Paper to be presented at the DRUID 2012

on

June 19 to June 21

at

CBS, Copenhagen, Denmark,

The organization of industrial R&D and scientific disclosure

Markus Simeth
Ecole Polytechnique Fédérale de Lausanne
CDM - CEMI
markus.simeth@epfl.ch

Stéphane Lhuillery
ICN Business School, Metz, France

Stephane.lhuillery@icn-groupe.fr

Abstract

Many profit orientated companies are not only sourcing knowledge from academic partners but are also active contributors to the stock of scientific knowledge. However, there is only little empirical evidence concerning the nature and organization of research activities which enable these firms to pursue strategies of scientific disclosure. Our paper aims to open the ?black box? of R&D and analyses empirically the relationship between R&D and scientific disclosure along the dimensions scope of research, researcher demographics and modes of knowledge production. The econometric analysis is performed using unique firm-level data from the French R&D survey 2007 and matched scientific publications for a sample of 2.070 firms. We find strong support for the relevance of R&D demographics that represent the abilities and motivations of firm scientists. Interestingly, we find little support for internal upstream research and geographical concentration of R&D as drivers for scientific disclosure whereas external research and funding links are found to have a strong positive impact.

Jelcodes:L20,-

The organization of industrial R&D and scientific disclosure

Preliminary version: please do not quote

This version: May 2012

Abstract

Many profit orientated companies are not only sourcing knowledge from academic partners but are also active contributors to the stock of scientific knowledge. However, there is only little empirical evidence concerning the nature and organization of research activities which enable these firms to pursue strategies of scientific disclosure. Our paper aims to open the "black box" of R&D and analyses empirically the relationship between R&D and scientific disclosure along the dimensions scope of research, researcher demographics and modes of knowledge production. The econometric analysis is performed using unique firm-level data from the French R&D survey 2007 and matched scientific publications for a sample of 2.070 firms. We find strong support for the relevance of R&D demographics that represent the abilities and motivations of firm scientists. Interestingly, we find little support for internal upstream research and geographical concentration of R&D as drivers for scientific disclosure whereas external research and funding links are found to have a strong positive impact.

Keywords: R&D, Scientific disclosure, Industrial Science, Econometric evidence

JEL classification: O3

1 Introduction

Open innovation is a popular paradigm that centers on opening up organizations for creating valuable innovations (Chesbrough, 2003). In this strand of literature, knowledge exchanges among organizations are crucial and firms not only use external knowledge from other organizations to enhance their innovation creation processes but also to commercialize innovations externally for expanding the income potential. However, Open Innovation is often reduced to collaborative innovation agreements and explicit market transactions, whereas there is less emphasis on more subtle mechanism, namely voluntary knowledge disclosure, that may drive and leverage knowledge exchanges (see Dahlander and Gann 2010). Many interesting results on Open Innovation with focus on knowledge inflows (e.g. Laursen and Salter 2006) are close to studies on absorption in the vein of Cohen and Levinthal (1989), where the access to knowledge is not subject to constraints imposed by the institution of origin. In such an environment, voluntary disclosure of own research outcomes is not required and appears counterintuitive since competitors may benefit from spillovers (see Arrow 1963). However, there are rationales to believe that knowledge inflows are not exogenous but can be facilitated by the provision of own research outcomes (von Hippel 1987; Schrader 1991; Lhuillery 2006; Hicks 1995; Penin 2007; Henkel 2009; Simeth and Raffo 2012).

One important dimension of voluntary knowledge disclosure is revealing of R&D outcomes in a scientific format. Here, particular difficulties are to consider that should make scientific contributions challenging for companies: First, firms are directing their research agendas not towards the provision of scientific contributions but on the development of innovative outcomes (Aghion et al. 2008; Dasgupta and David 1994). Second, the publication process may require specific organizational capabilities and resource allocations that are in potential conflict with other firm activities (see Kinney 2004; Cockburn and Henderson 1999). Nevertheless, it has to be acknowledged that firms are considerably contributing to the stock of scientific knowledge (Hicks 1995; Stephan 1996), where the number of firms is increasing over time (Simeth and Raffo 2012) and PhD trained firm scientists are more likely to be author on a scientific publication than inventor on a patent (Sauermann and Stephan 2011). Given the rationales to disclose and its practical occurrence on the one hand and the particular difficulties of scientific disclosure on the other hand, the question arises *how* firms organize for scientific disclosure.

The existing evidence on this question is scarce and focuses on a very limited set of dimensions. First, some authors propose that firms' basic research activities are a strong driver of disclosure (see Rosenberg 1990; Belenzon and Patacconi 2010). However the assumed link between basic research in firms and scientific publications lacks profound empirical foundation. To our knowledge, Adams and Clemmons (2008) provide the only study that examines this relationship, using a sample of the leading 200 American

R&D performers. Apart from the traditional basic-applied research separation, it is widely neglected that R&D activities of firms can also be characterized along additional dimensions, like the operational modes of knowledge production processes or different Human Resource compositions. In this paper, we will address this gap. We contend that firms intentionally allocate and organize R&D resources for being able to conduct scientific research and achieve academic publication of research outcomes. We elaborate original and multidimensional dimensions characterizing the allocation and the organization of R&D. Doing so, we aim at validating different hypotheses on the distinctive traits of publishing firms.

We draw from studies that analyse the impact of certain R&D characteristics on inventive output for the general framing of our study. A key characteristic of firm's R&D organization is the centralization degree, which refers to the concentration of decision power with respect to research agendas or the budget allocation, and where the choice impacts the possibilities of exploiting local knowledge and relative coordination costs (Alonso et al. 2008; Colombo and Delmastro 2004; Argyres and Silvermann 2004; Arora et al. 2011). A related but not necessarily equivalent R&D organization characteristic to centralization is the geographical dispersion of R&D facilities, where dispersed R&D facilities enable for a superior leveraging of local market and scientific knowledge, while coordination and communication costs increase (Leiponen and Helfat 2010; Furman et al., 2006; Singh, 2008). Some authors emphasize the role of researchers on the choice of firm policies on scientific disclosure (Stern 2004; Hicks 1995, Rosenberg 1990; Gans et al. 2011). Scientists moving to industry often value academic norms like the right to publish research outcomes in scientific journals. As a result, firms can leverage researcher motivations for boundary spanning activities in order to access external knowledge. This aspect implies for our study not to overlook researcher characteristics (that represent distinctive abilities and motivations) as determinant of scientific disclosure. An analogy for the organization of boundary spanning activities can also be identified from studies examining commercial activities by universities. Some studies show that the existence, organization structure and certain favorable practices of Technology Transfer offices enable academic researchers to engage in patenting activities (e.g. Azoulay et al. 2007, Bercovitz et al. 2001).

Building on this literature, we propose to characterize corporate R&D along the dimensions scope of research, researcher demographics and modes of knowledge production. The scope of research considers the firms inclination towards upstream research and external research relations, whereas the researcher demographics refer to the personal composition of R&D with respect to academic degree, age, gender and nationality of the researchers and engineers. Our third dimension, the modes of research, is concerned with the use of modern research equipment, the geographical dispersion and the formalization of R&D activities.

Our empirical analysis relies on rich information from the French R&D survey 2007 and matched scientific publication data for a sample of 2.070 R&D performing firms. The results suggest that all three dimensions proposed are relevant for the determination of scientific outcomes. With respect to the nature of research, there is interestingly little evidence for upstream research as a strong predictor for scientific disclosure, while the existence of external research links is found to be an important driver. The composition of R&D in terms of researcher demographics is a significant predictor in the dimensions academic degree, age and interestingly also gender where higher shares of female researchers show a positive relationship to disclosure. This suggests that researchers may play a considerable role in enabling strategies of scientific openness. Against our expectations, the use of modern equipment has as well as the geographical distribution of research facilities are without strong impact on the disclosure intensity.

