# THE UNEQUAL BENEFITS OF ACADEMIC PATENTING FOR SCIENCE AND ENGINEERING RESEARCH

Mario Calderini*, Chiara Franzoni*, Andrea Vezzulli**

*DISPEA, Polytechnic University of Turin, Corso Duca degli Abruzzi 24b, 10121 Torino, Italy
**CESPRI, Bocconi University, Via Sarfatti, 25, 20136 Milano, Italy

**Abstract**

We analyzed the scientific productivity of a sample of academic scientists that contribute to the field of Materials Science in the post-patenting period, by means of several econometric techniques suitable to treat unobserved heterogeneity, excess zeros and incidental truncation. Although patents do not alter the track of publications in the overall sample, we show this effect to be generated by two opposite effects: Materials Engineers increase their publications after patenting, whereas Materials Chemists experience a decrease. Besides, Materials Engineers who were academic inventors have a higher impact factor than their non-inventors colleagues, although the positive effect tends to vanish both for very basic publications and for serial inventions. Finally, a clearly negative effect is registered when we consider only very basic publications made by Materials Chemists. We interpret our findings as depending on different epistemologies of scientific and engineering research and discuss the implications for both university managers and policy makers.

Keywords: academic patenting; science and engineering research; technology transfer; science policy; university management

JEL: O31; O33, I23; I28

## 1. Introduction

In recent years academic institutions have become increasingly involved with technology transfer and academic patenting. During the last two decades, institutional and legal changes similar to the 1980 USA Bahy-Dole Act have been debated and approved in many countries, mirroring the political consensus behind the new role given to universities of becoming professional traders of technologies and of applications for industry1.

Current US figures are impressive: according to the AUTM survey (AUTM, 2005), in 2005, US member universities have filed 15,115 patent applications (grown at an average rate of 55% year by year in the last 5 years), which resulted in 3,278 patents granted. They created 628 new companies (nearly two every day), held more than 28 thousands active licenses, nearly 5 thousands of which started in 2005, and were responsible for the launch of 527 new products on the market.

Although aggregate statistics are not available for European countries and generalization is made complicated by several differences in regulations and practices, European universities have experienced a similar trend, although perhaps reduced in magnitude (Geuna and Nesta, 2003; see also ProTon, 2007).

From both sides of the Atlantic, this overall trend has initially raised different feelings. On the one hand, many members of the academic community, especially the senior professors, have showed considerable reticence to immoderate commercial openness, for fear that the pressures of market might be at odds with freedom of science, and raised concerns that education of students might also be disregarded under the burden of more lucrative activities (see for instance Lee, 1996). On the other hand, academic patenting and technology transfer in general were seen as a viable way of easing communication between science and market, unlock science from an ivory-tower position that allowed considerable independence, but limited impact on social and economic wealth, and favored mutually beneficial exchange of ideas and competencies.

Along the years, the initial skepticism eventually faded and it is now common place for universities to advertise their linkages with industrial partners and their commercial activities as a signal of good quality and prestige of the institution. Among the reasons behind this shift were several pieces of evidence presented in the latest years, which reveal that, in fact, top-rated universities for both research and education were among the best performers for number of patents issued, and license income (Henderson et al., 1998; Milken Institute, 2006). Besides, several preliminary analyses conducted at the level of individual scientists, rather than institutions, have confirmed that patenting does not jeopardize publications and is even likely to increase productivity in the post-

[1] After the 1980 Bahy-Dole Act, similar lows that assign de jure IPRs from publicly-funded research to the principal-investigator's institutions were approved in Canada, UK, and nearly all western European countries, except from Italy, Finland and Sweden (OECD, 2003).

patenting period. While a clear interpretation of this supposed "complemetarity effect" is still debated, the evidence provided so far suffered several limitations. First, the lack of good metrics to measure the character of research (quality, scope, decay, etc.) does not allow to rule out the existence of unobserved effects, such as, for instance, the advocated deterioration in the generality and scope of research. Second, empirical analyses are disproportionately based on Life Sciences and lack an appreciation of the differences related to aim and scope of the sub-field to which a scientist contributes. Our work aims at contributing to the latter point, by offering an empirical analysis of the post-patenting effect on productivity and character of research, based on a sample of researchers in Materials Science.

We use an unbalanced panel of 1276 Italian scientists working in the field of Science of Materials of which we gathered complete data on scientific activity (number, level of basicness ad impact of publications), patent applications, subfield and other personal data (gender, seniority, affiliation). Our time-span window covers since the conventional entrance in the academic career (23rd year for all) until the end of 2003. The effect of post-patenting on scientific performance is studied by means of several econometric techniques suitable to treat different problems that usually affect this kinds of analysis, given the characteristics of the experimental design (endogenous selection into patenting activity, unobserved heterogeneity) and the features of the dependent variable (positive integers with excess zeros, incidental truncation). In particular, given lack of good instrumental variables in this setting, we adopted the Inverse Probability of Treatment Weighted (IPTW) approach (Azulay et al., 2006; Breschi et al. 20067), in order to address the problem of endogenous selection into patenting

By separating our Materials scientists among the sub-samples of Materials Chemists and Materials Engineers, we show that Engineers get benefits form patenting, while Chemists do not. We further disentangled our analysis by considering both qualitative effects and serial vs. occasional patenting.

The paper is organized as follows. Section §2 describes the terms of the debate upon the effect of academic patenting on scientific publication, presents the results of the empirical investigations provided so far and states the research question addressed in following sections. In Section §3 we describe the dataset, in Section §4 we present the indicators and the models used to address unobserved heterogeneity, excess zeros and incidental truncation affecting this kind of analyses. Section §4 presents the models and results. Section §5 summarizes the findings and draws conclusion and policy implications.

## 2. The state-of-art debate and our research hypothesis

University patents have increased drastically in the last 25 years. In the USA, the post-Bahy Dole Act period was characterized by a steep surge of university patents (steeper than the surge of corporate patents) (Henderson et al., 1998). From the one side, this trend was shown to be largely supported by increased opportunities to patent in biotechnology and ICT sectors that opened-up new areas of patentability[2] and convinced researchers of the opportunity to patent their discoveries (Mowery et al., 2001; Hall, 2005). From the other side, internal universities policies provoked an increasing number of disclosures that resulted in more patents issued and licensed (Thursby and Kemp, 2002).

With regard to patent quality, empirical investigations based on both USPTO and EPO data have indicated that universities and public research organizations (PROs) produced patents of higher quality, when compared to the private sector patents (Henderson et al., 1998; Mowery and Ziedonis, 2002; Bacchiocchi and Montobbio, 2006). The methodology developed to draw this conclusions is based on citation analyses of patent data developed in the later 90s, that assess importance as the number of forward citations received by a patent, and generality as the number of different patent classes from which a patent receives citations (Trajtenberg et al., 1997). This difference was shown to be caused mostly by patents produced by USA universities in chemical, drug and medical classes, while European and Japan academic patents do not substantially differ from corporate patents (Bacchiocchi and Montobbio, 2006).

Comparisons of the pre and post Bahy-Dole Act figures have shown that, whereas the top performing institutions per number of patents were concentrated among the top research universities both before and after the 1980, smaller and newcomer institutions (i.e. universities that never patented before 1980) started to produce patents of lower importance and generality later on (Henderson et al., 1998)[3].

Because the higher quality of university patents was seen as depending on the wider scope and longer decay rate of academic inventions versus firms inventions, this evidence provoked concern that commercialization of science was associated to a deterioration of the breadth and basicness of academic research.

The debate upon the independence of science from markets to protect the natural sources of curiosity of scientists and their long term benefits has deep roots in science policy (for a complete

[2] Among the key determinants were the US supreme courts decisions to allow patentability of genetically-modified organisms (Diamond vs. Chakrabarthi, 1980), software codes (Diamond vs. Dieh, 1981) and business methods (State Street & ATT vs. Excel, 1998). See Hall (2005).

[3] Mowery and Ziedonis (2002) find a non-decrease in generality and importance in a study limited to Stanford, Berkeley and Columbia universities. Although Columbia started patenting only after 1980, they found no evidence of lower quality, such as those found by Henderson, Jaffe and Trajtenberg (1998).

discussion see Nelson, 2004). The terms of the debate, as discussed by many contributions of the last years, can be briefly sketched as follows: academic institutions traditionally received funds from public and non-profit sources in exchange of scientific discoveries, later to become technological change, economic development and eventually increased social wealth. Discoveries were disseminated through publications in open-science, which, in principle, ensured free access to everybody, but at the same time, created an appropriability problem (Arrow, 1962). Policies such as the Bahy-Dole Act indeed rely on the assumption that direct firm investments in technologies disclosed by universities were going unexploited because of lack of incentives to bear the costs of development, when knowledge was set open to everybody's use. Consequently, potentially valuable applications were left on book's pages because of lack of incentives for firms to take them up. In contrast, patents would offer the advantage of temporary monopoly concession, while at the same time ensuring some kind of disclosure.

The concern of those that see patenting and publishing as rival activities is based on a number of arguments. To begin with, the problem of going beyond simple open-science dissemination, is that open science dissemination, despite several limitations and pitfalls[4], also incorporates a number of unwritten rules regulating the functioning of the scientific community in a certain desirable fashion, including a) incentives to prompt disclosure, b) a mechanism for validation strictly internal to the community of peers, and c) a distribution of rewards based on scientific merits (Dasgupta and David, 1987).

Overcoming open science publications as the main scientist's goal -it is warned- will imply diverting from the previous rules, with potentially negative consequences, especially in the long term. With regard to point a), the argument goes that, while it is always in a scientist's interest to disseminate as much as possible his or her own publications, it is generally the interest of a patent holder that patents stay unnoticed, even after publication ceases to jeopardize novelty. Hence, a first matter of concern is that patenting may refrain scientists to publish or at least slow down dissemination thus reducing the pace of knowledge advance[5]. Point b) relates to the fact that patent, unlike publications, are not being discussed and validated by the scientific community, because patent examiners are called to check novelty and replicability, rather than validity, importance and scientific method, and this might jeopardize the quality and reliability of the

[4] By way of example, well documented effects of cumulative advantages (such as the "Matthew Effect"), causing unequal returns of effort and merit are attributed to the fact that articles are much more numerous than what a scholar can read, which gives known names more chances to be picked-up (Merton, 1968). The so-called Plank's Principle (Levin et al., 1995) also reinforces the idea that attribution of scientific merits is affected by political influence (Hagstrom, 1965).

[5] In principle, publications are delayed as a minimum until the filing of a patent, in those systems such as EPO and WIPO that do not accept the "grace period" exception. In practice, this is likely to occur also in countries that recognize the grace period if the inventor wants to keep the option to extent patents beyond the national borders later on.

knowledge disclosed (Myer, 1995). Finally, the concern raised under point c) is based on the fact that science and market differ in their appraisal of fundamental contributions. Hence market payoffs, such as those associated to a successful patent, may divert scientists from their traditional goals of pure research and teaching. As Merton was first to articulate, the fact that scientists produce knowledge that is diffused as a public good does not exempt them from chasing their (private) benefits resulting from discovery (Stephan, 2004). The strength of the scientific community was ultimately based on providing a regulating mechanism (alternative to market), to distribute merits and recognition in a way that fosters the production of fundamental knowledge, for which market alone offers little incentives. Allowing commercialization of scientific results hence looked to many like discarding this strength from its very basic foundations.

