **$R^2$ between groups** | 0.47 | 0.24 | 0.43 | 0.39 | 0.35 | 0.35 | 0.19 | 0.25 | 0.24 | 0.28 | 0.38 | 0.35 | 0.34 |

[11] The definition of regression tags indicating the sample: A= Medical, B= Health Science, c= Business and Management, D= Social Science, E= Education, F= Humanities, G= Non-health professions, H= Engineering, I= Science.

The focus of this section is to analyse the two-stage least square regression (table 3) which addresses the problem of endogeneity. The regressions imply that, *ceteris paribus*, scientific publications of female scientists are cited in the same manner as men's publication. The variable of [*dFemale*] is not significant but there is a negative effect of being female only in the humanities (the small sample size and the fact that research output is badly measured by the Web of Science, we will not dwell on this result). This finding seems to be different from what Aksnes et al. (2011) and Larivière et al (2013) show about the underperformance of women.

The number of articles [ln(*nbArticle*)] has a significant positive effect except for education and the humanities. Hence, in general, a greater visibility of scientists, as determined by their number of articles, is associated with a greater citation rate.

Having added the interaction between journal impact factor [*ln(Impactfactor)*] and number of articles [ln(*nbArticle*)] to the regression, the results still show that not only there is a positive effect of articles count (due to the author's visibility), but also those papers in higher impact factor journals receive more citations (all domains except non-health professions). In addition, a higher number of articles in more higher impact factor journals (interactive variable) results in more citations than the same number of articles in a less prestigious journal (this effect is significant for medical, health science, non-health professions, and science). Consequently, it is possible to argue that hypothesis 1 becomes validated as the higher visibility, measured by the number of articles and the journal impact factors, can positively affect the number of citations.

Turning now to the size of research team, our results show that articles with more authors [ln(*nbAuthor*)] are more likely to be cited. It indicates that hypothesis 2 is validated. This finding is compatible with evidence in literature indicating positive effect of collaboration on research quality (Johnes 1988; Melin 1996). As a justification, in a research team with numerous researchers, tasks are done collectively and by different scientists. It probably provides some sort of knowledge spillover or tacit knowledge transfer, which improves capability of researchers in conducting high impact research.

Moreover, funding does not have a major effect on the number of citations. For example, [ln(*PublicfundingO*)] has a positive significant effect only for science and social science. The effect of [ln(*PrivatefundingO*)] is only significant for the medical (positive effect) and science (negative effect) domains. The coefficient of [ln(*NFPfundingO*)] is not significant at all. The empirical and theoretical evidence in literature supporting the positive effect of funding on publication performance of scientists are known (Arundel and Geuna 2004; Harman 2000; Pavitt 2000, 2001), but our results indicate that higher funding does not necessarily results in publications which are more cited. This paper does not contradict the positive effect of funding on scientific productivity but it indicates that higher funding is not a determinant of article citation counts. As a result, hypothesis 3 cannot be validated because there is no significant and comprehensive evidence for positive effect of funding from different sources (public, private, or NFP organization) on research quality.

## 5 Policy Implications

Assuming that the number of citations is a good proxy for research impact and, in turn, for a certain kind of *quality*, we propose some policy advice to address the issues discussed in the paper. First, it seems that collaborative works (measured by the size of research team) can influence the quality of the research, and policy makers should therefore encourage research of a more collaborative nature. The measure of such research collaboration is not only limited to the number of authors in each articles but can also measure the quality, extent, and durability of collaboration. However, the only variable we have on hand, which measure the research collaboration, is the number of individuals in the authors list.

Second, the significance level of the funding effect on research quality (which is not the same for all domains) does not necessarily imply that funding is ineffective for the knowledge production process/chain – this paper only investigates researchers' scientific impact and not their research productivity, which his considered as an input here. There is strong evidence in the literature about the significant effect of different funding types on scientific productivity (Manjarrés-Henríquez et al. 2008; Pavitt 2001; Salter and Martin 2001) to which our research contributes: public funding has a positive and significant effect for "Science" and "Social Science", while the private funding effect is positive and significant for "Science" and "Medical Science". These results point towards domain-specific policies and incentives because the effect of funding is domain-dependent, *ceteris paribus.* Hence a domain-specific policy can be an effective tool for improving the research quality in specific domains, without the need for general policy making, which may require holistic manipulation of science policy.

The third policy implication is to incentivise researchers to publish in journals with a high impact factor. Such journals have more visibility and their articles are widely read and used by the scientific community, more so than articles in other journals with lower impact factor. Moreover, the impact factor of a journal is a proxy for research quality because journals with higher impact factor have more submissions and editors are able to choose higher quality papers. Although a greater number of articles in the past contributes to improving the visibility of articles in the future (and hence their perceived research quality), the positive and significant effect of the interactive variable, which measures the modulating effect of the number of articles on the journal impact factor, on research quality implies that past articles in journals with a higher impact factor can reflect the intrinsic research quality of individuals. We can also argue that there is a learning experience from the past collaboration, especially if that collaboration led to a highly cited paper.