The remainder of the paper is structured as follows: In Section 2, we provide an in-depth discussion about the relationship between the three R&D dimensions proposed and scientific disclosure. Section 3 describes the data and sample composition as well as the econometric setting. In Section 4, results and implications are discussed whereas conclusions, limitations and propositions for further research are covered in the final Section 5.

2 R&D and scientific disclosure

2.1 Scope of research activities

In order to disclose in a scientific format, firms need to produce research which is eligible for academic publication and of interest for the scientific community. In general, companies have incentives to focus on applied research activities (Aghion et al. 2008). On the other hand it can be observed that some firms are deliberately involved in fundamental research and academic activities since it allows to be at the technological frontier, to absorb external knowledge (Cohen and Levinthal, 1989; Cockburn and Henderson 1998) be connected to scientific networks (Rosenberg, 1990; Pavitt, 1991; Liebeskind et al., 1996; Zucker et al., 2002), and to get public funding (Audretsch and Stephan, 1996). A distinction between upstream research and development can thus be expected to be influential for determining the degree of scientific outcomes[1]: firms involved in upstream research activities are more likely to produce breakthrough knowledge and to be closer to academic research than other firms and thus more likely to be able to publish at lower costs in academic journals. From an empirical point of view, this link between

[1] The symmetry with patent publications does not hold here since fundamental results are more difficult to patent (Pavitt, 1991).

generic research by industry and academic publications is hardly explored. Based on a sample of top 200 U.S. R&D firms, Adams and Clemmons (2008) proposed a production function for industrial science at the firm level. The authors show that firm's past internal basic research is positively related to its current scientific production, suggesting that industrial science is a cumulative activity. Further empirical evidence on the individual scientist level shows that the relationship of the nature of research in terms of basic-, appliedness is indeed a predictor for individual scientists' patent and publication outputs, but only of moderate strength (Sauermann and Stephan, 2011). It is to stress that not only pure basic research might be adequate since many research projects lead to both fundamental insights and applied outcomes (Stokes, 1997; Murray, 2002). For example, in the bio-scientific fields, potential applications can be closely related to fundamental insights and publications. Nevertheless we expect that input allocation to basic research enhance the possibilities for academic publication.

Second, it is also to consider that firms not only publish research that has been created entirely in-house but in collaboration with other institutions, particularly academic research labs or through support of public R&D funding institutions. Given the existence of collaborative links, the basic eligibility of outcomes for publication is expected to be larger (Fontana et al. 2006) and the marginal costs of codification lower, leading to higher publication propensities. Collaborative links with universities or public research institutions like the CNRS indicate that the firm aims to access valuable knowledge from the scientific community. The PROs-industry cooperation is likely to generate more flows of tacit knowledge among bodies but also imposes a certain codification level during the generation, the exchange of knowledge and the monitoring of outcomes (See Cockburn & Henderson, 1998). It is also to consider that the academic partner has specific codification know-how that lowers the publication costs for firms and that the academic partner has an inherent desire for codification since its reward systems root on the disclosure of research outcomes (Dasgupta and David, 1994). Depending on the funding constellation and the resources provided by the firm, universities can have strong negotiation positions (Lacetera 2009) and may insist on the right to publish, leading to co-publications with authors from the university and the firm (see also Simeth and Raffo 2012)[2].

Similarly, the potential aim of firms to access public R&D subsidies is expected to influence the research scope and the chosen codification degree in order to signal the firm's abilities (see Cowan and Foray, 1997; Cohendet and Meyer-Krahmer, 2001). Particularly the access to funding programs with upstream focus should induce more publications since corresponding subsidies are typically provided to firms being

[2] The literature stresses predominantly the negative impact of industry funding on publications of academic researchers (See Czarnitzki et al. (2011) and references therein) than on the positive impact of academic labs on openness of industry science.

at the scientific and technological frontier and the support is also based on the ability to combine industrial and academic knowledge production. Public subsidies can also be considered as a means to alleviate resources constraints of researchers and to allocate some resource slack to exploration tasks (Nohria and Gulati, 1996) and publication. Besides the project scopes of public subsidies with upstream focus, the provision of financial support may be directly related to the disclosure of project outcomes. For instance the subsidies from the European Union framework programs contain explicit disclosure requirements that impose the codification of knowledge (EC, 2003a). Therefore, a positive relationship between external research links or research funding and scientific publication is assumed[3].

H1a: Companies conducting upstream research internally show greater disclosure intensities

H1b: Companies maintaining external research and funding links show greater disclosure intensities.

2.2 R&D demography

The firm decides also about the personal composition in the R&D function in order to fulfil the planned research tasks. The composition reflects the abilities and motivations of the scientists employed not only for creating innovative outcomes but also for generating and publishing scientific outcomes. The successful conduction of research and development is heavily depending on the abilities of the researchers including their ability to interact. R&D requires creative work and the re-combination of existing and generation of new knowledge is done by highly qualified individuals. Connected with the need of the firm to develop new knowledge and inventions, contributions to the knowledge stock of Open Science should depend on the corresponding capabilities and motivations of the firm scientists and engineers (Sauermann and Cohen 2010; Stern 2004). The capacity of teams and especially R&D teams to produce knowledge can be influenced by its composition (Lazear, 1999; Bercovitz and Feldman, 2011) but this dimension is hardly explored for industrial R&D staffs. We propose in the following to focus on industrial R&D teams to investigate more precisely the link between researchers' characteristics defined on their levels of academic background, their gender, their age and the internationalization and their scientific disclosure orientation.

The PhD degree of an industrial researcher is expected to be a strong driver of knowledge transfer including tacit knowledge (Stephan et al., 2004) and networking activities (Dasgupta and David, 1994; Mangematin, 2000; Lam, 2007). However, it does not seem to be acknowledged as an important source of

[3] However if resources are provided by the Department of Defence, openness is expected to be a marginal practice.

industrial R&D project completion (See Cohen et al. 2002). Still, PhD should be a very strong driver of academic knowledge, methods and "culture" in the firm since PhD trained scientists have usually experienced the publication process and spent a sufficient time in academic training and research in order to adopt academic values. Given the desire of many researchers in industry to publish in referred journals, companies can leverage the right to publish as an active HR instrument (Hicks 1995; Ritti 1968; Stern 2004; Sauermann and Roach, 2011). Or expressed in the reverse logic, firms that regard strategies of scientific disclosure as beneficial have an interest in hiring researcher with academic mindset. Besides individual motivations, this is also plausible along the abilities dimension, where scientists and engineers with master degrees should lack the training and experience for the publication process. From an empirical point of view, high-level industrial scientists are indeed found to be the one publishing articles (See Halperin and Chakrabarti, 1987). This point may be even more relevant in France due to its dual system of higher education where the engineers graduated from the "Grande Ecoles", who are usually not trained towards academic research, are usually the one chosen to become industrial researchers (Beret et al., 2004).