In addition to the previous, concerns were raised with regard to the problem known as the "anti-commons effect", which arises when some relevant resource, such as a research method or material, is property of many different owners, having rights to exclude others from its use. In this case, multiple and conflicting ownership and transaction costs may cause underuse of the resource, since no one person can use the whole (Heller and Eisenberg, 1998)[6].

Several empirical investigations have been recently conducted to test the effect of patenting on subsequent scientific activity, based on both comparisons of institutions and individuals and on both cross-sections and longitudinal data. So far, two main findings have emerged quite consistently. First, by looking cross-sectionally at the group of scientists that ever patented vs. those that didn't, all studies show that the academic inventors, despite representing a small proportion of the population (10-15%), are disproportionately concentrated among the most productive in research. Fabrizio and Di Minin (2005) find a positive correlation between actual and lagged numbers of papers and patents, in a sample of 150 inventors and 150 controls. Breschi, Lissoni and Montobbio (2007) show that the academic inventors published on average one paper more than a matched-pair sample of researchers that never patented and that this difference is higher for serial inventors. Stephan, Black, Sumell and Gurmu (2007) run a zero-inflated negative binomial regression in a large sample of doctorate-recipients and find that patent counts and publication counts are positively related after controlling for field, seniority and other institutional and job characteristics. Carayol (2007) finds similar results for a sample of scientists at Louis Pasteur University. Among the institution-level analyses, Van Looy et al. (2004) find that

[6] In the USA increasing concern upon the availability of research instruments was raised after the Madey vs. Duke decision (*Madey v. Duke U.*, F.Supp. 2d 420 (M.D.N.C. 2001)), that substantially reduced the experimental use exception, i.e. the right of a third party to "use a patented invention without inventor authorization for purposes of philosophical experimentation, to satisfy curiosity, or ascertain functionality of the patent" (*Whittemore v. Cutter*, 29 F.Cas. 1121). The reason to reject the experimental exception right raised by Duke was indeed that the university was no longer recognized as having a non-profit, educational mission (Lowry, 2005).

researchers who were systematically involved in contract research published more than the colleagues in the control sample and argue for a complementarity of research and application[7]. Although longitudinal evidence is still preliminary and suffered of several problems in the treatment of data, available studies to date hinted that patents might not only be invented by the most productive in research, but can also be associated to an increase of publications (Azoulay et al, 2006; Breschi et al, 2006; Fabrizio and Di Minin, 2005)[8]. The effect seems non-negligible, in terms of magnitude, and occurs either in the year of the invention, or in the following one or two years, which by and large corroborated the idea that patenting and publishing may be complementary, mutually sustaining, activities (Azoulay et al., 2006; Breschi et al. 2005; Fabrizio and Di Minin, 2004).

With regard to the spectrum of university activities, the previous investigations were based upon various scientific fields of S&E (Chemistry, Physics, Life Sciences, Computer Sciences, Mechanical and Electronic Engineering), although, at present, Life Sciences happens to be the most widely analyzed field, which advices some cautions in the generalization of results to other disciplines. With regard to the anti-commons hypotheses, a survey of Walsh, Cho and Cohen conducted among biomedical scientists revealed that scientists did not claim to suffer any strong change in the attitude to share materials and methods (Walsh et al., 2006). Besides, evidence that the citations received from papers associated to patented materials and methods decreased after the patent was being issued were found by Murray and Stern (2007), in a sample of patent-paper pairs in biotechnologies.

The empirical evidence discussed so far has made a very impressive job in putting forth new issues and discarding unsupported preconceptions. At the same time, several issues are left open and deserve further investigation. Our idea is that two areas of improvement demand specific attention: first, very few analyses encompass assessments of the character of the knowledge disclosed in the post-patenting period[9], and this mirrors a fundamental paucity of both metrics and theoretical concepts to characterize research beyond sheer productivity. Second, little consideration has been devoted to how differences in the nature and scope of the various fields of S&E to which a researcher contributes might affect the relation between his/her scientific and inventive work. Among the empirical analyses mentioned before there was no attempt to separate

[7] They compare the performances of a unit departments rather than of individuals. The departments were part of the contract research units of Catholic University of Leuven (BE) and the department controls were made of (pure research) faculties of the same university in the same research fields.

[8] An exception is the study by Agrawal and Henderson (2002), that finds no statistically significant effect in a sample of MIT scholars.

[9] To the best of authors' knowledge, only Azoulay, Ding and Stuart (2006) and Breschi, Lissoni and Montobbio (2006) use metrics specifically aiming at measuring qualitative features of research. Agrawal and Henderson (2002) and Fabrizio and Di Minin (2005) make use of citations counts, which however depend on the article age, as well as on the patterns of citations that in turn are journal-specific. As such, their use as an indicator of quality is debated.

the effect according to field or subfield, in part due to the small numbers of patents found, that did not advice further breakdowns.

Our paper is especially aimed at addressing the latter issue. We expect that no unique impact is linking the inventive activity of a scientist with the post-patenting performance, but rather that this relation is at least partially field-dependent. The starting point to build our hypothesis would be to consider that not all disciplines stand in the same relation and earn equal benefits from serving practical ends. A first rough, but quite clear-cut distinction can be made between Hard Science and Engineering.

Although Epistemology of Engineering is still regrettably quite-undeveloped, a key difference of doing research in engineering, as opposed to hard science, is that, whereas science is aimed at the understanding of phenomena, and somehow sees technology as instrumental to that end, engineering is in its fundamental and epistemological essence a science applied in scope, i.e. a discipline that addresses and aims to solve problems of industrial (practical) relevance, by means of a rigorous scientific method (Vincenti, 1990). By "applied in scope" we do not mean to suggest that engineering is a an applied science, in the sense of being deductive, i.e. a discipline that applies findings of a hierarchically-dominant scientific domain into practice, such as conventional wisdom suggests[10]. Rather, we mean that the application to solving a practical problem is the engine that moves the investigation.

For instance Walter G. Vincenti says:

> *I have never attempted to design an airplane in my entire career as a research engineer (although I participated in planning and designing large aeronautical research facilities). The atmosphere in which I worked, however, and the knowledge I helped produce, were conditioned by the needs of airplane designers who visited our laboratory. My collegues and I were keenly and continuously aware of the practical purposes we served.* [Vincenti, 1990:7]

In a survey of university and firm collaboration, Mansfield (1995) found that university scientists were very frequently conducting academic research on problems and ideas that they became aware of while doing industrial consulting. In the interviews (a large proportion of interviewed scientists in fact happened to be engineers), researchers reported that the contribution of firms and users could vary from being very marginal up to being fundamental in indicating the problems and the direction of research.

Following this line of reasoning, it is hence consequent that working on practical problems such as those posed by inventing a new functional tool can be in principle more fertile of ideas for engineering than for science. Our hypothesis is hence that engineers would be more likely to benefit from working on practical problems than their chemists colleagues.

---

[10] An epistemological discussion of the argument would exceed the purposes of the present work. See Walter G. Vincenti (1990) and Edwin T. Layton (1974) for a more comprehensive discussion.

## 3. Sample and Data

The database used for the present study was based on a list of scientists members of an Italian association for research in Materials Science, called INSTM (Consortium of Italian Universities for Science and Technology of Materials). The association gathered, at the end of 2003, over 1660 researchers, belonging to 42 Italian universities and public research centers, which virtually represent all universities and public research units working in the field of Science of Materials throughout Italy.

According to the Carnegie Mellon Survey, academic research in Materials Science is perceived by firms among those that contribute more substantially to industrial R&D (Cohen et al., 2002). Given that admittance of researchers to INSTM association is individual and voluntary, and requires paying an annual membership fee, scientists are self-selected as those working in the area of Materials Science. We took all members at the end of 2003 that were born in 1954 or later, which resulted in a final list of 1323 names and eliminated the lab engineers and technician, which leaves us with a list of 1276 names. Materials Science is a considerably homogeneous field, and its scientific community gathers contributions from several mother disciplines: mainly Chemistry, Engineering, Physics, and, more rarely, Mineralogy and Geology. Our sample of scientists mirrors this organization: observed scientists resulted to be distributed in the following proportions: 919 Materials Chemists (72%), 309 Materials Engineers (34%), 35 Materials Physics (3%), plus 12 scientists (1%) from several other sub-fields[11].

Our sampled scientists were in 2003 tenured professors, as well as untenured researchers, PhD students and research assistants, thus providing a good representation of the variety of roles and types of professionals working for the Italian public research system. To the best of our knowledge, we are not aware of any selection bias affecting stratification of our sample.

For each of the 1247 names we collected all papers published in open science journals (as listed by ISI Science Citation Index) and all USPTO or EPO patent issued (from Delphion Thomson)12. See patent descriptive statistics in Table 1.

We take as a conventional starting observation time ($t_{i0}$) the year in which the scientist was 23 (which is the minimum age to obtain an MS degree in the Italian education system) and collected all information from that year to the end of 2001. Publication lags in Materials Science range from four weeks to six months; therefore, we can take the publication year as a proxy of the discovery date. Similarly, we take patent priority date as the proxy of the invention date. Given that the ISI

[11] Based on classifications of the Italian Ministry of Research (http://sito.cineca.it/murst-daus/docenti/docenti.shtml)

[12] Extensions of patents from EPO to WIPO or vice versa were checked and duplicates were eliminated (only the original patent was kept).

database allows only querying for full surname, plus name initials of the author, the case of including homonyms is highly frequent. To cope with this problem, we filtered the resulting list of papers on the basis of coherence of scientific fields (Materials Science) of the reviews, according to the ISI Journal of Citation Report (JCR) taxonomy (multidisciplinary fields included).

We apprised basic/fundamental vs. applied orientation of research by means of the IpIQ ranking of journal Level, which is an indicator expressed in a 1 to 4 rank, where "very basic, untargeted research" is set equal to Level 4 (Narin et alt. 1967).

In order to appraise the quality of the scientific papers, we used the Impact Factor (IF) of the scientific journals where the articles were published[13] (for general information on the index and on citation-based indicators see Diamond, 1986; Narin and Hamilton, 1996). Usage of the journal's Impact Factor as a proxy of quality of the published article equals to making the assumption that good journals only publish good papers and vice versa.

Table 2 (a and b) provides a complete explanation and summary statistics of the dataset variables.

**4. Methods and results.**

In this section we study the effect of post-patenting on the scientific performance by means of a number of econometric techniques in order to account for the multiple problems that usually affect this kind of analysis, given the characteristics of the experimental design and the nature of the dependent variable under study in each of the settings considered.

In general, the estimation of the causal effect of a treatment (patenting) on a variable of outcome (quantity, basicness and quality of scientific production) can be difficult in non-controlled studies for the presence of confounding variables (or confounders) which both affect the outcome of interest and the probability of being treated.

We then adopt the Inverse Probability of Treatment Weighted (IPTW) approach (Azulay et al., 2007; Breschi et al. 2007), a method that is widely accepted in biostatistics for estimating Average Treatment Effects (ATE) in observational studies (Robins et al., 2000; Hernan et al. 2001), which address the problem of endogenous selection into treatment in a similar way to others propensity-scores matching techniques (Rosembaum and Rubin, 1983).