The last but not the least, our research did not find any gender bias in terms of research quality (except for the domain of Humanities). This contrasts with some evidence in literature that points towards the relative under-performance of women in terms of number of articles and research quality. However, it is not possible to make a policy conclusion in this regard because there may be a great number of reasons that may explain why women are less cited in one specific domain. First, they publish less and are thus less visible. Second, there may be conscious or unconscious discrimination. Third, women may be involved in more multidisciplinary research that is harder to publish. Fourth, women may spend more time involved in other duties at university. Although more investigations are needed, some incentive programs can be to encourage women to apply for more funding, and to go to more conferences (to be more visible). In addition, mentorship programs should be put in place where the gap is significant.

## 6 Conclusion

The paper investigates the determinants of citation counts as an indicator of research quality. To reach that aim, three propositions were set to be validated: one on the effect of funding, one on the effect of research team size, and another one on the effect of articles count and journal impact factor. The last two have been validated and the first one is rejected. In conclusion, this paper shows that the number of articles and the visibility of a researcher, the impact factor of the journal, the size of the research team, and the institutional setting of the university (fixed effect) are the important determinants of citation counts. In addition, the regressions show that there is no significant effect of public funding and gender in most of the domains examined. Furthermore, it should be noted that for the domains of Education, Humanities, and Non-health professions, hypothesis 1 (about the effect of articles count and journal impact factor) is not validated. Moreover, hypothesis 2 (about the effect of research team size) for the domains of Education and Humanities is not validated.

In terms of validity of abovementioned interpretations, it should be noted that the study only covers Quebec scientists and some data entries are missing in original dataset. In addition to using more comprehensive and complete data set, there are also some other suggestions for future works in this subject. First, it would be interesting to investigate the citations and group them to distinguish self-citations, citations due to high research quality, and citations for bringing evidence from literature. Each group of citations may have different sets of determinants. Second, it would also be of importance to investigate the time trend of citation whether it is possible to make some arguments about timing of citation accumulation of each scientist or each paper, based on different events and different factors.

# 6- Appendices:

**Appendix 1 – Variable description (Number of observations = 31,563)**

| Variable name | Variable description |
|---|---|
| **ln(*nbCitation*)$_{it}$** | Natural logarithm of number of citations of papers published by scientist i in year t (10 years following publication year) divided by the average citation rate of the papers published in the same year in the same discipline |
| **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(*PublicfundingI*)$_{it}$** | Natural logarithm of the three-year average up to year t of public sector funding for the purpose of buying instruments for 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 number of articles published in year t by researcher i |
| **ln(*nbAuthor*)$_{it}$** | Natural logarithm of the three-year average up to year t of number of authors in the papers of researcher i |
| **ln(*Impactfactor*)$_{it}$** | Natural logarithm of the five-year average up to year t of journal impact factor in which the scientist publishes |
| **ln(*nbArticle*)$_{it}$* ln(*Impactfactor*)$_{it}$** | Interaction between ln(*nbArticle*)$_{it}$ and ln(*Impactfactor*)$_{it}$: ln(*nbArticle*)$_{it}$ X ln(*Impactfactor*)$_{it}$ |
| ***dFemale$_i$*** | Dummy variable taking the value 1 if the scientist is a woman and 0 otherwise |
| ***Age$_{it}$*** | Age of a researcher i at year t |
| ***d2000, d2001, ….*** | Dummy variables indicating the year |
| **ln(*nbArticleAvg3*)$_{it}$** | Natural logarithm of the three-year average up to year t of articles published by researcher i |

**Appendix 2 - Summary statistics (Number of observation = 31,563) – the variables are not summarized in logarithmic scale and they are raw amount[12]**

| variable | Minimum | Mean | Maximum | Standard Deviation |
|---|---|---|---|---|
| ***nbCitation*** | 0 | 1.183177 | 74.575 | 1.74 |
| ***Age*** | 16 | 50.62187 | 92 | 9.26 |
| ***dfemale*** | 0 | 0.228939 | 1 | 0.42 |
| ***nbArticle*** | 0 | 3.469125 | 84 | 3.66 |
| ***Impactfactor*** | 0.016 | 1.130298 | 12.476 | 0.63 |
| ***nbAuthor*** | 1 | 7.269147 | 3174.5 | 52.02 |
| ***PublicfundingO*** | 0 | 112969.6 | 1.01E+07 | 203169.30 |
| ***PrivatefundingO*** | 0 | 21543.06 | 4077193 | 96437.42 |
| ***NFPfundingO*** | 0 | 21294.22 | 8720387 | 128658.50 |
| ***PublicfundingI*** | 0 | 30482.68 | 1.28E+07 | 222562.50 |