With respect to academic research positions, the transition rate of women to higher academic positions (i.e. tenure track) is smaller than for men (Ginther and Kahn 2009), female scholars publish less than male scholars due to lower access to hierarchical positions and resources (Xie and Shauman, 1998) and have lower degrees of involvement in industry relationships (Murray and Graham 2007), inducing lower patenting activities and entrepreneurial activities (Whittington and Smith-Doerr, 2008; Stuart and Ding, 2006). Due to this gender discrimination occurring in PROs, a considerable number of highly qualified female should be available for industrial R&D occupations with potentially higher skills than the average male researcher. However, there are several arguments that suggest an opposite relationship. Literature examining research outcomes in industry suggests that women are less output orientated as patenting activities in industry are disproportionally done by males (Giuri et al. 2007; Whittingthon 2009; see however Ejermo and Jung, 2012 on Swedish teams). Correspondingly, firm publications should be done predominantly by male researchers too. Female industrial researchers should also get lower resources, lower positions, less strategic R&D projects and then therefore have less patent or article opportunities. Moreover, female researchers have a less diverse network (Renzulli et al., 2000), with stronger ties and more feminine ties than men (Ding et al., 2010). On a single firm, Liu and Stuart (2011) find however that publishing industrial researchers have indeed significantly more correspondents in universities but do not find a significant and superior ability of women to create and maintain links with academic researchers[4].

A third demographic component concerns the age composition of the scientists and engineers in a firm where a couple of mechanisms may apply. New graduates are aware of the latest knowledge in their field and have typically less developed an appropriate understanding of business necessities (Allen and Katz, 1992). Moreover, young graduates starting in industry have not necessarily decided if they want to return to the academic sector at a later point in their career. In such a case, keeping an academic track record should have a high importance to them since it keeps future career options open. Apart from this argument, publications can be more simply a good means for young researcher to provide a signal of quality to industrial colleagues and thus to build an internal and external network in a faster way (see Gambardella 1992). The arbitrage between academic and industrial career is however not straightforward and evolve among cohorts. Industrial R&D has become over time more attractive to leading researchers due to the increasing gap between resources available in industry and academic labs (Sauermann and Stephan, 2011; Roach and Sauermann 2010) as well as the scarcity of tenure track positions (Fox and Stephan, 2001). Even if we acknowledge that intrinsically motivated scientists prefer a career in academia, young PhD scientists with high research ability and publication skills should be more prone to pursue their career in industrial firms than some time ago. Finally, young graduates either with a Master of Science or a PhD degree are likely publishing their articles with a time delay and might additionally mention the firm affiliation. A certain share of publications done by firm scientists should thus reflect the rise of scientific publications done by young graduated people. The management literature on the influence of team composition on performances proposes however a more cautious view: age diversity may be also associated with a lower communication and higher conflicts within the R&D group or among cohorts. Zajac et al. (1991) for example find that age diversity is negatively linked to innovation performances. Despite some possible negative effects, we content that academic values as well as publishing opportunities based on the position in firms should be stronger for “younger” R&D teams than for more experienced personnel.

The role of foreign born scientists and inventors has been underlined in the literature on high-tech entrepreneurship (e.g. Hunt and Marjolaine, 2009) and academic production (Stephan 2012). Up to now a

[4] A huge literature on small group psychology suggests that gender composition may influence team performances either positively (See Wood, 1987) or negatively (See Pfeffer, 1985). Combining the two antagonist effects, an inverted U shape is often expected. Empirical evidence are however scarce when knowledge production is considered. Gratton et al. (2007) confirm that teams with gender diversity are more likely to take risks and to innovate. Their results further suggest that gender parity is the best team composition for innovation. Based on 2575 US industrial researchers working in 24 firms, Cordero et al. (1997) failed however to find such an inverted U shape or even any robust significant link between gender and innovation or patents.

mere attention has be paid to industrial R&D teams. A higher proportion of foreign researchers in R&D labs is however the result of active HR strategies to hire international experts. Ethnic minority or foreign researchers are indeed chased by firms in order to acquire complementary knowledge likely to improve productivity (Cordero 1997; Lazear 1999). Basically, foreign based researchers must be more visible than local ones and thus likely have already an academic track record and "taste" for scientific publication. International teams of researchers should also allow for accessing broader scientific communities as they can leverage their personal network. This social capital can be valuable at the input or output level of the publication process as an international network is useful to stay at the scientific knowledge frontier. On the output side, an article with foreign co-authors should be more likely to be cited. Also from a pure linguistic point of view, hiring international researcher may have a positive impact on disclosure. If an English speaking researcher is hired, her bilingualism is costly (Lazear, 1999) but may raise the likelihood to publish in international journals since it is easier to comply with the language requirements of those journals.

H2: Companies occupying greater shares of young, PhD trained and foreign born researchers show greater disclosure intensities, while there is a negative effect of greater shares of female researchers

2.3 Modes of knowledge creation

The utilization of modern equipment and ICT technologies is expected to influence R&D capabilities and the ability to publish results. The adoption of these new technologies brought important improvements in R&D productivity in many industries, for instance in the pharmaceutical (Nightingale, 2000), biotech (Ding, 2010) chemistry (Dodgson, 2006), computer (Thomke, 1998a) or automotive industry (Thomke and Fujimoto, 2000; Vaccaro et al., 2009). With respect to new equipment, three main mechanisms can be identified. First, as argued by Nightingale (2000) for the pharmaceutical sector, craft-based laboratory experiments have been replaced by semi-automatic "high-throughput screening" techniques and visualization tools that allow also for a parallel conduction of experiments. The new equipment thus allows firms to produce knowledge earlier and faster than competitors due to the enhanced screening abilities concerning knowledge combinations and solutions (Rosenberg, 1974; Arora and Gambardella, 1994; Nightingale, 2000; Becker et al., 2005). Second, besides the increasing performance, the implementation of new technologies leads also to a standard labor-capital substitution where previously labor-intensive experiments executed by R&D technicians are replaced by equipment and specific utilization skills of researchers (Baba and Nobeoka, 1998; Nightingale 2000; Bresnahan et al. 2002). Here, the change in labor force composition towards scientifically trained people may also lead to lower communication and coordination costs among researchers that allows for a further specialization. As a consequence, firms are more able to use and explore abstract and science based knowledge (Arora and Gambardella, 1994). Third, the new knowledge production technologies promote also the codification of knowledge (Arora and Gambardella, 1994) at every stage of the R&D process (including inputs, procedures and outputs) and are corresponding to lower costs of codification (see Cowan and Foray 1997; Dasgupta and David, 1994). Therefore, we expect that the utilization of modern equipment increases the ability of the firm to participate in academic publication.

A main internal organizational dimension is the formalization of R&D activities. In general, formalization represents the extent of rules, procedures and instructions that are applied in an organization (Pugh et al., 1968). In the context of R&D, it can be questioned how formal or informal the research activities are structured. For enhancing scientific creativity in firms, flexible and informal structures are often more suitable for generating original scientific ideas and for adjusting to rapid environmental changes (Liebeskind et al., 1996; Gambardella 1995; Sorensen and Stuart 2000). Hierarchical structures on the other hand are more effective for the conduction of routine tasks, for instance clinical trials in pharmaceutical research (Gambardella, 1995). We propose to regard the weight of the R&D administration as sub-dimension of formalization, where the potential influence towards academic publication is ambiguous. On the one hand, administration may disburden scientists from time-consuming routine work that is not directly associated with research, like correspondence with funding authorities or a professional processing of patent applications (Kornhauser 1962). If scientists can focus on their core tasks, the freedom for academic activities should be larger. In this connection, an analogy can be drawn to the Technology Transfer offices in the academic domain which facilitate academic patenting activities by scholars (Azoulay et al. 2007, Bercovitz et al. 2001). On the other hand, a larger R&D administration may also represent a larger bureaucracy and formal procedures where many actions have to be approved explicitly, restricting creativity (see Holstrom, 1989; Cardinal et al. 2004; Sorensen and Stuart 2000). This could target not only questions like project selection or monitoring but also the procedures associated with scientific publication if the administration and legal staff review scientific documents very restrictive based on the aim to avoid or minimize spillovers.