This method relies on the crucial assumption that the selection into treatment is based on observables variables and that the modeling structure of selection is correctly specified (see Azoulay at al. 2007 for details). It nonetheless brings the considerable advantage of non requiring exclusion restrictions for identification, unlike in the Instrumental Variable approach, so that there is no need of instruments (which are not easy to find in this context).

[13] Impact Factor figures were taken from the 2002 edition of JCR.

With IPTW the role of confounders is neutralized by weighting each observation with its (stabilized) inverse-probability of treatment and it can be interpreted as the inverse of a subject's conditional probability of receiving her treatment history up to time *t*, given past treatment history and others "prognostic" factors.

We implement this procedure by estimating a logit model on the probability of applying for a patent for the first time. Logit formulation and estimates are reported in Table 10-11.

Weights obtained from the logit analysis will then be used to weight each observation, when regressing the outcome variable of interest $Y$ on the set of covariates X and on the treatment variables Z.

The set of covariates X will include $\text{SENIORITY}_{it}$, which measures the number of years a scientists had spent in academia up to year $t$ , $\text{EXPTTOMA}_{it}$, which proxies the experience of the institution in patenting, and hence captures environmental effects (measured as the total number of patents granted to the institution in the previous 5 years), and a dummy variable for gender ($\text{GENDER}_i$). The set of treatment variables Z includes $\text{PAT}_{it}$ (flow treatment indicator), expressed as the number of patents granted to scientist *i* at time *t* (where *t* is the year of priority of the first application for patents that has scientist *i* among the inventors), $\text{POSTPAT}_{it}$, (regime treatment indicator), a dummy equal to 1 if scientist *i* has at least one patent up to year *t*, and $\text{CUMPAT}_{it}$ (cumulative treatment indicator) as the total number of patents granted to scientist *i* up to year *t*.

### 4.1 Scientific productivity: quantity of scientific production.

The first question we want to investigate is the effect of patenting on the quantity of scientific production measured by the number of articles published in a year.

We first consider the raw number of (authored and co-authored) scientific papers published by scientist *i* in year *t* ($\text{PUBL\_M}_{it}$). As visible from Since this is a count variable showing a disproportional amount of zeroes (more than 40%) (see Table 2) the natural choice for modeling it is a Zero Inflated Negative Binomial (ZINB) model. This model entails two different regressions because it assumes two different processes governing the dependent variable: one for the inflation part (zero outcome) and the other for the count outcome (without extra-zeroes). Moreover it allows for unobserved heterogeneity among subjects by assuming individual gamma distributed random effects. This model is estimated via iterative Maximum Likelihood techniques (Wooldridge, 2002) with robust standard errors clustered across subjects. Table 13 shows the estimating results. Looking at the whole sample no statistically significant effect is exerted by either the lagged patent regime variable ($\text{POSTPAT(-1)}_{it}$) and by the cumulative number of past patents ($\text{CUMPAT(-1)}_{it}$).

However, since the average number of annual scientific publications depends largely on the researcher's scientific field (see Table 6), we run separate regressions for the two sub-samples (ENGINEERS and CHEMISTS) that offer a fairly numerous number of observations. The results of this estimates are reported in Table 13. After the first patent, engineers tend to have a greater yearly number of publications than non-patenters (although the positive estimate on $POSTPAT_{it-1}$ is significant only at 10% level). Conversely, for chemists we find a negative impact, significant at 10%. Besides, we also see a positive effect of past cumulated patents (CUMPAT) on articles productivity, which tends to overwhelm the former negative effect after the 3-4$^{th}$ patent granted, although this counter-effect of CUMPAT is relevant only for a small proportion of the observations (for instance, in 2001, only the 2% of chemists had more than 3 patents granted, as shown in Table 9).

The estimated sign of the controls are quite simply explained: articles productivity first increases with seniority ($SENIORITY_{it}$) and eventually declines at a later stage of career ($SENIORITYsq_{it}$). Men tend to have a higher productivity than their female colleagues ($GENDER_{it}$), while the overall number of past patents owned by the institution of affiliation ($EXPTTOMA_{it}$), which captures the institutional/environment effect, has a positive impact on a scientist's productivity. The calendar-time dummies ($DUMYEAR*_{it}$) show that, on average, publications have increased in recent years.

We then consider a different measure of scientific productivity which takes into account co-authorships, i.e. shared articles, and build an alternative weighted indicator of publications by dividing the number of yearly publications by the average number of authors ($WPUBL\_M_{it}$). Because this new variable is no longer a positive integer, we are free to use a standard linear model. We partially recover its skewedness (due to the excess of zeroes) by means of the following transformation $LWPUBL\_M_{it} = \log(WPUBL\_M_{it} + 1)$. For the sake of comparison, we also estimate a similar model for the original un-weighted variable $LPUBL\_M_{it} = \log(PUBL\_M_{it} + 1)$. This "linearization" of the former model has the advantage of allowing the application of a linear Fixed Effect (FE) estimation method which is more robust (although less efficient) than ML methods. The results are reported in Table 15-16. The linear model with the weighted dependent variable basically confirms the findings of the ZINB model, whereas the comparison model with the un-weighted dependent variable confirms the findings only for the engineers, while for the chemists the estimated relation of publications and patents is not statistically significant.

**4.2 Scientific productivity: basicness of scientific production.**

The second question we want to investigate is whether patenting hampers or boosts basic scientific research. We follow an approach similar to the previous subsection, but take as dependent variables the raw number of scientific papers authored (or co-authored) by scientist *i* in year *t*, which resulted to be ranked as "very basic" (level 4) in the IpIQ classification ($PUBLBAS4_{it}$), its log-linear transformation ($LPUBLBAS4_{it}$) and its author-weighted version ($LWPUBLBAS4_{it}$) which takes into account co-authorship. Again the analysis is further disentangled between engineers and chemists and models are estimated by ZINB-ML (Table 14) and standard linear OLS-FE (Tables 17-18) techniques respectively. The only notable and statistically significant effect relates to the sub-sample of chemists. While in fact patenting does not seem to impact the basic scientific output of engineers, for chemists we find in both the equations a negative impact of patenting on the basic scientific output.

### 4.3 Scientific productivity: quality of scientific production.

The final question of our analysis concerns the effect of patenting on the quality of a scientist's research. To answer this question, we first have to overcome the problem of finding an appropriate measure of scientific quality. Several approaches has been proposed by the scholars, the most common of which is to capture the quality of a researcher's output by counting the total number of citations received (Agrawal and Henderson, 2002; Breschi et al. 2006; Fabrizio and Di Minin, 2007). However this method is not immune from drawbacks, since it can be dramatically affected by the specific characteristics of the scientific field considered such as different publications rates, different cross-citing practices, different citation trajectories along time and so on.

Azoulay et al. (2007) tried to overcome such drawbacks by constructing two alternative metrics: the first is based on the proportion of publications in which the researcher appears in first and last position of the authors' list. The second (also adopted by Calderini et al., 2007) is based on the average journal impact factor (IF) of the articles published in a given year. We follow the latter approach, although in a slightly different way.

Given the different distribution of the average IF among journals of different scientific field (as outlined by Table 3), in addition to run separate regression according to the researcher's main scientific field (as in the previous sub-sections), we also standardize the IF score assigned to each publication as follows: $STDIFAC = [(IF_{it}\text{-mean}(IF))/\text{std. dev}(IF)]$, where the mean and the standard deviation of IF are calculated with respect to the journal scientific field on which the article appeared. Thus $STDIFAC_{it}$ is the average of the standardized journal impact factor index (STDIFAC) for the articles published in year *t* by scientist *i*.

This dependent variable is clearly affected by incidental truncation, since it the Impact Factor is only observable when the researcher has at least one publication in year $t$ (i.e. if the dummy variable $DUMPUBL_{it}$ is equal to 1). We treat truncation by means of a Heckman selection equation (based on $DUMPUBL_{it}$ as dependent variable) and a truncated regression (based on $STDIFAC_{it}$ as dependent variable) that are estimated simultaneously along with the variance of the error component $u_1$ of the outcome equation $\sigma$ (the variance of the error component in the selection $u_2$ is set to 1) and the correlation $\varrho$ between $u_1$ and $u_2$ (see Heckman, 1979 and Amemiya, 1985 for details). Results are reported in Table 19[14].

In the estimates based on all observation we find that researchers, after the first patent granted, tend to publish in journals with (average) higher IF scores than non-patenters, which mirrors an increase in their ability to publish on higher-impact journals.

However, for engineers, we find that the positive and significant coefficient associated to ($DUMPAT_{it}$) is counterbalanced by the negative and significant coefficient of the cumulative number of patents granted ($CUMPAT_{it}$), which suggests that this increase in performance comes at a decreasing marginal rate, and would eventually be neutralized and overwhelmed after about the 3rd patent granted.

## 5. Comments of results and conclusions

In recent years academic institutions have become increasingly involved with technology transfer and academic patenting. The reasons behind this institutional and managerial shifts of universities have been largely discussed by the scholars of science and innovation, which have also debated extensively the potential benefits and risks. From a broader perspective, science is less and less seen as an instrument of political competition between nations worldwide, as during the second half of the XX Century, and is more and more perceived as a means to foster the competitiveness and wealth of the economies at a national and local level. To ensure a high degree of communication and exchange of knowledge between science and market is as important as having a first-class research to support local firm's competitiveness.

In the European Union, a considerable number of policy actions of the last decade have been addressed to increasing the dissemination of fundamental research results, based on the (true or mistaken) assumption that the quality of basic science itself was satisfying, while much of the potential of new knowledge made available got unexploited and never left the labs.

[14] We included individual dummy variables (estimates not reported) in the outcome equation to control for potential sources of unobserved heterogeneity (gender is omitted to avoid multicollinearity, given its non-time varying nature).

All those policies have been based on the assumption that collaboration between science and market does not significantly jeopardize the ability of scientists to do fundamental research, disclose their achievements on open-science journals and chose their own topics of inquiry independently. Scientists –it is believed- can in principle patent and *sell* IPRs as a by-product of their normal activity, as much as they *sell* teaching and education services to their colleges.
This assumption has proved to hold in preliminary empirical evidence (Agrawal and Henderson, 2002), which additionally highlighted an unexpected boosting effect on publications in the post-patenting period (Azoulay et al., 2006; Fabrizio and Di Minin, 2005; Breschi et al., 2006). However, we claimed that at least two important pieces of information are missing to enlighten those findings: 1) qualitative assessments of the knowledge disclosed in the post-patenting period, and 2) an appreciation of the differences of the aim and scope of the sub-field to which a scientist contributes. Besides, the state-of-art evidence is disproportionately based on the field of Life Sciences, which in the last decades experienced a pretty unique contamination of private and public R&D (Mowery and Ziedonis, 2002).
This paper has contributed to both points, by offering an empirical investigation based on a large sample of scientists in Materials Science, an academic discipline deemed of key importance by industry (Cohen et al., 2002) and at the same time gathering contributions from Chemistry, Physics and Engineering. Indicators of quality of publications used include impact factor and level of the journal.