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

## Appendix 3 - Correlation table - the stars show 1% significance (No. observation: 31,563)

| | ln(*nbCitation*)$_{it}$ | ln(*PublicfundingI*)$_{it-1}$ | *Age*$_{it}$ | ln(*nbArticleAvg3*)$_{it-1}$ | *dFemale*$_{i}$ | ln(*nbArticle*)$_{it}$ | *ln*(Impactfactor)$_{it}$ | ln(*nbArticle*)$_{it}$*ln(Impactfactor)$_{it}$ | ln(*nbAuthor*)$_{it}$ | ln(*PublicfundingO*)$_{it}$ | ln(*PrivatefundingO*)$_{it}$ | ln(*NFPfundingO*)$_{it}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ln(*nbCitation*)$_{it}$** | 1 | | | | | | | | | | | |
| **ln(*PublicfundingI*)$_{it-1}$** | 0.0733* | 1 | | | | | | | | | | |
| ***Age*$_{it}$** | -0.0377* | -0.1247* | 1 | | | | | | | | | |
| **ln(*nbArticleAvg3*)$_{it-1}$** | 0.1945* | 0.2000* | 0.1130* | 1 | | | | | | | | |
| ***dFemale*$_{i}$** | -0.0318* | -0.1020* | -0.1197* | -0.1203* | 1 | | | | | | | |
| **ln(*nbArticle*)$_{it}$** | 0.2359* | 0.1912* | 0.0136 | 0.6728* | -0.0995* | 1 | | | | | | |
| **ln(*Impactfactor*)$_{it}$** | 0.4686* | 0.1029* | -0.0440* | 0.2209* | -0.0406* | 0.2192* | 1 | | | | | |
| **ln(*nbArticle*)$_{it}$*ln(*Impactfactor*)$_{it}$** | 0.4695* | 0.1114* | -0.0409* | 0.2772* | -0.0449* | 0.3005* | 0.9087* | 1 | | | | |
| **ln(*nbAuthor*)$_{it}$** | 0.2724* | 0.0039 | 0.0668* | 0.3098* | 0.0046 | 0.2898* | 0.2382* | 0.2794* | 1 | | | |
| **ln(*PublicfundingO*)$_{it}$** | 0.0707* | 0.2905* | -0.1044* | 0.2571* | -0.0218* | 0.2347* | 0.1094* | 0.1163* | -0.0002 | 1 | | |
| **ln(*PrivatefundingO*)$_{it}$** | 0.0723* | 0.1415* | -0.007 | 0.2141* | -0.1219* | 0.2163* | 0.0487* | 0.0550* | 0.1385* | 0.1294* | 1 | |
| **ln(*NFPfundingO*)$_{it}$** | 0.1017* | 0.1145* | -0.0603* | 0.2470* | -0.0178* | 0.2339* | 0.1216* | 0.1422* | 0.1952* | 0.1963* | 0.2227* | 1 |

## Appendix 4 - Regression results (first stage regression) - *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively – Year dummies and university dummies are significant[13]