A main organizational arbitrage on R&D is the choice to disperse geographically the R&D facilities or to centralize it in the same geographical area. Geographically dispersed R&D should allow researchers to tap more easily into complementary local knowledge sources in order to adapt products to local market needs (Leiponen and Helfat, 2011) but also access more easily into local academic knowledge sourcing (Furman et al., 2006; Alcazer and Chung, 2007). In the former case, less publications are expected that in the later R&D strategy. Both R&D strategies induce potential additional costs such as search costs (Brandon and

Hollingshead, 2004) or communication costs due to a lack of face-to-face communication, but both strategies are not similar regarding the risks of outgoing spillovers. Leading firms that are more likely to publish articles encounter more risks of leakages than followers. Several works reveal that firms with geographically dispersed R&D facilities implement specific appropriation strategies (Singh, 2007; Alcazer and Chung, 2007; De Faria and Sofka, 2011) or even refrain from decentralizing R&D activities (Kogut and Chang 1991) in order to prevent local outgoing spillovers. Empirical evidence converges towards the finding that geographically dispersed R&D facilities are hampering the number and quality of inventions (Furman et al., 2006; Singh, 2008). Leiponen and Helfat (2011) show that firms with at least two R&D locations in Finland generate more imitative innovations but not radical ones. Based on this empirical evidence, new knowledge should be produced more often in geographically centralized R&D since it allows for coordination and appropriation benefits. However also an alternative view can be established: If firms can benefit from informal knowledge exchanges (see Schrader 1991, von Hippel 1987), they may allow decentralized researchers to disclose outcomes in a scientific format in order to control and document outflows (that inevitable occur on an informal base anyway) more explicitly. Overall, we assume nevertheless a negative link between decentralization and disclosure.

> H3: Companies relying more strongly on modern equipment and ICT reveal greater disclosure intensities, whereas geographically dispersed R&D and stronger formalization of R&D leads to less disclosure

3 Data and methodology

3.1 Data sources

The firm level information is utilized from the French annual R&D survey (Statistical department of the French Ministry of Higher Education and Research) of the year 2007 and its extension “chercheur” covering the demographics of researchers and engineers. The R&D dataset disaggregates the R&D expenditures along multiple dimensions and includes also information about the origin of the R&D budget. In principal, the R&D survey in France is exhaustive for firms with an internal R&D spending of more than 250,000 Euro. The link between individual firms and business groups is made using the Liaison Financière (LIFI) dataset provided by INSEE. This dataset contains information about the different affiliates within a business group[5]. Our analysis includes both manufacturing and service sectors, however

[5] A firm is considered in LIFI to be controlled if an owner holds 50% of the shares. In the year 2007, 261,869 firms located in France were listed to belong to 41,837 business groups.

we remove the industry of R&D services (two digit NACE classification “73” in version 1.1.), all start-ups, R&D affiliates from business groups, industry-shared labs (Centre techniques) and R&D service firms on the identifying characteristic of a R&D/Sales ratio larger than 0.5. All in all, the final sample contains 2.070 firms.

The second main data source are the scientific publications retrieved from Elsevier’s Scopus, which represents the most comprehensive database available that reveals an even better coverage than the also commonly used Thomson Web of Science database (Flanagas et al., 2008). Before downloading the publication information, we cleaned carefully the firm names in order not to miss any correct hits. We removed the legal endings of the firms (SA, SARL etc.) and checked the names for spelling mistakes. Moreover, if the firm names contained simultaneously abbreviations and full name, both have been queried separately. On the other hand, after the retrieval and matching process, we applied automatic cleaning procedures to remove incorrect matches and duplicates per firm (e.g. if authors mentioned both full name and abbreviation or in the case that two sample firms are co-authoring a publication). A common source of error is academic institutions with an identical name component like the firm of interest. These false positive hits have been removed by an automatic cleaning procedure using a diverse multilingual set of name terms indicating academic institutions. After the automatic database cleaning, we screened in a time-consuming process all remaining firm-publications allocations manually in order to ensure a high quality matching. Thanks to the manual cleaning, we identified and solved several ambiguous cases, for instance where two firms with equal or similar names (that often belong to the same group) received the same or overlapping matches. Here, we used the address information of the firms (including the addresses of known affiliates) and the scientific publications to verify the correct allocation. In a few cases, could not solve the ambiguities which led us to fully exclude those firms.

3.2 Variables

To measure the relative orientation of a firm towards scientific openness, we use the number of publications and related it to the amount of total R&D expenditures. Since one can expect a time-lag between the knowledge creation and the disclosure in a referred scientific journal or proceeding, the publications of the post-survey period 2008 and 2009 are used to compute our measure. Moreover, only articles and conference proceedings have been considered since only these represent original research contributions.

For performing the robustness tests on inventive outcomes measured by patent applications, the self-declared number of EPO patent applications has been chosen and is basically computed in an equivalent way than the publication variable. However the reference year is only 2008 since our patent information is incomplete for later years. For these robustness tests, also the publications are restricted to the year 2008 in order to ensure a meaningful comparison. With respect to the patent variable it is to stress that the measure is self-declared by respondents which might be subject to bias. On the other hand it might contain correct values that are not visible in patent statistics if patents are assigned to the parent company (if firm is part of a group) instead of the firm that actually performs the research. All in all, we are convinced that our measure of inventive performance is valid for our purpose of performing a meaningful robustness test.

Our explanatory variables capture the dimensions scope of research, R&D demographics and modes of knowledge creation. The R&D demographics (Doctoral degree, Age, Internationality, Female) represent the respective shares in relation to the total number of researchers and engineers in a firm. For instance if four out of ten researchers in a firm have a doctoral degree, the corresponding variable receives the value 0.4. Since researchers with a PhD degree and those with a doctor title in medical fields may show a different research orientation, we distinguish these groups and introduced separate shares. For distinguishing the age variable between young researchers and the comparison group, the age margin of 35 years is imposed.

With respect to the scope of research, we introduce measures to account for the dimensions internal upstream research and external research links. The first dimension is measured with a dummy variable that captures if a firm is spending (internal) R&D in upstream research. Even though the dataset provides the respective shares of the internal R&D budget that is directed to basic, applied and development R&D, we include only a dummy variable since the shares might be subject to measurement error as the boundaries between these dimensions are not necessarily clear (see Stokes 1997). For reflecting the relative influence of external research links, the respective shares of external expenditures directed to upstream/downstream partner in relation to total R&D expenditures are computed[6]. Similarly, we have relative measures that capture the importance of public R&D funding coming from institutions that focus by agenda on upstream or downstream research[7].

The third dimension examined is modes of knowledge creation. Our measurement cover here the use of equipment, the relative formalization of R&D processes and the geographical dispersion of R&D. Concerning the first, we do not have an exact measurement about the value of equipment but can only rely on a proxy. Following our theoretical discussion in section 3.3, we assume that the use of modern equipment requires scientific and engineering personnel with advanced application skills and formal

---

[6] Typical upstream research links include universities and public research institutions that focus on fundamental research, most notably the CNRS (Centre de la Recherche Scientifique).

[7] For instance research funding originating from the European Framework programs is typically directed to R&D activities that are conducted with academic institutions.

training, which should be reflected in a reduced reliance on technicians that perform craft-based work. Concerning the formalization, we employ the proxy of the relative size of R&D administration in relation to the number of R&D performing scientists and engineers (R&D administration/scientist ratio). Our measure concerning the geographical dispersion of R&D is a standard Herfindhal-index that captures the concentration of researchers among different geographical departments in France, based on the actual researcher numbers.