**Figure 1. Summary of post-patent effects: sign and significance.**

| | | QUANTITY | | | BASICNESS | | | QUALITY |
|---|---|---|---|---|---|---|---|---|
| | | IPTW - ZINB | IPTW - OLS_FE | OLS_FE weighted | IPTW - ZINB | IPTW - OLS_FE | OLS_FE weighted | HECK_ML |
| ALL | postpat | + | + *** | + | − | − | − | + ** |
| | cumpat | + | − | + | − | − | − | − |
| ENGINEERS | postpat | + | + *** | + ** | − | − | − | + ** |
| | cumpat | − | − | − | − | − | + | − *** |
| CHEMISTS | postpat | − ** | + | − ** | − ** | − ** | − *** | + |
| | cumpat | + | + | + | − | + | + | + |

The models we run were suitable to treat unobserved heterogeneity, excess zeros and incidental truncation. A summary of the effects estimated through the diverse models is presented in Figure 1. Our results on the overall sample suggest that patenting does not substantially alter a scientist's publication track. However, when we separate the sample by sub-fields groups and run separate analyses for Engineers and Chemists, we found that Materials Engineers experience an increase of quantity and quality of publications after the invention, while Chemists of Materials might experience a decline in quantity of overall publications (not supported by all estimating techniques) and quantity of very basic level publications.

We showed that the scientists that were contributing to the Engineering-side of Material Science research were experiencing improved performances when working on industrial applications, whereas this was not the case for those that were contributing to Materials Science as Chemists. We interpreted our results to be depending on the different epistemology of science and engineering. Although Epistemology of Engineering is still regrettably undeveloped, a key difference of doing research in engineering, as opposed to hard science, is that, whereas science is aimed at the understanding of phenomena, and somehow sees technology as instrumental to that end, engineering is in its fundamental and epistemological essence a science applied in scope, i.e. a discipline that addresses and aims to solve problems of industrial (practical) relevance, by means of a rigorous scientific method (Vincenti, 1990). Engineering in fact is inherently scoped to problems of industrial relevance, while this is not necessarily the case of disciplines, such as Physics and Chemistry, aimed at the general understanding of processes.

In principle, an alternative explanation of the findings could be that scientists and engineers have a different attitude or policies in the disclosure of research associated to patenting, for instance that Chemists might overlook publishing in open science the content of patented research. We however consider the latter explanation less plausible.

Our results are based on Italian academia and on the field of Materials Science and allow therefore limited generalization. If similar results would be confirmed by other studies, several important implications should be derived for university managers and policy makers alike. For instance frequent and intense collaborations of faculties and firms, share of research project and joint funding should be more actively encouraged in the engineering schools, than in non-engineering departments. In principle, it is plausible that similar results of different returns from industrial inspiration to different subfield-disciplines would be found in other subject domains, such as mathematics and computing, or biology and biotechnology, which might derive unequal inspiration from scoping research to industrial problems.

## 6. References


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

[2] Aitchison J., Aitken C.G.G., 1976. "Multivariate binary discrimination by kernel method", Biometrika, 63(3):413-420.

[3] Amemiya, T., 1985. "Advanced Econometrics", Cambridge, MA: Harvard University Press.
[4] Arrow K., 1962. "Economic Welfare and the Allocation of Resources for Invention", in: Nelson R.R. (ed.) "The Rate and Direction of Inventive Activity", Princeton University Press, Princeton, NJ.
[5] Azoulay P., Ding W., Stuart T., 2006. "The impact of academic patenting on the Rate, Quality, and Direction of (Public) Research Output, NBER WP N.11917, January 2006.
[6] Bacchiocchi E., Montobbio F., 2006. "Knowledge diffusion from university and public research. A comparison between US, Japan and Europe using patent citations", CESPRI WP n. 193, Bocconi University.
[7] Breschi S., Lissoni F., Montobbio F., 2006. "University Patenting and Scientific Productivity. A quantitative study of Italian Academic Inventors", CESPRI WP n.189, November 2006, Bocconi University.
[8] Breschi S., Lissoni F., Montobbio F., 2007. "The Scientific Productivity of Academic Inventors: New Evidence from Italian Data", Economics of Innovation and New Technology, 16(2):101-18.
[9] Carayol N., 2007. "Academic Incentives, Research Organization and Patenting at a Large French University", Economics of Innovation and New Technology, 16(2):119-138.
[10] Dasgupta P., David P.A., 1987. "Information Disclosure and the Economics of Science and Technology," in Feiwel G.R. (ed.), "Arrow and the Ascent of Modern Economic Theory", New York University Press.
[11] Diamond A.M., 1986. "What is a Citation Worth?", The Journal of Human Resources, 21(2):200-215.
[12] Fabrizio K.R., Di Minin A., 2005. "Commercializing the Laboratory: The Relationship Between Faculty Patenting and Publishing", mimeo.
[13] Geuna A., Nesta L., 2003. "University Patenting and its Effects on Academic Research", SEWPS Paper n.99.
[14] Hagstrom W.O., 1965. "The Scientific Community", Basic Books Inc., New York-London.
[15] Hall B.H. (2005), "Exploring the Patent Explosion", The Journal of Technology Transfer, 30(2_2): 35-48.
[16] Heckman J. J., 1979. " Sample Selection Bias as a Specification Error", Econometrica 47, 153-161.
[17] Heller M.A. Eisenberg R.S., 1998. "Can patent deter innovation? The Anti-commons in biomedical research", Science , 280(5364):698-701

[18] Henderson R., Jaffe A.B., Trajtenberg M., 1998. "Universities as a Source of Commercial Technology: A detailed analysis of University Patenting, 1965-1988". Review of Economics and Statistics 80 (1):119-127.

[19] Hernan M.A., Brumback B. and Robins J.M., 2001. "Marginal Structural Models to Estimate the Joint Causal Effect of Nonrandomized Treatments", Journal of the American Statistical Association, 96:454, 440-48.

[20] Layton E.T. J., 1974. "Technology as Knowledge", Technology & Culture, 15( 1):31-41.

[21] Lee, Y.S., 1996. 'Technology transfer' and the research university: a search for the boundaries of university–industry collaboration", Research Policy, 25(6):843–863.

[22] Levin S.G, Stephan P.E., Walker M.B., 1995. Planck's Principle Revisited: A Note, Social Studies of Science, 25: 275-83.

[23] Li Q., Racine J.S., 2004. "Nonparametric estimation of regression functions with both categorical and continuous data", Journal of Econometrics, 119(1):99-130.

[24] Lowry S., 2005. "Madey vs. Duke University Clarifies Patent Protection and Bolsters Patentee Rights", mimeo.

[25] Mansfield E., 1995. "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing", Review of Economics and Statistics, 77(1):55-65.

[26] Merton R.K.,1968. "The Matthew effect in science", Science, 159(3810): 56-63.

[27] Milken Institute, 2006. "Mind to market: A Global Analysis of University Biotechnology Transfer and Commercialization", September 2006.

[28] Mowery D.C., Nelson R.R, Sampat B.N., Ziedonis A.A., 2001. "The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh-Dole act of 1980", Research Policy 30 (2):99-119.

[29] Murray, F. and Stern, S., 2007. "Do formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the anti-Commons Hypothesis", Journal of Economic Behavior and Organization, 63/4):648:687.

[30] Myer G., 1995. "From Discovery to Invention. The Writing and Rewriting of Two Patents", Social Studies of Science, 2:57-105

[31] Narin F., Hamilton K.S., 1996. "Bibliometric Performance Measures", Scientometrics 36(3):293-310.

[32] Narin F., Pinski G., Gee H.H., 1976. "Structure of the Biomedical Literature", Journal of the American Society for Information Science, January-February, 25:45.

[33] Nelson R.R., 2004. "The market economy, and the scientific commons", Research Policy 33:455–471.

[34] OECD, 2003. "Turning Science into Business. Patenting and Licensing at Public Research Organizations", Organisation for Economic Co-operation and Development, Paris.

[35] ProTon, 2005. "The ProTon Europe 2005 Annual Survey Report", ProTon, April 2007.

[36] Robins J.M., Hernan M.A. and Brumback B., 2000. "Marginal Structural Models and Causal Inference in Epidemiology", Epidemiology, 11:5, 550-60.

[37] Rosembaum P.R. and Rubin D.B., 1983. "The Central Role of the Propensity Scores in Observational Studies for Causal Effects", Biometrika 70:1, 41-55.

[38] Stephan P.E., 2004. "Robert K. Merton's perspective on priority and the provision of the public good knowledge", Scientometrics, 60(1):81-87.

[39] Stephan P.E., Gurmu S., Sumell A.J., Black G., 2004. "Who's patenting in the university? Evidence from a Survey of Doctorate Recipients", Economics of Innovation and New Technology, 16(2):71-99.

[40] Thursby J.G., Kemp S., 2002. Growth and productive efficiency of university intellectual property licensing, Research Policy, 31:109–124.

[41] Trajtenberg M., Henderson R., Jaffe A.B., 1997. "University versus Corporate Patents: A Window on the Basicness of Invention", Economics of Innovation and New Technology, 5:19-50.

[42] Van Looy, B., Ramga, M., Callaert, J., Debackere, K., Zimmermann, E., 2004. "Combining entrepreneurial and scientific performance in academia: towards a compounded and reciprocal Matthew-effect?", Research Policy, 33(3):425-441.

[43] Vincenti W.G., 1990. "What Engineers know and how they know it. Analytical studies from aeronautical history", John Hopkins University Press, Baltimore.

[44] Walsh J.P., Cho C., Cohen W.M., 2005. "Patents, Material Transfers and Access to Research Inputs in Biomedical Research", Report to the National Academy of Sciences' Committee Intellectual Property Rights in Genomic and Protein-Related Inventions.

[45] Wooldridge J.M., 2002. "Econometric Analysis of Cross Section and Panel Data", Cambridge, MA: The MIT Press.

**Appendix: IPTW computation and “robustness” checks.**

Given the logit estimates of Tables 10, let $\hat{p}_{it}$ the predicted probability for subject *i* of being treated for each year *t*. Our regime of treating specification assumes that the status of “patenter” for each researcher lasts until the end of the “follow-up”, then $\hat{p}_{it}$ equals one for each year after the first patent. The denominator[15] of the weights are computed as follows:

1) Calculate the probability of each subject *i* to receive the observed treatment at time *t* :

$$IPTW_{it}^{*} = \hat{p}_{it} * POSTPAT_{it} + (1-\hat{p}_{it}) * (1-POSTPAT_{it})$$

2) Estimate each subject’s probability of complete treatment history up to each year *t*:

$$IPTW_{it} = \prod_{k=0}^{t-1} IPTW_{it-k}^{*}$$

As stated in section 4 the reliability of the IPTW approach relies on the strong assumptions that there are no unobserved confounders and that the selection equation used to estimate the weights is correctly specified. Although the first assumption cannot be tested, we can relax the second one by re-estimating the predicted probabilities $\hat{p}_{it}$ by means of nonparametric approach which do not require any specification of the likelihood structure of the model (eg. logit distribution) nor any specific functional form which links the regressors to the dependent variables (eg. additive linear, with quadratics, with interactions, an so on). Several kernel estimators for categorical data has been proposed in literature (Aitchison and Aitken, 1976), in particular we follow the method proposed by Li and Racine (2004) which is for estimating density function defined over both discrete $(x^d)$ and continuous $(x^c)$ variables using the following joint kernel density estimator:

$$\hat{f}(x^d, x^c) = \frac{1}{nh_x}\sum_{i=1}^{n} L(X_i^d = x^d) W\left(\frac{X_i^C - x^C}{h_x}\right)$$

where $L(X_i^d = x^d)$ is a categorical data kernel function, $W\left[\left(X_i^C - x^C\right)/h_{x^C}\right]$ is a continuous data kernel function and $h_x$ is the bandwidth for the continuous variable chosen via cross-validation methods (see Li and Racine, 2004 for further details).