| Dependent variable: *PublicfundingO* | Domain | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | A+B | C | D | C+D | E | F | G | H | I | H+I | All |
| ***dFemale***$_i$ | 0.2876 *** | 0.2993 * | 0.4328 *** | 0.2781 | 0.2907 ** | 0.2834 *** | -0.1251 | 0.1731 | 0.9591 ** | -0.0572 | -0.0723 | -0.0683 | 0.1568 *** |
| | (0.1103) | (0.1687) | (0.0938) | (0.2072) | (0.1213) | (0.1064) | (0.4732) | (0.2831) | (0.3760) | (0.1405) | (0.0996) | (0.0806) | (0.0492) |
| **ln(*nbArticle*)**$_{it}$ | 0.4649 *** | 0.3078 ** | 0.4427 *** | 0.7069 *** | 0.4027 *** | 0.4487 *** | 0.4014 | 0.3092 | 0.2613 | 0.1989 *** | 0.3317 *** | 0.2856 *** | 0.3922 *** |
| | (0.0815) | (0.1534) | (0.0725) | (0.2701) | (0.1174) | (0.1077) | (0.5902) | (0.5991) | (0.3978) | (0.0667) | (0.0562) | (0.0429) | (0.0343) |
| **ln(*Impactfactor*)**$_{it}$ | 0.6793 *** | 0.1979 | 0.5225 *** | 0.1001 | 0.2232 | 0.2232 | 0.2278 | 0.1913 | 0.4793 | 0.2884 ** | 0.4161 *** | 0.3557 *** | 0.3085 *** |
| | (0.1670) | (0.3367) | (0.1496) | (0.4328) | (0.2017) | (0.1796) | (0.6987) | (0.6751) | (0.6336) | (0.1364) | (0.1216) | (0.0905) | (0.0658) |
| **ln(*nbArticle*)**$_{it}$***ln(*Impactfactor*)**$_{it}$ | -0.2200 | -0.1469 | -0.1574 | 0.2390 | -0.0709 | -0.0426 | -0.2590 | -0.0838 | -0.3334 | -0.1483 | -0.2412 ** | -0.1954 ** | -0.1326 ** |
| | (0.1394) | (0.3063) | (0.1265) | (0.4913) | (0.2157) | (0.1947) | (0.8194) | (0.8580) | (0.7133) | (0.1231) | (0.1044) | (0.0792) | (0.0594) |
| **ln(*nbAuthor*)**$_{it}$ | -0.4378 *** | -0.1446 | -0.4527 *** | -0.2270 | 0.1611 * | 0.1010 | 0.6436 * | 0.3161 | -0.6015 ** | -0.0372 | -0.2073 *** | -0.1671 *** | -0.3551 *** |
| | (0.0710) | (0.1445) | (0.0639) | (0.2269) | (0.0912) | (0.0847) | (0.3632) | (0.2572) | (0.2847) | (0.0716) | (0.0456) | (0.0375) | (0.0286) |
| **ln(*PrivatefundingO*)**$_{it}$ | -0.0382 *** | 0.0556 *** | -0.0318 *** | 0.1253 *** | 0.0663 *** | 0.0893 *** | 0.0786 | 0.1491 * | 0.0189 | 0.0962 *** | 0.0676 *** | 0.0776 *** | 0.0267 *** |
| | (0.0082) | (0.0184) | (0.0075) | (0.0261) | (0.0191) | (0.0153) | (0.0887) | (0.0779) | (0.0475) | (0.0071) | (0.0068) | (0.0049) | (0.0038) |
| **ln(*NFPfundingO*)**$_{it}$ | 0.0895 *** | 0.0451 *** | 0.0800 *** | 0.0897 *** | 0.0490 *** | 0.0568 *** | 0.1222 *** | 0.1051 * | 0.1543 *** | 0.0489 *** | 0.0601 *** | 0.0560 *** | 0.0574 *** |
| | (0.0077) | (0.0157) | (0.0070) | (0.0247) | (0.0131) | (0.0116) | (0.0562) | (0.0589) | (0.04210 | (0.0080) | (0.0066) | (0.0051) | (0.0036) |
| **ln(*PublicfundingI*)**$_{it-1}$ | 0.1217 *** | 0.1007 *** | 0.1191 *** | 0.0719 *** | 0.0769 *** | 0.0728 *** | 0.0089 | 0.0837 * | 0.1662 *** | 0.0666 *** | 0.0754 *** | 0.0717 *** | 0.0927 *** |
| | (0.0088) | (0.0209) | (0.0081) | (0.0244) | (0.0129) | (0.0113) | (0.0631) | (0.0451) | (0.0568) | (0.0065) | (0.0058) | (0.0043) | (0.0036) |
| **Age**$_{it}$ | 0.1304 *** | 0.4570 *** | 0.1872 *** | 0.1685 * | 0.0468 | 0.0370 | 0.5601 *** | -0.1593 | 0.3943 ** | 0.1431 *** | 0.2126 *** | 0.1886 *** | 0.1550 *** |
| | (0.0399) | (0.0823) | (0.0359) | (0.0951) | (0.0474) | (0.0425) | (0.1889) | (0.1288) | (0.1627) | (0.0331) | (0.0296) | (0.0223) | (0.0171) |
| **Age**$^2_{it}$ | -0.0017 *** | -0.0048 *** | -0.0022 *** | -0.0023 *** | -0.0007 | -0.0007 | -0.0061 *** | 0.0014 | -0.0041 ** | -0.0016 *** | -0.0024 *** | -0.0021 *** | -0.0018 *** |
| | (0.0004) | (0.0008) | (0.0003) | (0.0010) | (0.0005) | (0.0004) | (0.0018) | (0.0013) | (0.0016) | (0.0003) | (0.0003) | (0.0002) | (0.0002) |
| **ln(*nbArticleAavg3*)**$_{it-1}$ | 1.5659 *** | 0.8759 *** | 1.4303 *** | 1.1544 *** | 1.1444 *** | 1.1767 *** | 0.8231 * | 1.2543 *** | -0.1928 | 0.6174 *** | 0.7661 *** | 0.7135 *** | 0.9977 *** |
| | (0.0816) | (0.1489) | (0.0725) | (0.2222) | (0.1085) | (0.0969) | (0.4539) | (0.4056) | (0.3582) | (0.0663) | (0.0568) | (0.0432) | (0.0339) |
| ***Constant*** | 4.4273 *** | -2.4294 0.2430 | 3.2488 *** | 4.9033 ** | 6.9708 *** | 7.3620 *** | -6.1381 | 11.3988 *** | -0.3949 | 5.7952 *** | 4.0233 *** | 4.6705 *** | 4.9276 *** |
| | (1.0653) | (2.0788) | (0.9517) | (2.3117) | (1.1890) | (1.0572) | (4.9201) | (3.2269) | (4.0890) | (0.8470) | (0.7648) | (0.5738) | (0.4411) |
| **Number of observations** | 9026 | 1761 | 10787 | 1071 | 2944 | 4015 | 253 | 369 | 547 | 4029 | 7261 | 11290 | 31563 |
| **χ2** | 1770 *** | 256 *** | 1906 *** | 277 *** | 510 *** | 704 *** | 75 *** | 63 *** | 124 *** | 983 *** | 1490 *** | 2398 *** | 4942 *** |

[13] The definition of regression tags indicating the sample: A= Medical, B= Health Science, c= Business and Management, D= Social Science, E= Education, F= Humanities, G= Non-health professions, H= Engineering, I= Science