In order to control for firm level heterogeneity, several controls are included. Obvious control variables are the R&D intensity (Log of total R&D divided by employee number) and the firm size (log of employees) since they control for the overall size and relative importance of the research activities. We also include a dummy variable that captures if a firm runs a specialized research centre. Since research centres can be very heterogeneous in terms of autonomy, decision making processes or scope of research, we only use it as a control variable. Moreover, we use information about the financial links to consider if a firm belongs to a French or foreign business group. To control for the dynamics of different research fields, our models contain standard industry controls computed on the NACE-2 level. Appendix A.1 provides an overview about all variables used in the analysis.

3.3 Descriptive statistics

From the firms in our sample, 18.4% have at least one scientific publication in the estimated post-survey period 2008/2009. The high level of publishing companies is in line with previous works underlying the important share of academic publications coming from industrial scientists (Stephan, 1996; Godin, 1996) and the high percentage of industrial researcher publishing (Whittington, 2009; Sauermann and Stephan 2011). The Table 1 presents descriptive statistics concerning all variables used in the analysis where we computed also the means separately for publishing and non-publishing firms. With respect to the R&D demographics, it can be seen that the share of foreign researcher is small and female researchers are underrepresented since they just account for 20% of all researcher, whereas young researcher account for 41% on average. However it is interesting that female researchers have a larger share in firms that publish where they account in average for 28% of the researchers, while only for 18% in non-publishing firms. Concerning researcher with doctoral degree (PhD and Medical Doctors), there is a large difference between publishing and non-publishing firms since their average share accounts for 25% for the former group while only 11% for the non-publishing ones.

Only a moderate difference can be seen on the question if a firm directs internal R&D towards upstream research, where 86% of the publishing firms and 81% of the non-publishing ones perform upstream research internally. With respect to the technician ratio that is used to proxy for the use of modern equipment, it can be seen that publishing firms have remarkable lower ratios (0.79 vs. 1.18). An exhaustive overview about all variables is provided in the Table 1.

-- Insert Table 1 about here --

In Appendix A.2 the univariate correlations of all variables are shown. The correlations are not problematic for the interpretation of our results since we did not detect any high correlation. Therefore, the correlations are not further discussed.

3.4 Econometric model

In the empirical analysis, the relative degree of scientific disclosure of a firm is estimated as a function of scope of research, R&D demographics, modes of knowledge creation and a set of controls. As outlined in section 4.1, our dependent variable is an intensity measure which relates the number of scientific contributions to the firm size measured by employee number. For all estimations, a time lag is imposed between the firm-level variables and the scientific publications since it is plausible to assume a timely gap between the actual knowledge creation and the appearance of the outcomes in a scientific journal or conference proceeding. We employ several models in order to get a comprehensive view about the robustness of results. Since 82% of our firms in the sample have no single publication in the estimated period 2008/2009, we employ first a standard Tobit model that accounts for the left-censored nature of the dependent variable. As alternative specifications, we use also a selection model that more explicitly accounts for the binary decision to publish and a count data model for estimating the absolute number of scientific contributions. Moreover, all estimations are performed with robust standard errors to avoid inconsistent estimations due to heteroskedastic errors. For testing the robustness of our results, we tried alternatively to cluster the standard errors by business-group-id in order to account explicitly for any biasing effects imposed by the heterogeneity at the business group level, which we are unable to observe. However, since the results remain unchanged, we do not report this specification here. We further propose a bivariate Tobit model to account for the possibility of an interdependent determination of scientific openness and inventive outcomes. This model should also provide insights if scientific disclose requires distinct organizational capabilities. With respect to the implementation, it is to stress that the patent data is incomplete for later years than 2008. Therefore, the two outcome variables are restricted to the year 2008. The results of our econometric models are discussed in the sections 4.2 and 4.3.

4 Results and discussion

4.1 Intensity of scientific disclosure

The results of the regression models are depicted in Table 2, where a Tobit model (1), a selection model (2a & 2b), and a count data model (3) are presented. The equations (4a) and (4b) are commented in the following section 4.2. Accordingly to our theoretical discussion we have focused on the three dimensions scope of research, R&D demographics and modes of knowledge creation.

-- Insert Table 2 about here --

With respect to the scope of research, we find surprisingly no evidence that upstream research leads to higher degrees of scientific openness (H1a). Even though the coefficient has a positive sign in all specifications, it is never significant. In order to ensure robustness of this finding, alternative measures have been tested. We introduced a dummy variable indicating if the firm performs any basic research (instead of upstream that contains basic and applied research) and we also computed two ratios of research vs. development and basic research vs. applied research and development. However, also these alternative measures have no statistical significant effect on scientific disclosure. This finding supports the view that scientific publications created by firms are more frequently originating from research that can be placed into the "Pasteur's quadrant" (using the classification of Stokes 1997) rather than pure fundamental endeavours. Moreover, it is also an indirect sign that scientific publications are not only targeted to academic stakeholder but also used for realizing further signalling benefits. However it is to consider that external research links to upstream partners lead to higher degrees of scientific openness as the models quite consistently show (H1b). This finding is in line with our assumptions that external links are associated with a different nature of research projects, which are inherently more codified since academic collaboration partner have a strong interest to publish outcomes. The result is also consistent with the findings of Simeth and Raffo (2012) about the impact of concessions and reciprocal behaviour, even though our measurement has the limitation that it is based on the existence of financial resources provided by the firm. Nevertheless, as one could expect that the negotiation position of firms improves once the financial resource provision increases, the positive and significant effect of upstream links provides further support that reaching leading academic institutions requires additional concessions. Similarly, there is a positive and significant effect of public R&D funding. Interestingly, there is a positive effect not only from funding originating from institutions that have typically an upstream focus but also from those with downstream focus. Therefore, it seems that the positive impact of public funding is not only driven by the research scope imposed by the funding but also increased requirements of codification which should also be valid in the case for downstream funding.

Concerning the R&D demographics, firms with higher shares of PhD researchers and medical doctor degree have a positive significant impact on scientific disclosure. This finding supports our assumptions concerning the impact of academic socialization, as those groups should have a stronger desire ("taste for science") to keep links with the scientific community. According to our theoretical discussion, also firms with higher shares of young researchers should have more scientific contributions since they have the latest academic knowledge and have been less exposed to the requirements of business firms. This prediction receives also strong empirical support since our variable is significant across all models, where the variable is only in the continuous part of the selection model not significant at the 5% level but on the 10% level only. Interestingly, we find also a relative stable and positive effect of higher shares of female researchers which is contradictory to our derived prediction from literature. Given the (limited) literature concerning female researchers in industry, we would have predicted a negative effect. At this point, we can only speculate about an explanation for this effect, but possible reasons could lie in a higher average academic socialization of female researchers in industry (due to discrimination for higher positions in academia) or in a task separation in industry where male researchers focus more on the creation of inventive outputs. Another and more trivial explanation could be based on sector heterogeneity that is only imperfectly captured by our controls. Firms with higher shares of female researchers might reflect (sub-)sectors that are more prone to publish due to other reasons. Nevertheless, this finding proposes in any case to examine gender issues in industrial research more explicitly. Finally, there is not much evidence for a greater openness if firms occupy more researchers with international experience since our variable is only significant in the count model (3). We introduced also alternatively a measure that is based on international study degrees instead of the nationality, but the results reveal an identical picture and are therefore not reported here (available upon request). All in all, our hypothesis H1 finds support in the dimensions academic degree and age, whereas the shares of female and international researchers are not along our predictions. Given these results, it can be emphasized that the personal composition is an important driver for scientific outcomes, reflecting the impact of scientists' abilities and motivations for the possibilities of firms to provide scientific contributions.