Given the computationally intensive nature of these procedures which complexity increases exponentially with the sample size, we estimated the nonparametric version of the IPTW only for the subset of the engineers and the chemists and not for the whole sample.

[15] The numerator is computed similarly using the predicted probabilities according to the model in Table 11.

We then re-run all the ZINB-ML, OLS-FE and Heckman-FE models with robust IPTW estimated non-parametrically. The findings are similar to the previous ones[16] and are summarized in Figure 2 below:

[16] Detailed results and estimation routines are available upon request.

# Tables

**Table 1 - Distribution of patent per year in classes of applicants and inventor age at the time of priority.**

| year | inventor | research institution | private company | inventor age at priority date | | | | | | | total |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | 24 - 30 | 31 - 35 | 36 - 40 | 41 - 45 | 46 - 50 | 51 - 55 | 56 - 59 | |
| 1971 | 0 | 0 | 0 | 0 | | | | | | | 0 |
| 1972 | 0 | 0 | 0 | 0 | | | | | | | 0 |
| 1973 | 0 | 0 | 0 | 0 | | | | | | | 0 |
| 1974 | 0 | 0 | 0 | 0 | | | | | | | 0 |
| 1975 | 1 | 1 | 0 | 2 | 0 | | | | | | 2 |
| 1976 | 0 | 0 | 2 | 2 | 0 | | | | | | 2 |
| 1977 | 0 | 0 | 0 | 0 | 0 | | | | | | 0 |
| 1978 | 1 | 0 | 2 | 2 | 1 | | | | | | 3 |
| 1979 | 0 | 0 | 2 | 0 | 2 | | | | | | 2 |
| 1980 | 0 | 0 | 2 | 2 | 0 | 0 | | | | | 2 |
| 1981 | 0 | 0 | 1 | 0 | 0 | 1 | | | | | 1 |
| 1982 | 1 | 1 | 2 | 1 | 1 | 2 | | | | | 4 |
| 1983 | 0 | 3 | 3 | 1 | 2 | 3 | | | | | 6 |
| 1984 | 1 | 2 | 1 | 1 | 3 | 0 | | | | | 4 |
| 1985 | 0 | 4 | 10 | 6 | 1 | 6 | 1 | | | | 14 |
| 1986 | 0 | 3 | 6 | 6 | 1 | 2 | 0 | | | | 9 |
| 1987 | 0 | 0 | 4 | 3 | 0 | 1 | 0 | | | | 4 |
| 1988 | 0 | 0 | 8 | 2 | 3 | 3 | 0 | | | | 8 |
| 1989 | 1 | 0 | 19 | 5 | 5 | 6 | 4 | | | | 20 |
| 1990 | 1 | 0 | 10 | 2 | 7 | 0 | 1 | 1 | | | 11 |
| 1991 | 1 | 2 | 19 | 6 | 5 | 4 | 6 | 1 | | | 22 |
| 1992 | 0 | 3 | 18 | 6 | 6 | 5 | 4 | 0 | | | 21 |
| 1993 | 0 | 1 | 21 | 5 | 9 | 3 | 3 | 2 | | | 22 |
| 1994 | 0 | 2 | 15 | 6 | 6 | 2 | 1 | 2 | | | 17 |
| 1995 | 0 | 1 | 20 | 5 | 3 | 7 | 1 | 4 | 1 | | 21 |
| 1996 | 1 | 1 | 17 | 0 | 8 | 6 | 3 | 1 | 1 | | 19 |
| 1997 | 0 | 2 | 36 | 9 | 10 | 6 | 9 | 4 | 0 | | 38 |
| 1998 | 0 | 2 | 17 | 4 | 8 | 4 | 3 | 0 | 0 | | 19 |
| 1999 | 0 | 0 | 18 | 6 | 3 | 2 | 3 | 2 | 2 | | 18 |
| 2000 | 0 | 2 | 8 | 2 | 0 | 4 | 2 | 2 | 0 | 0 | 10 |
| 2001 | 0 | 2 | 4 | 2 | 3 | 0 | 0 | 0 | 1 | 0 | 6 |
| **total** | **8** (2.62%) | **32** (10.49%) | **265** (86.89%) | **86** (28.20%) | **87** (28.52%) | **67** (21.97%) | **41** (13.44%) | **19** (6.23%) | **5** (1.64) | **0** 0% | **305** |

**Table 2(a) - Summary of variables used in the analysis.**

| Variable | Description |
|---|---|
| **Time varying** | |
| T | Year |
| Scientific productivity | |
| publ_m | Number of scientific publications (authored & co-authored) in year t |
| publ_msq | Publ_m squared |
| cumpubl_m | Total number of scientific publications since entering in academia (authored & co-authored) up to year t |
| publbas4 | Number of scientific publications (authored & co-authored) in year t in basic research |
| aut_m | Average number of authors for scientific pubblications in year t |
| aut_m4 | Average number of authors for scientific pubblications in year t in basic research |
| wpubl_m | Publ_m/aut_m (=0 if publ_m =0) |
| wpublbas4 | publbas4/aut_m4 (=0 if publbas4=0) |
| lpubl_m | log(publ_m+1) |
| lpublbas4 | log(publbas4+1) |
| lwpubl_m | log(wpubl_m+1) |
| lwpublbas4 | log(wpublbas4+1) |
| dumpubl_m | = 1 if she has at least 1 pubblication in year t; = 0 otherwise |
| Stdifac | Average journal IF (standardized by scientific field of the journal) for publications in year t |
| Patenting activity: | |
| Pat | Number of patents in year t |
| Postpat | = 1 for years during and after the first patent; = 0 for years before the first patent |
| Control variates: | |
| Seniority | Years spent in academia up to year t |
| Senioritysq | seniority squared |
| Cumpat | Total number of patents up to year t |
| Expttoma | Total number of patents assigned to the institution of affiliation between year t-4 and year t |
| Expttomasq | expttoma squared |
| dumyear75_79 | = 1 t is between 1975-1979 (calendar effect); = 0 otherwise |
| dumyear80_84 | = 1 t is between 1980-1984 (calendar effect); = 0 otherwise |
| dumyear85_89 | = 1 t is between 1985-1989 (calendar effect); = 0 otherwise |
| dumyear90_94 | = 1 t is between 1990-1994 (calendar effect); = 0 otherwise |
| **Non time varying** | |
| Gender | = 1 if male, = 0 if female |
| Scientific field dummies: | |
| dumSF1 | = 1 if CHEMISTRY; = 0 otherwise |
| dumSF2 | = 1 if ENGINEERING; = 0 otherwise |
| dumSF4 | = 1 if PHYSICS; = 0 otherwise |
| dumSF3 | = 1 if OTHER; = 0 otherwise |

**Table 2(b) - Summary of variables used in the analysis.**

| Variable | ALL (1276 scientists) | | | | | ENGINEERS (309 scientists) | | | | | CHEMISTS (919 scientists) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Obs | Mean | Std. | Min | Max | Obs | Mean | Std. | Min | Max | Obs | Mean | Std. | Min | Max |
| **Time varying** | | | | | | | | | | | | | | | |
| t | 22385 | 1992.73 | 7.55 | 1975 | 2003 | 5341 | 1992.8 | 7.509 | 1975 | 2003 | 16092 | 1992.6 | 7.589 | 1975 | 2003 |
| Scientific | | | | | | | | | | | | | | | |
| publ_m | 22385 | 1.27 | 2.25 | 0 | 25 | 5341 | 0.771 | 1.756 | 0 | 25 | 16092 | 1.457 | 2.379 | 0 | 25 |
| publ_msq | 22385 | 6.66 | 25.18 | 0 | 625 | 5341 | 3.678 | 19.54 | 0 | 625 | 16092 | 7.783 | 26.843 | 0 | 625 |
| cumpubl_m | 23284 | 11.17 | 21.43 | 0 | 292 | 5492 | 5.973 | 13.01 | 0 | 127 | 16188 | 13.615 | 23.670 | 0 | 292 |
| publbas4 | 22385 | 0.58 | 1.51 | 0 | 25 | 5341 | 0.097 | 0.464 | 0 | 7 | 16092 | 0.753 | 1.686 | 0 | 25 |
| aut_m | 9828 | 3.15 | 1.82 | 1 | 20 | 1654 | 3.296 | 1.982 | 1 | 20 | 7878 | 3.106 | 1.765 | 1 | 20 |
| aut_m4 | 5493 | 4.86 | 1.89 | 1 | 20 | 328 | 5.461 | 2.667 | 1 | 20 | 5011 | 4.823 | 1.805 | 1 | 20 |
| wpubl_m | 22385 | 0.68 | 1.59 | 0 | 22.32 | 5341 | 0.402 | 1.235 | 0 | 21.55 | 16092 | 0.781 | 1.690 | 0 | 22.32 |
| wpublbas4 | 22385 | 0.13 | 0.35 | 0 | 6.485 | 5341 | 0.021 | 0.108 | 0 | 1.750 | 16092 | 0.172 | 0.387 | 0 | 6.485 |
| lpubl_m | 22385 | 0.53 | 0.69 | 0 | 3.258 | 5341 | 0.343 | 0.580 | 0 | 3.258 | 16092 | 0.600 | 0.713 | 0 | 3.258 |
| lpublbas4 | 22385 | 0.26 | 0.52 | 0 | 3.258 | 5341 | 0.054 | 0.227 | 0 | 2.079 | 16092 | 0.340 | 0.572 | 0 | 3.258 |
| lwpubl_m | 22385 | 1.68 | 1.59 | 1 | 23.32 | 5341 | 1.402 | 1.235 | 1 | 22.55 | 16092 | 1.781 | 1.690 | 1 | 23.32 |
| lwpublbas4 | 22385 | 1.13 | 0.35 | 1 | 7.485 | 5341 | 1.021 | 0.108 | 1 | 2.750 | 16092 | 1.172 | 0.387 | 1 | 7.485 |
| dumpubl_m | 23284 | 0.59 | 0.49 | 0 | 1 | 5492 | 0.491 | 0.500 | 0 | 1 | 16188 | 0.647 | 0.478 | 0 | 1 |
| stdifac | 9296 | -0.05 | 0.81 | -1.90 | 8.530 | 1536 | -0.272 | 0.678 | -1.900 | 4.921 | 7475 | -0.004 | 0.828 | - | 8.530 |
| Patenting | | | | | | | | | | | | | | | |
| pat | 19833 | 0.02 | 0.15 | 0 | 5 | 4723 | 0.019 | 0.169 | 0 | 3 | 14254 | 0.014 | 0.140 | 0 | 5 |
| postpat | 19833 | 0.07 | 0.25 | 0 | 1 | 4723 | 0.074 | 0.262 | 0 | 1 | 14254 | 0.068 | 0.252 | 0 | 1 |
| cumpat | 19833 | 0.14 | 0.72 | 0 | 16 | 4723 | 0.184 | 0.838 | 0 | 8 | 14254 | 0.133 | 0.637 | 0 | 10 |
| Control variates: | | | | | | | | | | | | | | | |
| seniority | 22385 | 12.61 | 8.53 | 1 | 38 | 5341 | 12.482 | 8.526 | 1 | 37 | 16092 | 12.687 | 8.573 | 1 | 38 |
| senioritysq | 22385 | 231.84 | 274.3 | 1 | 1444 | 5341 | 228.46 | 274.2 | 1 | 1369 | 16092 | 234.44 | 276.524 | 1 | 1444 |
| expttoma | 22385 | 3.05 | 5.13 | 0 | 30 | 5341 | 2.946 | 4.927 | 0 | 30 | 16092 | 3.075 | 5.204 | 0 | 30 |
| expttomasq | 22385 | 35.66 | 104.9 | 0 | 900 | 5341 | 32.947 | 99.64 | 0 | 900 | 16092 | 36.532 | 107.173 | 0 | 900 |
| dumyear75_79 | 22385 | 0.07 | 0.25 | 0 | 1 | 5341 | 0.066 | 0.248 | 0 | 1 | 16092 | 0.069 | 0.254 | 0 | 1 |
| dumyear80_84 | 22385 | 0.10 | 0.31 | 0 | 1 | 5341 | 0.101 | 0.301 | 0 | 1 | 16092 | 0.107 | 0.309 | 0 | 1 |
| dumyear85_89 | 22385 | 0.15 | 0.35 | 0 | 1 | 5341 | 0.144 | 0.351 | 0 | 1 | 16092 | 0.147 | 0.354 | 0 | 1 |
| dumyear90_94 | 22385 | 0.20 | 0.40 | 0 | 1 | 5341 | 0.195 | 0.396 | 0 | 1 | 16092 | 0.195 | 0.396 | 0 | 1 |
| **Time invariant** | | | | | | | | | | | | | | | |
| gender | 22385 | 0.69 | 0.46 | 0 | 1 | | | | | | | | | | |
| Scientific field | | | | | | | | | | | | | | | |
| dumSF1 | 21980 | 0.72 | 0.45 | 0 | 1 | | | | | | | | | | |
| dumSF2 | 21980 | 0.24 | 0.43 | 0 | 1 | | | | | | | | | | |
| dumSF4 | 21980 | 0.01 | 0.10 | 0 | 1 | | | | | | | | | | |
| dumSF3 | 21980 | 0.03 | 0.18 | 0 | 1 | | | | | | | | | | |