**Appendix 5 - regression result (OLS) - *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively – Year dummies and university dummies are significant**[14]

| Dependent variable: ln(*nbCitation*) | Domain | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A | B | A+B | C | D | C+D | E | F | G | H | I | H+I | All |
| ***dFemale**$_i$* | 0.0072 | 0.0001 | 0.0074 | 0.0229 | 0.0103 | 0.0112 | -0.0644 | -0.1348 * | -0.0498 | 0.0067 | -0.0245 | -0.0161 | -0.0049 |
| | (0.0123) | (0.0201) | (0.0106) | (0.0389) | (0.0212) | (0.0186) | (0.0824) | (0.0732) | (0.0389) | (0.0297) | (0.0161) | (0.0141) | (0.0070) |
| **ln(*nbArticle*)$_{it}$** | 0.0359 *** | 0.0755 *** | 0.0395 *** | 0.0981 ** | 0.0817 *** | 0.0812 *** | -0.0399 | 0.0416 | 0.0629 | 0.1203 *** | 0.1037 *** | 0.1083 *** | 0.0667 *** |
| | (0.0088) | (0.0186) | (0.0080) | (0.0393) | (0.0202) | (0.0178) | (0.0853) | (0.1732) | (0.0416) | (0.0150) | (0.0134) | (0.0104) | (0.0057) |
| **ln(*Impactfactor*)$_{it}$** | 0.2978 *** | 0.2194 *** | 0.2801 *** | 0.3672 *** | 0.2582 *** | 0.2928 *** | 0.2082 | 0.4411 ** | 0.0784 | 0.2249 *** | 0.2626 *** | 0.2497 *** | 0.2461 *** |
| | (0.0278) | (0.0444) | (0.0248) | (0.0687) | (0.0392) | (0.0328) | (0.1367) | (0.1770) | (0.0775) | (0.0297) | (0.0383 ) | (0.0284) | (0.0154) |
| **ln(*nbArticle*)$_{it}$*ln(*Impactfactor*)$_{it}$** | 0.1172 *** | 0.1185 *** | 0.1201 *** | -0.0227 | 0.0630 | 0.0362 | 0.2107 | -0.0940 | 0.1817 ** | 0.0641 ** | 0.0899 *** | 0.0800 *** | 0.1100 *** |
| | (0.0220) | (0.0388) | (0.0199) | (0.0781) | (0.0412) | (0.0350) | (0.1557) | (0.2287) | (0.0811) | (0.0256) | (0.0347) | (0.0259) | (0.0137) |
| **ln(*nbAuthor*)$_{it}$** | 0.1944 *** | 0.0499 * | 0.1782 *** | 0.0420 | 0.0802 *** | 0.0662 *** | 0.0186 | 0.0596 | 0.1710 *** | 0.1440 *** | 0.0728 *** | 0.0892 *** | 0.1090 *** |
| | (0.0141) | (0.0279) | (0.0131) | (0.0375) | (0.0187) | (0.0167) | (0.0712) | (0.0562) | (0.0465) | (0.0214) | (0.0177) | (0.0156) | (0.0086) |
| **ln(*PublicfundingO*)$_{it}$** | -0.0008 | -0.0001 | -0.0010 | 0.0085 * | 0.0092 *** | 0.0080 *** | 0.0021 | 0.0129 | -0.0014 | 0.0030 | 0.0040 | 0.0040 | 0.0001 |
| | (0.0013) | (0.0029) | (0.0012) | (0.0046) | (0.0031) | (0.0025) | (0.0114) | (0.0096) | (0.0059) | (0.0036) | (0.0030) | (0.0024) | (0.0009) |
| **ln(*PrivatefundingO*)$_{it}$** | 0.0043 *** | -0.0021 | 0.0036 *** | -0.0043 | 0.0040 | 0.0022 | 0.0126 | -0.0262 *** | -0.0048 | -0.0006 | -0.0013 | -0.0015 | 0.0013 ** |
| | (0.0010) | (0.0025) | (0.0010) | (0.0053) | (0.0033) | (0.0029) | (0.0138) | (0.0092) | (0.0053) | (0.0018) | (0.0015) | (0.0011) | (0.0007) |
| **ln(*NFPfundingO*)$_{it}$** | -0.0030 *** | 0.0031 | -0.0021 *** | 0.0072 | 0.0041 | 0.0050 ** | 0.0075 | -0.0148 | 0.0066 | -0.0006 | 0.0013 | 0.0010 | -0.0009 |
| | (0.0010) | (0.0024) | (0.0009) | (0.0044) | (0.0025) | (0.0022) | (0.0100) | (0.0118) | (0.0054) | (0.0019) | (0.0015) | (0.0012) | (0.0006) |
| ***Constant*** | 0.2768 *** | 0.3968 *** | 0.2945 *** | 0.2945 *** | 0.2575 *** | 0.2822 *** | 0.8582 *** | 0.4314 ** | 0.3331 ** | 0.1809 *** | 0.3295 *** | 0.2872 *** | 0.3584 *** |
| | (0.0290) | (0.0696) | (0.0269) | (0.0820) | (0.0463) | (0.0398) | (0.2167) | (0.1896) | (0.1301) | (0.0494) | (0.0429) | (0.0352) | (0.0175) |
| **Number of observations** | 10124 | 1954 | 12078 | 1243 | 3265 | 4508 | 281 | 420 | 613 | 4466 | 8089 | 12555 | 35201 |
| **Log likelihood** | -4776.06 | -825.66 | -5675.60 | -747.64 | -1753.85 | -2526.03 | -193.77 | -390.39 | -345.74 | -2162.48 | -4108.04 | -6312.27 | -17887.10 |
| **F-test** | . | 25.34 *** | 98.82 *** | 21.31 *** | 36.59 *** | 53.31 *** | 12.50 *** | 57.48 *** | 25.17 *** | 45.19 *** | 52.96 *** | 84.53 *** | 268.80 *** |
| **$R^2$** | 0.32 | 0.25 | 0.31 | 0.28 | 0.26 | 0.26 | 0.33 | 0.28 | 0.28 | 0.22 | 0.24 | 0.24 | 0.27 |