The third and last dimension examined is the modes of knowledge creation (H3). Here, we find in the specifications (2) and (3) a positive and significant effect of higher technician-scientist ratios that should reflect the use of less sophisticated equipment. However, given the limitation of this particular measurement, one has to be very cautious in interpreting the result which requires further empirical work. With respect to the relative size of the R&D administration that represents some formalization of research activities, we do not find any significant effect on scientific disclosure. Concerning the geographical dispersion of R&D facilities, we find interestingly no statistically significant effect on scientific openness. Previous literature proposes that centralized R&D is producing more often original ("radical") innovative outcomes, which would lead us to expect a similar effect on scientific outcomes.

The control variables R&D intensity and firm size behave along the (implicit) expectations and show across models a consistent positive significant relationship towards the scientific openness intensity, with the only exception at the continuous part of the selection model (2b). The implementation of a specialized research centre is found to be without impact on scientific disclosure which proposes a large heterogeneity in the design of research centres across firms.

4.2 Scientific vs. inventive outcomes

In this subsection, we address the question if strategies of scientific disclosure require distinctive organizational capabilities that are not necessary for creating inventive outcomes. For this purpose, a bivariate Tobit model is estimated using the patent intensity as second output measure. Since we only have the patent numbers for the year 2008, both intensity measures (scientific publications and patents) are restricted to this year. Moreover, the sample size decreases slightly as we do not have patent information for the full sample (1.767 instead of 2.070 observations). The results of these regressions are presented in the columns (4a) and (4b) in the table 2.

We find a significant and positive effect of greater shares of PhD trained, young and female researchers on the scientific disclosure intensity, while there is no effect of these shares on the patent intensity. A formal Wald-test is significant in the dimensions PhD trained and female researchers (both at 1% level). Based on the cross-sectional nature of our dataset, we have to be very cautious in terms of causal interpretations, but the result suggests that firms with the desire to publish in a scientific format need a distinctive R&D workforce composition in order to serve the scientific dimension. Concerning the scope of research, there is no effect for external research and funding links for a higher intensity of inventive outcomes. Also the conduction of upstream research remains without effect. With respect to the modes of knowledge creation, larger technician-scientist ratios have both for scientific and inventive outcomes a negative and significant sign, supporting the view that technicians are representing an R&D environment that relies less on cutting-edge equipment but is more craft-based. Greater R&D administration-scientist ratios have a positive effect on the patent equation, potentially reflecting organizational support for the professional filing of patent applications. However there is a negative effect for the publication intensity and the formal Wald test is significant, indicating some trade-off between scientific and technical outcomes. Finally, higher R&D

intensities and larger firm size have a positive impact both on scientific and inventive outcomes. Concerning the latter however, firm size is more influential for determining the scientific disclosure intensity (Wald test significant at 5% level).

Even though there are some dimensions that have a similar impact for both outcomes, there is some support for the assumption that scientific disclosure requires at some specific organizational capabilities and resource allocations, supporting the view that publication activities are not just a simple by-product.

4.3 Robustness tests

In order to learn about the robustness of our results, several additional tests have been performed. In Appendix A.3, intensity and count data models are shown which estimate scientific output measures building on the fractional publication output of firms (see model (5) and (6)), impact-factor weighted publication outputs (see (7) and (8)) and publication outputs without considering conference proceedings but only articles (see (9) and (10)).

There are different rationales for these dependent variables. First, fractional publication counts should reflect more precisely the work effort a firm contributed to a scientific publication. Second, an impact-factor weighted output measure accounts for the possibility that a higher effort is necessary to achieve sophisticated scientific performance and to reach high-quality journals. Third, considering only articles may account for the possibility that conference proceedings follow different pattern, e.g. what the required resource commitment concerns.

Interestingly, our findings remain very robust across all estimated models. Higher shares of PhD trained, international, young and female researchers are associated with higher disclosure intensities. Surprising is however the finding that upstream research does also not have an influence for determining impact-factor weighted publication outputs. This indicates that the boundaries between research and development can be quite fuzzy, where also more applied work can lead to interesting scientific findings (see Stokes, 1997).

5 Conclusion

This paper analyses the origins of scientific contributions provided by profit-orientated firms. In the analysis, the R&D function of firms is examined along the dimensions scope of research, researcher demographics and modes of knowledge creation. Using a unique firm-level dataset which contains detailed information about R&D performing firms in France and matched scientific publication data, we provide novel insights to the scarce empirical literature about firms that pursue strategies of "Open Science". As we consider firm level information which also contains aggregated researcher level information, we provide a complementary view towards recent studies that examine the origin of research outcomes on the individual scientist level (Sauermann and Stephan 2011; Sauermann and Cohen 2010). Moreover, scholarly research about science-based firms focused so far strongly on the wider life sciences industries whereas our study provides a more comprehensive picture by considering a broad range of manufacturing and service sectors.

The findings suggest that the abilities and motivations of firm scientists play an important role in determining the possibilities of firms to disclose in a scientific format. Apart from the question if researchers have a strong motivation to publish or if they need to get incentive, our results show that the scientific training of the firm scientists is crucial for achieving scientific outputs. This underlines the strong position of researchers and engineers in R&D performing organizations. Moreover, we show evidence that the scope of research is a strong predictor for scientific disclosure, however interestingly not in the dimension of internal upstream research. Concerning the modes of knowledge creation, we find only limited support for the equipment dimension but interestingly no effect of geographical concentration or formalization of research. However overall, our study proposes to consider not only the nature of research in terms of basicness but also further dimensions for future scholarly research on innovative firms.

Moreover, our equations in section 4.2 provide some evidence that the creation of scientific and technical outcomes may follow different pattern. This finding proposes that the creation of scientific outcomes in the corporate domains is associated with some particular costs and resource endowments. This result has implications for firm managers that are concerned about disclosure strategies for instance in the context of Open Innovation, since the design of the R&D function has to be adapted accordingly.

Our paper is not free of limitations. Even though a comprehensive and representative firm-level dataset is exploited, we have to be very cautious in terms of causal interpretations since we have only cross-sectional data. Therefore, further work is required to validate our findings. Further research might also continue on this direction and aim to capture more explicitly incentive systems which firms might have introduced for the firm scientists (see Cockburn et al. 1999). Since this information is hardly available from existing large-scale datasets, there is an opportunity to add valuable insights with qualitative studies.

References

Adams, J., Clemmons, R., 2008. The origins of industrial scientific discoveries. NBER Working Paper No. 13823. National Bureau of Economic Research, Cambridge

Aghion, P., Dewatripont, M., Stein, J.C., 2008. Academic freedom, private-sector focus, and the process of innovation. RAND Journal of Economics 39 (3), 617-635.

Alcácer J., Chung W., 2007. Location Strategies and Knowledge Spillovers, Management Science, Vol. 53, No. 5, May, pp. 760–776.

Allen, T.J., R. Katz. 1992. Age, education and the technical ladder. *IEEE Transactions on Engineering Management* 39 (3) 237-245

Argyres, N.S., Silverman, B.S., 2004. R&D, Organization structure, and the development of corporate technological knowledge. Strategic Management Journal 25 (8-9), 929-958.

Alonso, R., Wouter, D., Matouschek, N., 2008. When does coordination require centralization? The American Economic Review 98 (1), 145-179.

Arora, A., Belenzon, S., Rios, L., 2011. The Organization of R&D in American Corporations: The determinants and consequences of decentralization. Working paper

Arora, A., Gambardella, A., 1994. The changing technology of technological change: general and abstract knowledge and the division of innovative labour. Research Policy 23(5), 523–532.