**Table 3**

| | Journal IF of published articles | | |
|---|---|---|---|
| field | Mean | Std. Dev. | Freq. |
| | | | |
| **CHEMISTRY** | 2.367 | 1.349 | 14704 |
| **ENGINEERING** | 1.166 | 0.878 | 3466 |
| **OTHER** | 1.884 | 0.823 | 695 |
| **PHYSICS** | 2.676 | 2.965 | 5869 |
| | | | |
| Total | 2.258 | 1.875 | 24734 |

Test: Equality of populations (Kruskal-Wallis test)
Chi-squared = 3427.049; 3 d.f. prob = 0.0001
chi-sq with ties = 3427.720; 3 d.f.prob = 0.0001
Bartlett's test for equal variances:
chi2(3) = 9.1e+03 Prob>chi2 = 0.000

**Table 4**

| | Journal level of published articles | | |
|---|---|---|---|
| field | Mean | Std. Dev. | Freq. |
| | | | |
| **CHEMISTRY** | 3.531 | 0.537 | 16656 |
| **ENGINEERING** | 2.264 | 0.582 | 3770 |
| **OTHER** | 3.015 | 0.908 | 883 |
| **PHYSICS** | 3.643 | 0.490 | 6011 |
| | | | |
| Total | 3.364 | 0.712 | 27320 |

Test: Equality of populations (Kruskal-Wallis test)
Chi-squared = 7068.608; 3 d.f. prob = 0.0001
chi-squ with ties = 8616.147; 3 d.f.prob = 0.0001
Bartlett's test for equal variances:
chi2(3) = 831.3114 Prob>chi2 = 0.000

**Table 5**

| | Age of scientists | | |
|---|---|---|---|
| field | Mean | Std. Dev. | Freq. |
| | | | |
| **CHEMISTRY** | 35.687 | 8.573 | 16092 |
| **ENGINEERING** | 35.482 | 8.526 | 5341 |
| **OTHER** | 35.394 | 8.060 | 254 |
| **PHYSICS** | 35.032 | 7.611 | 698 |
| | | | |
| Total | 35.614 | 8.528 | 22385 |

Test: Equality of populations (Kruskal-Wallis test)
Chi-squared = 2.623; 3 d.f. prob = 0.4534
chi-sq with ties = 2.627; 3 d.f.prob = 0.4527
Bartlett's test for equal variances:
chi2(3) = 19.5610 Prob>chi2 = 0.000

**Table 6**

| | Annual number of publications | | |
|---|---|---|---|
| Field | Mean | Std. Dev. | Freq. |
| | | | |
| **CHEMISTRY** | 1.457 | 2.379 | 16092 |
| **ENGINEERING** | 0.771 | 1.756 | 5341 |
| **OTHER** | 0.327 | 0.765 | 254 |
| **PHYSICS** | 1.014 | 2.198 | 698 |
| | | | |
| Total | 1.267 | 2.249 | 22385 |

Test: Equality of populations (Kruskal-Wallis test)
Chi-squared = 536.295; 3 d.f. prob = 0.0001
chi-squ with ties = 653.910; 3 d.f.prob = 0.0001
Bartlett's test for equal variances:
chi2(3) = 932.9579 Prob>chi2 = 0.000

**Table 7**

| | Annual number of patents | | |
|---|---|---|---|
| field | Mean | Std. Dev. | Freq. |
| | | | |
| **CHEMISTRY** | 0.014 | 0.140 | 14254 |
| **ENGINEERING** | 0.019 | 0.169 | 4723 |
| **OTHER** | 0.000 | 0.000 | 228 |
| **PHYSICS** | 0.008 | 0.105 | 628 |
| | | | |
| Total | 0.015 | 0.145 | 19833 |

Test: Equality of populations (Kruskal-Wallis test)
Chi-squared = 0.300; 3 d.f. prob = 0.9600
chi-sq with ties = 7.934; 3 d.f.prob = 0.0474
Bartlett's test for equal variances:
chi2(3) = 379.2983 Prob>chi2 = 0.000

**Table 8**

| | Cumulative number of patents in 2001 | | |
|---|---|---|---|
| Field | Mean | Std. Dev. | Freq. |
| | | | |
| **CHEMISTRY** | 0.224 | 0.941 | 917 |
| **ENGINEERING** | 0.298 | 1.126 | 309 |
| **OTHER** | 0.000 | 0.000 | 13 |
| **PHYSICS** | 0.143 | 0.550 | 35 |
| | | | |
| Total | 0.237 | 0.977 | 1274 |

Test: Equality of populations (Kruskal-Wallis test)
Chi-squared = 0.608 ; 3 d.f. prob = 0.8946
chi-sq with ties = 2.191; 3 d.f.prob = 0.5338
Bartlett's test for equal variances:
chi2(3) = 32.3395 Prob>chi2 = 0.000

**Table 9 Distribution of cumpat in 2001**

| ALL | | | |
|---|---|---|---|
| **cumpat** | **Freq.** | **Percent** | **Cum.** |
| 0 | 1,143 | 89.72 | 89.72 |
| 1 | 71 | 5.57 | 95.29 |
| 2 | 19 | 1.49 | 96.78 |
| 3 | 16 | 1.26 | 98.04 |

| | | | |
|---|---|---|---|
| 4 | 9 | 0.71 | 98.74 |
| 5 | 5 | 0.39 | 99.14 |
| 6 | 5 | 0.39 | 99.53 |
| 7 | 2 | 0.16 | 99.69 |
| 8 | 3 | 0.24 | 99.92 |
| 10> | 1 | 0.08 | 100 |
| **ENGINEERING** | | | |
| **cumpat** | **Freq.** | **Percent** | **Cum.** |
| 0 | 273 | 88.35 | 88.35 |
| 1 | 20 | 6.47 | 94.82 |
| 2 | 4 | 1.29 | 96.12 |
| 3 | 2 | 0.65 | 96.76 |
| 4 | 4 | 1.29 | 98.06 |
| 6 | 3 | 0.97 | 99.03 |
| 8 | 3 | 0.97 | 100 |
| **CHEMISTRY** | | | |
| **cumpat** | **Freq.** | **Percent** | **Cum.** |
| 0 | 825 | 89.97 | 89.97 |
| 1 | 49 | 5.34 | 95.31 |
| 2 | 15 | 1.64 | 96.95 |
| 3 | 13 | 1.42 | 98.36 |
| 4 | 5 | 0.55 | 98.91 |
| 5 | 5 | 0.55 | 99.45 |
| 6 | 2 | 0.22 | 99.67 |
| 7 | 2 | 0.22 | 99.89 |
| 10> | 1 | 0.11 | 100 |

**The following Logit-ML models are estimated using, for each scientist, only observations up to the year of the first patent (included).**

**Table 10** IPTW Estimation. Probability of patenting, logit ML regression (dep. Variable **postpat**)

| | IPTW denominator (patent regime) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **(all)** | | | **(eng)** | | | **(chem)** | | |
| | **Coeff** | **Se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.132 | 0.048 | 0.007 | 0.161 | 0.105 | 0.127 | 0.139 | 0.056 | 0.013 |
| **senioritysq** | -0.005 | 0.002 | 0.002 | -0.006 | 0.004 | 0.088 | -0.005 | 0.002 | 0.004 |
| **gender** | 0.427 | 0.224 | 0.056 | -0.389 | 0.435 | 0.372 | 0.666 | 0.272 | 0.014 |
| **expttoma(-1)** | 0.116 | 0.066 | 0.079 | 0.248 | 0.198 | 0.211 | 0.076 | 0.074 | 0.302 |
| **expttomasq(-1)** | -0.010 | 0.006 | 0.112 | -0.034 | 0.028 | 0.221 | -0.006 | 0.006 | 0.320 |
| **publ_m(-1)** | 0.098 | 0.090 | 0.276 | 0.887 | 0.333 | 0.008 | 0.003 | 0.089 | 0.972 |
| **publ_msq(-1)** | -0.008 | 0.009 | 0.329 | -0.189 | 0.086 | 0.028 | 0.000 | 0.006 | 0.952 |
| **cumpubl_m(-1)** | 0.000 | 0.007 | 0.982 | 0.002 | 0.026 | 0.951 | -0.002 | 0.008 | 0.773 |
| **constant** | -6.016 | 0.335 | 0.000 | -5.641 | 0.641 | 0.000 | -6.125 | 0.418 | 0.000 |
| | Number of obs = 17317<br>Wald chi2(12) = 33.18<br>Prob > chi2 = 0.0009<br>Pseudo R2 = 0.0163<br>Loglikelihood = -738.5563 | | | Number of obs = 4016<br>Wald chi2(12) = 35.63<br>Prob > chi2 = 0.0004<br>Pseudo R2 = 0.0514<br>Log likelihood = -181.48308 | | | Number of obs = 12232<br>Wald chi2(12) = 25.50<br>Prob > chi2 = 0.0126<br>Pseudo R2 = 0.0201<br>Log likelihood = -511.44665 | | |