[14] The definition of regression tags indicating the sample: A= Medical, B= Health Science, c= Business and Management, D= Social Science, E= Education, F= Humanities, G= Non-health professions, H= Engineering, I= Science

**Appendix 6 - Regression results (One-stage panel regression) - *, **, and *** show the significance level at 0.05, 0.02, and 0.01 respectively – Year dummies and university dummies are significant[15]**

| **Dependent variable:** **ln(*nbCitation*)** | **Domain** | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **A** | **B** | **A+B** | **C** | **D** | **C+D** | **E** | **F** | **G** | **H** | **I** | **H+I** | **All** |
| ***dFemale*** $_i$ | 0.0061 | 0.0010 | 0.0058 | 0.0070 | 0.0100 | 0.0102 | -0.0733 | -0.1126 | -0.0612 | 0.0009 | -0.0268 | -0.0183 | -0.0063 |
| | (0.0137) | (0.0228) | (0.0121) | (0.0375) | (0.0201) | (0.0182) | (0.0918) | (0.0708) | (0.0473) | (0.0311) | (0.0179) | (0.0157) | (0.0083) |
| **ln(*nbArticle*)** $_{it}$ | 0.0234 *** | 0.0752) *** | 0.0288 *** | 0.0231 | 0.0689 *** | 0.0560 *** | -0.0466 | 0.0425 | 0.0819 * | 0.0753 | 0.0741 *** | 0.0732 *** | 0.0438 *** |
| | (0.0087) | (0.0189) | (0.0080) | (0.0396) | (0.0190) | (0.0171) | (0.1068) | (0.1327) | (0.0479) | (0.0130) | (0.0098) | (0.0077) | (0.0049) |
| **ln(*Impactfactor*)** $_{it}$ | 0.3027 *** | 0.2358 *** | 0.2889 *** | 0.3958 *** | 0.2709 *** | 0.3049 *** | 0.2484 * | 0.4423 *** | 0.1064 | 0.2359 *** | 0.2380 *** | 0.2346 *** | 0.2472 *** |
| | (0.0193) | (0.0446) | (0.0177) | (0.0640) | (0.0342) | (0.0299) | (0.1346) | (0.1628) | (0.0798) | (0.0280) | (0.0223) | (0.0173) | (0.0099) |
| **ln(*nbArticle*)** $_{it}$ ***ln(*Impactfactor*)** $_{it}$ | 0.0956 *** | 0.0928 ** | 0.0958 *** | -0.0771 | 0.0372 | 0.0081 | 0.1043 | -0.0995 | 0.1388 | 0.0523 ** | 0.0782 *** | 0.0704 *** | 0.0859 *** |
| | (0.0163) | (0.0405) | (0.0151) | (0.0728) | (0.0369) | (0.0327) | (0.1583) | (0.2059) | (0.0918) | (0.0256) | (0.0193) | (0.0152) | (0.0090) |
| **ln(*nbAuthor*)** $_{it}$ | 0.2045 *** | 0.0601 *** | 0.1884 *** | 0.0528 | 0.0853 *** | 0.0751 *** | 0.0308 | 0.0674 | 0.1675 *** | 0.1425 *** | 0.1124 *** | 0.1239 *** | 0.1376 *** |
| | (0.0082) | (0.0195) | (0.0076) | (0.0339) | (0.0151) | (0.0139) | (0.0712) | (0.0573) | (0.0352) | (0.0151) | (0.0083) | (0.0072) | (0.0044) |
| **ln(*PublicfundingO*)** $_{it}$ | -0.0006 | -0.0001 | -0.0008 | 0.0089 * | 0.0073 ** | 0.0069 *** | 0.0060 | 0.0120 | -0.0039 | 0.0027 | 0.0064 *** | 0.0054 *** | 0.0012 |
| | (0.0012) | (0.0031) | (0.0011) | (0.0048) | (0.0030) | (0.0026) | (0.0131) | (0.0124) | (0.0056) | (0.0030) | (0.0020) | (0.0017) | (0.0008) |
| **ln(*PrivatefundingO*)** $_{it}$ | 0.0032 *** | -0.0003 | 0.0029 *** | -0.0064 | 0.0027 | 0.0002 | 0.0005 | -0.0240 | -0.0076 | -0.0005 | -0.0008 | -0.0011 | 0.0003 |
| | (0.0010) | (0.0025) | (0.0009) | (0.0045) | (0.0032) | (0.0026) | (0.0180) | (0.0183) | (0.0059) | (0.0015) | (0.0013) | (0.0010) | (0.0006) |
| **ln(*NFPfundingO*)** $_{it}$ | -0.0021 ** | 0.0013 | -0.0016 * | 0.0088 ** | 0.0026 | 0.0042 ** | 0.0020 | -0.0169 | 0.0050 | 0.0000 | 0.0010 | 0.0008 | -0.0009 |
| | (0.0009) | (0.0021) | (0.0008) | (0.0041) | (0.0022) | (0.0020) | (0.0114) | (0.0149) | (0.0055) | (0.0017) | (0.0012) | (0.0010) | (0.0006) |
| ***Constant*** | 0.2690 ** | 0.3782 * | 0.2848 ** | 0.3647 * | 0.2731 ** | 0.3033 ** | 0.7501 | 0.4237 | 0.3376 * | 0.2333 ** | 0.2918 ** | 0.2705 ** | 0.3384 ** |
| | (0.0233) | (0.0540) | (0.0216) | (0.0793) | (0.0418) | (0.0368) | (0.1817) | (0.1739) | (0.0990) | (0.0429) | (0.0289) | (0.0240) | (0.0134) |
| **Number of observations** | 10124 | 1954 | 12078 | 1243 | 3265 | 4508 | 281 | 420 | 613 | 4466 | 8089 | 12555 | 35201 |
| **Number of scientists** | 1330 | 313 | 1643 | 336 | 708 | 1044 | 111 | 225 | 191 | 679 | 1131 | 1810 | 5634 |
| **χ2** | . *** | 537.41 *** | 4479.74 *** | 390.77 *** | 923.51 *** | 1258.02 *** | 92.61 *** | 144.61 *** | 190.72 *** | 880.63 *** | 1841.27 *** | 2734.85 *** | 9506.55 *** |
| **Average number of years** | 7.61 | 6.24 | 7.35 | 3.70 | 4.61 | 4.32 | 2.53 | 1.87 | 3.21 | 6.58 | 7.15 | 6.94 | 6.25 |
| **$R^2$ within groups** | 0.25 | 0.19 | 0.23 | 0.18 | 0.15 | 0.15 | 0.25 | 0.23 | 0.19 | 0.13 | 0.13 | 0.13 | 0.17 |
| **$R^2$ overall** | 0.32 | 0.25 | 0.31 | 0.28 | 0.26 | 0.26 | 0.31 | 0.28 | 0.28 | 0.21 | 0.24 | 0.23 | 0.27 |
| **$R^2$ between groups** | 0.49 | 0.26 | 0.44 | 0.33 | 0.38 | 0.35 | 0.27 | 0.25 | 0.27 | 0.28 | 0.41 | 0.37 | 0.35 |