Arrow, K.J. 1963. Economic welfare and the allocation of resources for innovation. In: The rate and direction of inventive activity: Economic and social factors, 609-25, Princeton University Press, Princeton.

Audretsch, D., Stephan, P., 1996. Company-Scientist locational links: The case of biotechnology. The American Economic Review 86 (3), 641-652.

Azoulay, P., Ding, W., Stuart, T., 2007. The determinants of faculty patenting behavior: Demographics or opportunities. Journal of Economic Behavior & Organization (63), 599-623.

Baba, Y., Nobeoka, K. 1998. Towards knowledge-based product development: the 3-D CAD model of knowledge creation. Research Policy, 26(6): 643-659.

Becker, M.C., Salvatore, P., Zirpoli, F., 2005. The impact of virtual simulation tools on problem-solving and new product development organization. Research Policy 34 (9), 1305–1321.

Belenzon, S., Patacconi, A., 2010. Performance, firm size, and the basicness of research. Working paper.

Berkovitz, J., Feldman, M., 2011. The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. Research Policy 40 (1), 81-93

Bercovitz, J., Feldman, M., Feller, I., Burton, R., 2001. Organizational structure as a determinant of academic patent and licensing behavior: An exploratory study of Duke, Johns Hopkins, and Pennsylvania State Universities. Journal of Technology Transfer 26, 21-35.

Béret P., Giret J.F., Recotillet I. (2004) « L'évolution des débouchés professionnels des docteurs: les enseignements de trois enquêtes du Céreq », *Education et Formations, n°67,* pp.109-116.

Brandon, D.P., Hollingshead, A.B., 2004. Transactive memory systems in organizations: Matching tasks, expertise, and people. Organization Science 15(6) 633–644.

Bresnahan, T.F., Brynjolfsson, E., Hitt, L.M., 2002. Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. The Quarterly Journal of Economics 117 (1), 339-376.

Cardinal, L.B., Sitkin, S.B., Long, C.P., 2004. Balancing and rebalancing in the creation and evolution of organizational control. Organization Science 15 (4), 411-431.

Chesbrough, H., 2003. Open Innovation – The New Imperative for Creating and Profiting from Technology, Harvard Business School Press, Boston.

Cockburn, I.M., Henderson, R.M., 1998. Absorptive capacity, coauthoring behavior, and the organization of research in drug discovery. The Journal of Industrial Economics XLVI (2), 157-182.

Cockburn, I.M., Henderson, R.M., Stern, S., 1999. Balancing Incentives: The tension between basic and applied research. NBER Working Paper 6882, National Bureau of Economic Research, Cambridge.

Cohen, W., Levinthal, D., 1989. Innovation and Learning: The two faces of R&D. The Economic Journal 99 (397), 569-596.

Cohendet, P., Meyer-Krahmer, F., 2001. The theoretical and policy implications of knowledge codification. Research Policy 30, 1563-1591

Cohen, Wesley, Richard Nelson and John Walsh (2002). "Links and Impacts: The Influence of Public Research on Industrial R&D." *Management Science* 48(1) 1-23.

Colombo, M.G., Delmastro, M., 2004. Delegation of authority in business organizations: An empirical test. The Journal of Industrial Economics 52 (1), 53-80.

Cordero, R., DiTomaso, N., Farris, G.F., 1996. Gender and race/ethnic composition of technical work groups: relationship to creative productivity and morale. Journal of Engineering and Technology Management 13, 205-221.

Cowan, R., Foray, D. (1997) The economics of codification and the diffusion of knowledge. Industrial and Corporate Change, 6, 595–622.

Czarnitzki D., Grimpe C., Toole A. 2011. Delay and Secrecy: Does Industry Sponsorship Jeopardize Disclosure of Academic Research? ZEW, Discussion Paper No. 11-009.

Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Research Policy 23 (5), 487-521.

Dahlander, L., Gann, D.M., 2010. How open is innovation? Research Policy 39, 699-709.

De Faria, P., Sofka, W., 2010. Knowledge protection strategies of multinational firms--A cross-country comparison, Research Policy, 39(7) 956-968.

Ding, W., Levin, S., Stephan, P., Winkler, A., 2010. The impact of information technology on academic scientists' productivity and collaboration pattern. Management Science 56 (9), 1439-1461.

Dodgson, M., Gann, D., Salter, A. (2006), 'The Role of Technology in the Shift Towards Open Innovation: The Case of Procter & Gamble,' R&D Management, 36(3), 333–346.

EC (European Commission). 2003a. Model Contract for 6th Framework Program, Annex II.34.2. WebURL:*http://ec.europa.eu/research/fp6/model-contract/pdf/annex-ii-general-conditions_en.pdf.* Last access: 15.11.2011.

Ejermo, O., Jung, T., 2012. Demographic patterns and trends in patenting: Gender, age, and education of inventors. Working paper.

Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., Pappas, G., 2008. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. The FASEB Journal 22 (2), 338-342.

Fontana, R., Geuna, A., Matt, M., 2006. Factors affecting university-industry R&D projects: The importance of searching, screening and signaling. Research Policy 35 (2), 309-323.

Fox, M.F., Stephan, P.E., 2001. Careers of young scientists. Social Studies of Science 31 (1), 109-122.

Furman, J., Kyle, M., Cockburn, I., Henderson, R., 2006. Public and private spillovers, location & the productivity of pharmaceutical research. NBER Working Paper w12509

Gambardella, A., 1992. Competitive advantages from in-house scientific research: The US pharmaceutical industry in the 1980s. Research Policy 21, 391-407.

Gambardella, A., 1995. Science and Innovation. The US pharmaceutical industry during the 1980s. Cambridge University Press.

Gans, J., Murray, F., Stern, S., 2011. Contracting over the disclosure of scientific knowledge: Intellectual property protection and academic publication. SSRN Working paper 1559871.

Ginther, D., Kahn, S., 2009. Does Science promote women? Evidence from Academia 1973-2001 in: Freeman, R., Goroff, D. (eds): Science and Engineering careers in the United States: An analysis of Markets and Employment. National Bureau of Economic Research.

Giuri, P., Mariani, M., Brusoni, S., Crespi, G., Francoz, D., Gambardella, A., Garcia-Fontes, W., Geuna, A., Gonzales, R., Harhoff, D., Hoisl, K., Le Bas, C., Luzzi, A., Magazzini, L., Nesta, L., Nomaler, Ö., Palomeras, N., Patel, P., Romanelli, M., Verspagen, B., 2007. Inventors and invention processes in Europe: Results from PatVal-EU survey. Research Policy 36, 1107-1127.

Godin, B., (1996) Research and the practice of publication in industries. Research Policy 25, 587-606.

Gratton L., 2007. Innovative Potential: Men and Women in Teams. Lehman Brothers Centre for Women in Business, LSE

Fox, M.F., Stephan, P.E., 2001. Careers of young scientists: preferences, prospects and realities by gender and field. Social studies of Science 31, 109-122.

Halperin, M.R., Chakrabarti, A.K., 1987. Firm and industry characteristics influencing publications of scientists in large American companies. R&D Management 17 (3), 167-173.

Henkel, J. 2009. Champions of revealing: the role of open source developers in commercial firms. Industrial and Corporate Change, Oxford University Press, vol. 18(3), pages 435-471, June.

Hicks, D., 1995. Published papers, tacit competencies and corporate management of the public/private character of knowledge. Industrial and Corporate Change 4 (2), 401-424.

Holstrom, B., 1989. Agency costs and Innovation. Journal of Economic behaviour and Organization 12, 305-327.