* Calendar dummy variables included

**Table 11** Probability of patenting, logit ML regression (dep. Variable **postpat**)

| | IPTW numerator (patent regime) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **(all)** | | | **(eng)** | | | **(chem)** | | |
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.144 | 0.046 | 0.002 | 0.216 | 0.100 | 0.030 | 0.138 | 0.053 | 0.009 |
| **senioritysq** | -0.005 | 0.002 | 0.001 | -0.008 | 0.003 | 0.028 | -0.005 | 0.002 | 0.003 |
| **gender** | 0.430 | 0.223 | 0.054 | -0.375 | 0.438 | 0.393 | 0.661 | 0.271 | 0.015 |
| **expttoma(-1)** | 0.118 | 0.066 | 0.076 | 0.253 | 0.197 | 0.197 | 0.076 | 0.074 | 0.304 |
| **expttomasq(-1)** | -0.010 | 0.006 | 0.109 | -0.035 | 0.027 | 0.198 | -0.006 | 0.006 | 0.322 |
| **constant** | -6.044 | 0.332 | 0.000 | -5.763 | 0.636 | 0.000 | -6.119 | 0.413 | 0.000 |
| | Number of obs = 17317<br>Wald chi2(9) = 29.36<br>Prob > chi2 = 0.0006<br>Pseudo R2 = 0.0155<br>Log likelihood = -739.13778 | | | Number of obs = 4016<br>Wald chi2(9) = 14.31<br>Prob > chi2 = 0.1118<br>Pseudo R2 = 0.0356<br>Log likelihood = -184.49896 | | | Number of obs = 12232<br>Wald chi2(9) = 25.04<br>Prob > chi2 = 0.0029<br>Pseudo R2 = 0.0200<br>Log likelihood = -511.48054 | | |

* Calendar dummy variables included

| **Table 13** | ZINB ML regression with clustered id robust SE dep. Variable **publ_m** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **(all)** | | | **(eng)** | | | **(chem)** | | |
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.098 | 0.013 | 0.000 | 0.101 | 0.048 | 0.035 | 0.093 | 0.012 | 0.000 |
| **senioritysq** | -0.002 | 0.000 | 0.000 | -0.003 | 0.001 | 0.036 | -0.002 | 0.000 | 0.000 |
| **gender** | 0.244 | 0.076 | 0.001 | 0.194 | 0.271 | 0.475 | 0.323 | 0.080 | 0.000 |
| **postpat(-1)** | 0.021 | 0.135 | 0.876 | 0.608 | 0.330 | 0.065 | -0.224 | 0.115 | 0.052 |
| **cumpat(-1)** | 0.034 | 0.038 | 0.371 | -0.022 | 0.077 | 0.775 | 0.075 | 0.040 | 0.062 |
| **expttoma(-1)** | 0.017 | 0.007 | 0.011 | 0.023 | 0.021 | 0.278 | 0.013 | 0.007 | 0.071 |
| **dumyear75_79** | -0.619 | 0.115 | 0.000 | -1.081 | 0.345 | 0.002 | -0.609 | 0.119 | 0.000 |
| **dumyear80_84** | -0.295 | 0.089 | 0.001 | -0.727 | 0.255 | 0.004 | -0.286 | 0.092 | 0.002 |
| **dumyear85_89** | -0.200 | 0.064 | 0.002 | -0.821 | 0.233 | 0.000 | -0.158 | 0.065 | 0.015 |
| **dumyear90_94** | -0.047 | 0.043 | 0.275 | -0.483 | 0.123 | 0.000 | -0.008 | 0.044 | 0.854 |
| **constant** | -0.599 | 0.133 | 0.000 | -0.684 | 0.499 | 0.171 | -0.477 | 0.131 | 0.000 |
| | | | | | | | | | |
| **inflate** | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | -0.971 | 0.079 | 0.000 | -0.829 | 0.170 | 0.000 | -1.010 | 0.089 | 0.000 |
| **senioritysq** | 0.023 | 0.002 | 0.000 | 0.021 | 0.004 | 0.000 | 0.023 | 0.003 | 0.000 |
| **gender** | 0.477 | 0.206 | 0.021 | -0.176 | 0.524 | 0.736 | 0.582 | 0.234 | 0.013 |
| **postpat(-1)** | -1.924 | 1.655 | 0.245 | 1.089 | 2.432 | 0.654 | -3.405 | 2.434 | 0.162 |
| **cumpat(-1)** | -0.402 | 0.300 | 0.180 | -1.450 | 3.033 | 0.633 | -0.195 | 0.343 | 0.571 |
| **expttoma(-1)** | 0.009 | 0.033 | 0.789 | -0.054 | 0.093 | 0.563 | 0.003 | 0.043 | 0.948 |
| **dumyear75_79** | -1.117 | 0.343 | 0.001 | -0.952 | 0.730 | 0.192 | -1.214 | 0.383 | 0.002 |
| **dumyear80_84** | -1.253 | 0.313 | 0.000 | -0.831 | 0.801 | 0.299 | -1.472 | 0.341 | 0.000 |
| **dumyear85_89** | -0.755 | 0.265 | 0.004 | -1.130 | 0.691 | 0.102 | -0.828 | 0.289 | 0.004 |
| **dumyear90_94** | -0.065 | 0.232 | 0.778 | -0.368 | 0.555 | 0.508 | -0.202 | 0.249 | 0.418 |
| **constant** | 4.087 | 0.315 | 0.000 | 4.699 | 0.598 | 0.000 | 4.163 | 0.369 | 0.000 |
| | | | | | | | | | |
| **/lnalpha** | 0.300 | 0.042 | 0.000 | 0.585 | 0.118 | 0.000 | 0.091 | 0.061 | 0.135 |
| **alpha** | 1.350 | 0.057 | | 1.795 | 0.212 | | 1.096 | 0.067 | |
| | Number of obs = 18559<br>Nonzero obs = 8373<br>Wald chi2(10) = 330.39<br>Prob > chi2 = 0.0000<br>Log likelihood = -26330.38 | | | Number of obs = 4414<br>Nonzero obs = 1374<br>Wald chi2(10) = 92.20<br>Prob > chi2 = 0.0000<br>Log likelihood = -4592.511 | | | Number of obs = 13337<br>Nonzero obs = 6745<br>Wald chi2(10) = 280.57<br>Prob > chi2 = 0.0000<br>Log likelihood = -20415.4 | | |

**Tab 14** ZINB ML regression clustered id robust SE dep. Variable **publbas4**

| | **(all)** | | | **(eng)** | | | **(chem)** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.094 | 0.019 | 0.000 | 0.082 | 0.448 | 0.854 | 0.085 | 0.020 | 0.000 |
| **senioritysq** | -0.002 | 0.001 | 0.000 | -0.003 | 0.013 | 0.816 | -0.002 | 0.001 | 0.001 |
| **gender** | 0.343 | 0.129 | 0.008 | -0.869 | 0.469 | 0.064 | 0.485 | 0.131 | 0.000 |
| **postpat(-1)** | -0.075 | 0.301 | 0.803 | -0.326 | 0.696 | 0.639 | -0.438 | 0.221 | 0.048 |
| **cumpat(-1)** | -0.094 | 0.090 | 0.296 | -0.072 | 0.161 | 0.654 | 0.012 | 0.077 | 0.877 |
| **expttoma(-1)** | 0.029 | 0.011 | 0.008 | -0.017 | 0.160 | 0.916 | 0.025 | 0.011 | 0.023 |
| **dumyear75_79** | -0.199 | 0.167 | 0.235 | -0.070 | 1.019 | 0.945 | -0.284 | 0.165 | 0.084 |
| **dumyear80_84** | 0.059 | 0.143 | 0.679 | -0.312 | 0.931 | 0.737 | -0.024 | 0.143 | 0.869 |
| **dumyear85_89** | 0.114 | 0.102 | 0.264 | -0.454 | 0.331 | 0.170 | 0.061 | 0.102 | 0.549 |
| **dumyear90_94** | 0.137 | 0.070 | 0.051 | -0.854 | 0.470 | 0.069 | 0.118 | 0.070 | 0.093 |
| **constant** | -1.521 | 0.243 | 0.000 | -1.274 | 5.217 | 0.807 | -1.227 | 0.243 | 0.000 |
| | | | | | | | | | |
| **inflate** | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | -1.064 | 0.123 | 0.000 | -0.624 | 0.824 | 0.449 | -1.095 | 0.125 | 0.000 |
| **senioritysq** | 0.025 | 0.004 | 0.000 | 0.016 | 0.020 | 0.407 | 0.025 | 0.004 | 0.000 |
| **gender** | 0.241 | 0.308 | 0.435 | -2.119 | 6.173 | 0.731 | 0.441 | 0.322 | 0.171 |
| **postpat(-1)** | -5.340 | 6.819 | 0.434 | -13.094 | 1.458 | 0.000 | -14.210 | 4.473 | 0.001 |
| **cumpat(-1)** | 0.366 | 0.214 | 0.087 | 0.075 | 0.542 | 0.891 | 0.589 | 0.247 | 0.017 |
| **expttoma(-1)** | 0.025 | 0.057 | 0.659 | -0.223 | 0.166 | 0.177 | 0.019 | 0.063 | 0.767 |
| **dumyear75_79** | -1.353 | 0.491 | 0.006 | 0.454 | 2.224 | 0.838 | -1.625 | 0.498 | 0.001 |
| **dumyear80_84** | -1.627 | 0.472 | 0.001 | -1.451 | 2.859 | 0.612 | -1.946 | 0.471 | 0.000 |
| **dumyear85_89** | -0.695 | 0.419 | 0.097 | -0.265 | 1.204 | 0.826 | -0.974 | 0.407 | 0.017 |
| **dumyear90_94** | -0.260 | 0.383 | 0.497 | -1.548 | 4.745 | 0.744 | -0.527 | 0.360 | 0.143 |
| **constant** | 4.817 | 0.417 | 0.000 | 5.787 | 4.147 | 0.163 | 5.022 | 0.439 | 0.000 |
| | | | | | | | | | |
| **/lnalpha** | 1.132 | 0.023 | 0.000 | 1.896 | 1.492 | 0.204 | 0.759 | 0.032 | 0.000 |
| **alpha** | 3.101 | 0.070 | | 6.659 | 9.937 | | 2.137 | 0.068 | |
| | Number of obs = 18559<br>Nonzero obs = 4744<br>Wald chi2(10) = 88.22<br>Prob > chi2 = 0.0000<br>Log likelihood = -17266.05 | | | Number of obs = 4414<br>Nonzero obs = 270<br>Wald chi2(10) = 24.34<br>Prob > chi2 = 0.0067<br>Log likelihood = -1232.765 | | | Number of obs = 13337<br>Nonzero obs = 4342<br>Wald chi2(10) = 92.84<br>Prob > chi2 = 0.0000<br>Log likelihood = -14609.22 | | |