[15] The definition of regression tags indicating the sample: A= Medical, B= Health Science, c= Business and Management, D= Social Science, E= Education, F= Humanities, G= Non-health professions, H= Engineering, I= Science

**References**:


Aaltojärvi, Inari, et al. (2008), 'Scientific productivity, web visibility and citation patterns in sixteen Nordic sociology departments', *Acta Sociologica,* 51 (1), 5-22.

Abramo, G., D'Angelo, C.A., and Di Costa, F. (2009), 'Research collaboration and productivity: is there correlation?', *Higher Education,* 57 (2), 155-71.

Adam, David (2002), 'Citation analysis: The counting house', *Nature,* 415 (6873), 726-29.

Aksnes, D.W., et al. (2011), 'Are female researchers less cited? A large ‐ scale study of Norwegian scientists', *Journal of the American Society for Information Science and Technology,* 62 (4), 628-36.

Ale Ebrahim, Nader, et al. (2014), 'Visibility and citation impact', *International Education Studies,* 7 (4), 120-25.

Amsterdamska, O. and Leydesdorff, L. (1989), 'Citations: indicators of significance?', *Scientometrics,* 15 (5), 449-71.

Appleyard, M.M. (1996), 'How does knowledge flow? Interfirm patterns in the semiconductor industry', *Strategic management journal,* 17, 137-54.

Arundel, A. and Geuna, A. (2004), 'Proximity and the use of public science by innovative European firms', *Economics of Innovation and new Technology,* 13 (6), 559-80.

Bar-Ilan, Judit, et al. (2012), 'Beyond citations: Scholars' visibility on the social Web', *arXiv preprint arXiv:1205.5611*.

Beise, M. and Stahl, H. (1999), 'Public research and industrial innovations in Germany', *Research Policy,* 28 (4), 397-422.

Bernier, C., Gill, W., and Hunt, R. (1975), 'Measures of excellence of engineering and science departments: a chemical engineering example', *Chemical Engineering Education,* 9, 194-202.

Bornmann, Lutz, et al. (2008), 'Citation counts for research evaluation: standards of good practice for analyzing bibliometric data and presenting and interpreting results', *Ethics in Science and Environmental Politics,* 8 (1), 93-102.

Broad, William J (1981), 'The publishing game: getting more for less', *Science,* 211 (4487), 1137-39.

Brown, Lawrence D and Gardner, John C (1985), 'Applying citation analysis to evaluate the research contributions of accounting faculty and doctoral programs', *Accounting Review*, 262-77.

Butler, Linda (2002), 'A list of published papers is no measure of value', *Nature,* 419 (6910), 877-77.

Calderini, M. and Franzoni, C. (2004), 'Is academic patenting detrimental to high quality research', *An empirical analysis of the relationship between scientific careers and patent applications. Bocconi University: Cespri Working Paper,* (162).

Cole, Jonathan and Cole, Stephen (1971), 'Measuring the Quality of Sociological Research: Problems in the Use of the" Science Citation Index"', *The American Sociologist,* 6 (1), 23-29.

Diamond, A.M. (1986), 'The life-cycle research productivity of mathematicians and scientists', *Journal of Gerontology,* 41 (4), 520.

Feist, Gregory J (1997), 'Quantity, quality, and depth of research as influences on scientific eminence: Is quantity most important?', *Creativity Research Journal,* 10 (4), 325-35.

Fowler, James H and Aksnes, Dag W (2007), 'Does self-citation pay?', *Scientometrics,* 72 (3), 427-37.

Godin, B. and Gingras, Y. (2000), 'Impact of collaborative research on academic science', *Science and Public Policy,* 27 (1), 65-73.

Gulbrandsen, M. and Smeby, J.C. (2005), 'Industry funding and university professors' research performance', *Research Policy,* 34 (6), 932-50.

Harman, G. (2000), 'Allocating research infrastructure grants in post-binary higher education systems: British and Australian approaches', *Journal of Higher Education Policy and Management,* 22 (2), 111-26.

Hayati, Zouhayr and Ebrahimy, Saeideh (2009), 'Correlation between quality and quantity in scientific production: A case study of Iranian organizations from 1997 to 2006', *Scientometrics,* 80 (3), 625-36.
Hayes, Steven C (1983), 'When more is less: Quantity versus quality of publications in the evaluation of academic vitae', *American Psychologist,* 38 (12), 1398.
Hesli, V.L. and Lee, J.M. (2011), 'Faculty Research Productivity: Why Do Some of Our Colleagues Publish More than Others?', *PS Political Science and Politics,* 44 (2), 393.
Johnes, G. (1988), 'Determinants of research output in economics departments in British universities', *Research Policy,* 17 (3), 171-78.
Katz, J.S. and Martin, B.R. (1997), 'What is research collaboration?', *Research Policy,* 26 (1), 1-18.
Kostoff, Ronald N (1998), 'The use and misuse of citation analysis in research evaluation', *Scientometrics,* 43 (1), 27-43.
--- (2002), 'Citation analysis of research performer quality', *Scientometrics,* 53 (1), 49-71.
Kyvik, S. and Olsen, T.B. (2008), 'Does the aging of tenured academic staff affect the research performance of universities?', *Scientometrics,* 76 (3), 439-55.
Larivière, Vincent and Gingras, Yves (2010), 'The impact factor's Matthew Effect: A natural experiment in bibliometrics', *Journal of the American Society for Information Science and Technology,* 61 (2), 424-27.
Leahey, E. (2006), 'Gender Differences in Productivity Research Specialization as a Missing Link', *Gender & Society,* 20 (6), 754-80.
Long, J.S. (1990), 'The origins of sex differences in science', *Social Forces,* 68 (4), 1297-316.
--- (1992), 'Measures of sex differences in scientific productivity', *Social Forces,* 71 (1), 159-78.
Manjarrés-Henríquez, L., Gutiérrez-Gracia, A., and Vega-Jurado, J. (2008), 'Coexistence of university-industry relations and academic research: Barrier to or incentive for scientific productivity', *Scientometrics,* 76 (3), 561-76.
Melin, G. (1996), 'The networking university', *Scientometrics,* 35 (1), 15-31.
Merton, Robert K (1968), 'The Matthew effect in science', *Science,* 159 (3810), 56-63.
Moed, H.F. (2005), *Citation analysis in research evaluation* (9: Springer).
Moed, H.F., et al. (1985), 'The use of bibliometric data for the measurement of university research performance', *Research Policy,* 14 (3), 131-49.
Moed, Henk F (2006), *Citation analysis in research evaluation* (9: Springer).
--- (2009), 'New developments in the use of citation analysis in research evaluation', *Archivum immunologiae et therapiae experimentalis,* 57 (1), 13-18.
Narin, Francis (1976), *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity* (Computer Horizons Washington, D. C).
Norris, Michael and Oppenheim, Charles (2003), 'Citation counts and the Research Assessment Exercise V: Archaeology and the 2001 RAE', *Journal of Documentation,* 59 (6), 709-30.
Pavitt, K. (2000), 'Why European Union funding of academic research should be increased: a radical proposal', *Science and Public Policy,* 27 (6), 455-60.
--- (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 (3), 761-79.
Phelan, TJ (1999), 'A compendium of issues for citation analysis', *Scientometrics,* 45 (1), 117-36.
Salter, A.J. and Martin, B.R. (2001), 'The economic benefits of publicly funded basic research: a critical review', *Research Policy,* 30 (3), 509-32.
Sammarco, Paul W (2008), 'Journal visibility, self-citation, and reference limits: influences on Impact Factor and author performance review', *Ethics in Science and Environmental Politics,* 8, 121-25.
Seglen, P.O. (1997a), 'Why the impact factor of journals should not be used for evaluating research', *Bmj,* 314 (7079), 497.

Seglen, Per O (1997b), 'Citations and journal impact factors: questionable indicators of research quality', *Allergy,* 52 (11), 1050-56.

Stegmann, Johannes and Grohmann, Guenter (2001), 'Citation rates, knowledge export and international visibility of dermatology journals listed and not listed in theJournal Citation Reports', *Scientometrics,* 50 (3), 483-502.

Sugimoto, Cassidy R, et al. (2013), 'Global gender disparities in science', *Nature,* 504 (7479), 211-13.