| Table 15 | OLS regression FE dep. Variable **lpubl_m** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **(all)** | | | **(eng)** | | | **(chem)** | | |
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.077 | 0.003 | 0.000 | 0.064 | 0.006 | 0.000 | 0.083 | 0.004 | 0.000 |
| **senioritysq** | -0.001 | 0.000 | 0.000 | -0.001 | 0.000 | 0.000 | -0.001 | 0.000 | 0.000 |
| **postpat(-1)** | 0.096 | 0.037 | 0.009 | 0.283 | 0.067 | 0.000 | -0.005 | 0.045 | 0.909 |
| **cumpat(-1)** | -0.005 | 0.014 | 0.733 | -0.011 | 0.020 | 0.593 | 0.005 | 0.018 | 0.784 |
| **expttoma(-1)** | 0.002 | 2.560 | 0.001 | 0.015 | 0.003 | 0.000 | 0.002 | 0.002 | 0.438 |
| **dumyear75_79** | 0.059 | 0.050 | 0.234 | 0.181 | 0.090 | 0.044 | 0.026 | 0.061 | 0.669 |
| **dumyear80_84** | 0.045 | 0.038 | 0.232 | 0.086 | 0.069 | 0.213 | 0.034 | 0.046 | 0.462 |
| **dumyear85_89** | -0.017 | 0.027 | 0.527 | -0.039 | 0.047 | 0.414 | -0.012 | 0.033 | 0.707 |
| **dumyear90_94** | -0.001 | 0.016 | 0.963 | -0.024 | 0.029 | 0.398 | 0.009 | 0.019 | 0.658 |
| **constant** | -0.217 | 0.025 | 0.000 | -0.284 | 0.033 | 0.000 | -0.185 | 0.031 | 0.000 |
| | Number of obs = 18559 | | | Number of obs = 4414 | | | Number of obs = 13337 | | |
| | R-squared = 0.5507 | | | R-squared = 0.5084 | | | R-squared = 0.5450 | | |

| Table 16 | OLS regression FE dep. Variable **lwpubl_m** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **(all)** | | | **(eng)** | | | **(chem)** | | |
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.102 | 0.008 | 0.000 | 0.091 | 0.015 | 0.000 | 0.108 | 0.010 | 0.000 |
| **senioritysq** | -0.001 | 0.000 | 0.000 | -0.001 | 0.000 | 0.000 | -0.002 | 0.000 | 0.000 |
| **postpat(-1)** | 0.060 | 0.100 | 0.549 | 0.550 | 0.164 | 0.001 | -0.293 | 0.121 | 0.015 |
| **cumpat(-1)** | 0.011 | 0.037 | 0.774 | -0.062 | 0.048 | 0.192 | 0.078 | 0.051 | 0.127 |
| **expttoma(-1)** | 0.004 | 0.005 | 0.387 | 0.014 | 0.008 | 0.089 | 0.003 | 0.001 | 0.036 |
| **dumyear75_79** | 0.034 | 0.135 | 0.803 | 0.450 | 0.231 | 0.052 | 0.001 | 0.006 | 0.814 |
| **dumyear80_84** | 0.008 | 0.102 | 0.938 | 0.246 | 0.182 | 0.176 | -0.097 | 0.166 | 0.560 |
| **dumyear85_89** | -0.057 | 0.073 | 0.434 | -0.035 | 0.118 | 0.768 | -0.082 | 0.126 | 0.516 |
| **dumyear90_94** | -0.014 | 0.043 | 0.748 | -0.109 | 0.068 | 0.108 | -0.079 | 0.093 | 0.392 |
| **constant** | 0.736 | 0.071 | 0.000 | 0.668 | 0.086 | 0.000 | 0.768 | 0.089 | 0.000 |
| | Number of obs = 18559 | | | Number of obs = 4414 | | | Number of obs = 13337 | | |
| | R-squared = 0.4258 | | | R-squared = 0.3701 | | | R-squared = 0.4275 | | |

| Table 17 | OLS regression FE dep. Variable **lpublbas4** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **(all)** | | | **(eng)** | | | **(chem)** | | |
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.035 | 0.002 | 0.000 | 0.013 | 0.002 | 0.000 | 0.042 | 0.003 | 0.000 |
| **senioritysq** | -0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -0.001 | 0.000 | 0.000 |
| **postpat(-1)** | -0.026 | 0.026 | 0.314 | -0.004 | 0.027 | 0.893 | -0.077 | 0.035 | 0.027 |
| **cumpat(-1)** | -0.001 | 0.008 | 0.938 | 0.001 | 0.006 | 0.900 | 0.010 | 0.012 | 0.406 |
| **expttoma(-1)** | 0.002 | 0.001 | 0.105 | 0.002 | 0.001 | 0.081 | 0.003 | 0.001 | 0.036 |
| **dumyear75_79** | -0.055 | 0.039 | 0.157 | 0.038 | 0.042 | 0.368 | 0.003 | 0.002 | 0.155 |
| **dumyear80_84** | -0.031 | 0.029 | 0.299 | 0.031 | 0.031 | 0.309 | -0.079 | 0.051 | 0.118 |
| **dumyear85_89** | -0.027 | 0.021 | 0.188 | 0.000 | 0.021 | 0.988 | -0.049 | 0.039 | 0.206 |
| **dumyear90_94** | -0.003 | 0.012 | 0.808 | -0.008 | 0.013 | 0.505 | -0.038 | 0.027 | 0.165 |
| **constant** | -0.100 | 0.020 | 0.000 | -0.051 | 0.014 | 0.000 | -0.120 | 0.026 | 0.000 |
| | Number of obs = 18559 | | | Number of obs = 4414 | | | Number of obs = 13337 | | |
| | R-squared = 0.5381 | | | R-squared = 0.3521 | | | R-squared = 0.5218 | | |

| Table 18 | OLS regression FE dep. Variable **lwpublas4** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **(all)** | | | **(eng)** | | | **(chem)** | | |
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.019 | 0.002 | 0.000 | 0.005 | 0.001 | 0.000 | 0.024 | 0.002 | 0.000 |
| **senioritysq** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | -0.001 | 0.000 | 0.000 |
| **postpat(-1)** | -0.019 | 0.019 | 0.301 | -0.014 | 0.011 | 0.203 | -0.073 | 0.020 | 0.000 |
| **cumpat(-1)** | -0.004 | 0.005 | 0.350 | 0.004 | 0.003 | 0.164 | 0.004 | 0.007 | 0.571 |
| **expttoma(-1)** | -0.001 | 0.001 | 0.349 | 0.001 | 0.001 | 0.200 | -0.001 | 0.001 | 0.278 |
| **dumyear75_79** | -0.027 | 0.029 | 0.355 | 0.011 | 0.022 | 0.619 | -0.037 | 0.038 | 0.335 |
| **dumyear80_84** | -0.004 | 0.022 | 0.862 | 0.014 | 0.014 | 0.315 | -0.013 | 0.029 | 0.668 |
| **dumyear85_89** | -0.010 | 0.016 | 0.555 | -0.001 | 0.010 | 0.947 | -0.017 | 0.021 | 0.441 |
| **dumyear90_94** | -0.002 | 0.009 | 0.866 | -0.005 | 0.006 | 0.388 | -0.004 | 0.012 | 0.769 |
| **constant** | 0.977 | 0.013 | 0.000 | 0.983 | 0.006 | 0.000 | 0.974 | 0.018 | 0.000 |
| | Number of obs = 18559 | | | Number of obs = 4414 | | | Number of obs = 13337 | | |
| | R-squared = 0.4830 | | | R-squared = 0.3195 | | | R-squared = 0.4647 | | |

| Table 19 | HECKMAN ML regression Dummy Variable Model | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Truncated equation:** dependent variable **stdifac** | | | | | | | | |
| | **all** | | | **engineers** | | | **chemists** | | |
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.158 | 0.008 | 0.000 | 0.019 | 0.013 | 0.141 | 0.146 | 0.008 | 0.000 |
| **senioritysq** | -0.003 | 0.000 | 0.000 | 0.000 | 0.000 | 0.836 | -0.003 | 0.000 | 0.000 |
| **postpat(-1)** | 0.159 | 0.062 | 0.011 | 0.278 | 0.118 | 0.018 | 0.035 | 0.075 | 0.638 |
| **cumpat(-1)** | -0.010 | 0.020 | 0.627 | -0.110 | 0.039 | 0.005 | 0.033 | 0.024 | 0.160 |
| **expttoma(-1)** | 0.012 | 0.003 | 0.000 | -0.004 | 0.005 | 0.505 | 0.013 | 0.003 | 0.000 |
| **_cons** | -0.844 | 0.067 | 0.000 | -1.036 | 0.140 | 0.000 | -0.676 | 0.072 | 0.000 |

| | **Selection equation:** dependent variable **dumpubl_m** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Coeff** | **se** | **P** | **coeff** | **se** | **P** | **coeff** | **se** | **P** |
| **seniority** | 0.239 | 0.005 | 0.000 | 0.265 | 0.012 | 0.000 | 0.248 | 0.006 | 0.000 |
| **senioritysq** | -0.006 | 0.000 | 0.000 | -0.007 | 0.000 | 0.000 | -0.006 | 0.000 | 0.000 |
| **gender** | 0.000 | 0.026 | 0.989 | 0.153 | 0.065 | 0.019 | 0.073 | 0.030 | 0.016 |
| **postpat(-1)** | 0.242 | 0.074 | 0.001 | 0.542 | 0.147 | 0.000 | -0.166 | 0.141 | 0.239 |
| **cumpat(-1)** | 0.057 | 0.031 | 0.061 | 0.011 | 0.044 | 0.801 | 0.408 | 0.096 | 0.000 |
| **expttoma(-1)** | 0.018 | 0.003 | 0.000 | 0.030 | 0.007 | 0.000 | 0.018 | 0.004 | 0.000 |
| **dumyear75_79** | -0.110 | 0.046 | 0.016 | -0.484 | 0.120 | 0.000 | -0.098 | 0.054 | 0.068 |
| **dumyear80_84** | -0.001 | 0.035 | 0.971 | -0.464 | 0.093 | 0.000 | 0.050 | 0.043 | 0.239 |
| **dumyear85_89** | 0.004 | 0.028 | 0.881 | -0.327 | 0.082 | 0.000 | 0.045 | 0.035 | 0.197 |
| **dumyear90_94** | 0.000 | 0.023 | 0.987 | -0.169 | 0.067 | 0.012 | 0.030 | 0.030 | 0.308 |
| **_cons** | -1.742 | 0.041 | 0.000 | -2.252 | 0.094 | 0.000 | -1.694 | 0.048 | 0.000 |
| | | | | | | | | | |
| **rho** | 0.944 | 0.012 | 0.963 | -0.010 | 0.064 | 0.115 | 0.924 | 0.015 | 0.948 |
| **sigma** | 0.804 | 0.021 | 0.847 | 0.498 | 0.017 | 0.534 | 0.786 | 0.021 | 0.828 |
| **lambda** | 0.759 | 0.028 | 0.814 | -0.005 | 0.032 | 0.058 | 0.726 | 0.029 | 0.783 |
| | Wald test indep. eqs.($\rho$=0)<br>$\chi^2(1)$ = 278.83: P > $\chi^2$ =0<br>Number of obs = 14784<br>Log likelihood = -14523.46 | | | Wald test indep. eqns. ($\rho$=0)<br>$\chi^2(1)$ = 0.02: P > $\chi^2$= 0.8745<br>Number of obs = 3433<br>Log likelihood = -2658.351 | | | Wald test indep. eqns. ($\rho$=0)<br>$\chi^2(1)$ = 263.35: P > $\chi^2$ = 0<br>Number of obs = 10707<br>Log likelihood = -11063.57 | | |