Hunt, J., Marjolaine, G.L., 2009. How much does immigration boost innovation. IZA discussion paper No. 3921.

Kinney, A.J., Krebbers, E., Vollmer, S.J., 2004. Publications from industry. Personal and Corporate Incentives. Plant Physiology 134 (1), 11-15.

Kogut, B., Chang, S. J., 1991. Technological capabilities and Japanese foreign direct investment in the United States. Review of Economics and Statistics 73(3) 401–413.

Kornhauser, W., 1962. Scientists in Industry – Conflict and Accomodation. University of California Press, Berkeley and Los Angeles.

Lazear, E.P., 1999. Personnel economics: Past lessons and future directions. NBER working paper w6957.

Lacetera, N., 2009. Different missions and commitment power in R&D organizations: Theory and evidence on university-industry alliances. Organization Science 20 (3), 565-582.

Lam, A. (2007) Knowledge networks and careers: academic scientists in industry-university links, Journal of Management studies, 44(6), 993-1015.

Laursen, K., Salter, A., 2006. Open for innovation: The role of openness in explaining innovation performance among U.K. manufacturing firms. Strategic Management Journal 27 (2), 131-150.

Leiponen, A., Helfat, C.E., 2011. Location, Decentralization, and Knowledge Sources for Innovation. Organization Science 22(3): 641-658.

Liebeskind, J.P., Zucker, O., Brewer, M., 1996. Social networks, learning and flexibility: Sourcing scientific knowledge in new biotech firms. Organization Science 7 (4), 428-443.

Liu and Stuart (2010), Boundary spanning in a for-profit research lab: An exploration of the interface between commerce and academe Working Paper 11-012, Harvard Business School.

Lhuillery, S., 2006. Voluntary technological disclosure as an efficient knowledge management device: an empirical study. Economics of Innovation and New Technology 15 (4/5), 465-491.

Mangematin, V. (2000), "Ph.D. job market: professional trajectories and incentives during the Ph.D.", *Research Policy*, 29 (6), pp. 741-756.

Murray, F., 2002. Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. Research Policy 31 (8-9), 1389-1403.

Murray, F., Graham, L., 2007. Buying science and selling science: gender differences in the market for commercial science. Industrial and Corporate change 16 (4), 657-689.

Nightingale, P., 2000. Economics of scale in experimentation: Knowledge and technology in pharmaceutical R&D. Industrial and Corporate Change 9 (2), 315-359.

Nohria, N., Gulati, R., 1996. Is slack good or bad for innovation? The Academy of Management Journal 39 (5), 1245-1264.

Pavitt, K., 1991. What makes basic research economically useful? Research Policy 20, 109-119.

Penin, J., 2007. Open knowledge disclosure: An overview of the evidence and economic motivations. Journal of Economic Surveys 21 (2), 326-347.

Pugh, D.S., Hickson, D.J., Hinings, C.R., Turner, C., 1968. Dimensions of Organization Structure. Administrative Science Quarterly 13 (1), 65-105.

Renzulli, Linda A., Howard Aldrich and James Moody. 2000. "Family Matters: Gender, Networks, and Entrepreneurial Outcomes." Social Forces 79: 523-46.

Ritti, R. 1968. Work Goals of Scientists and Engineers. *Industrial Relations: A Journal of Economy and Society* 7(2) 118-131

Roach, M., Sauermann, H., 2010. A taste for science? PhD scientists' academic orientation and self-selection into research careers in industry. Research Policy 39, 422-433.

Rosenberg, N . 1974, *Science, innovation and economic growth*, The Economic Journal 84, 90-108.

Rosenberg, N., 1990. Why firms do basic research (with their own money). Research Policy 19 (2), 165-174.

Sauermann, H., Cohen, W., 2010. What makes them tick? Employee Motives and Firm Innovation. Res 65 (12), 2134-2153.

Sauermann, H., Roach, M., 2011. Not all scientists pay to be scientists: Heterogeneous preferences for publishing in industrial research. SSRN Working paper 1696783.

Sauermann, H., Stephan, P. E., 2011. A multidimensional view of industrial and academic science. Working paper.

Schrader, S., 1990. Informal technology transfer between firms: Cooperation through information trading. Research Policy 20, 153-170.

Simeth, M., Raffo, J., 2012. What makes companies pursue an Open Science strategy? EPFL Working paper No. 167693.

Singh, J., 2007. Asymmetry of knowledge spillovers between MNCs and host country firms. Journal of International Business Studies 38, 764-786.

Singh, J. 2008. Distributed R&D, cross-regional knowledge integration and quality of innovative output. Research Policy 37(1) 77–96.

Sorenson, J.B., Stuart, T.E., 2000. Aging, Obsolescence, and Organizational Innovation. Administrative Science Quarterly 45 (1), 81-112.

Stephan, P.E.,1996. The Economics of Science. Journal of Economic Literature 34 (3), 1199-1235.

Stephan, P., Sumell, A., Black, G., Adams, J., (2004).”Doctoral Education and Economic Development: The Flow of New Ph.D.s to Industry”. *Economics Development Quarterly*, 18(2), 151-167.

Stephan, P.E., 2012. How Economics shapes science. Harvard Press.

Stern, S., 2004. Do Scientists pay to be scientists? Management Science 50 (6), 835-853.

Stokes, D.E., 1997. Pasteur’s quadrant: Basic science and technological innovation. Washington DC.

Stuart, T.E., Ding W.W., 2006. When do scientists become entrepreneurs? The social structural antecedents of commercial activity in the academic life sciences. American Journal of Sociology 112 (1), 97-144.

Thomke, S.H., 1998. Managing experimentation in the design of new products. Management Science 44 (6), 743-762.

Thomke, S.H., Fujimoto, T. 2000. The Effect of "Front-Loading" Problem-Solving on Product Development Performance. Journal of Product Innovation Management, 17: 128-142.

Vaccaro, A., Veloso, F., Brusoni, S., 2000. The impact of virtual technologies on knowledge-based processes: An empirical study. Research Policy 38, 1278-1287.

Von Hippel, E., 1987. Cooperation between rivals: Informal know-how trading. Research Policy 16, 291-302.

Whittington, K., Smith-Doerr, L., 2005. Gender and commercial science: Women’s patenting in the life sciences. Journal of Technology Transfer 30, 355-370.

Whittington, K., 2009. Patterns of male and female scientific dissemination in public and private science. in: Freeman, R., Goroff, D. (eds): Science and Engineering careers in the United States: An analysis of Markets and Employment. National Bureau of Economic Research.

Xie, Y., Schauman, K.A., 1998. Sex differences in research productivity: New evidence about an old puzzle. American Sociological Review 63, 847-870.

Zajac, E., Golden, B.R., Shortell, S.M., 1991. New organizational forms for enhancing innovation: the case of internal corporate joint ventures. Management Science 37 (29, 170-184.

Zucker, L.G., Darby, M.R., Armstrong, J.S., 2002. Commercializing Knowledge: University Science, Knowledge Capture, and Firm performance in biotechnology. Management Science 48 (1), 138-153.

Tables

Table 1 – Descriptive statistics

	Full Sample		Non-publishing firms		Publishing firms	
N	**2,070**		**1,689**		**381**	
Variable	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
Pubcount0809	2.11	19.37	0.00	0.00	11.47	43.98
PhD	0.10	0.20	0.08	0.19	0.19	0.23
MedDoc	0.03	0.12	0.02	0.11	0.06	0.16
International	0.06	0.17	0.05	0.17	0.06	0.13
Young	0.41	0.28	0.41	0.30	0.42	0.22
Female	0.20	0.25	0.18	0.25	0.28	0.22
Upstream research	0.82	0.38	0.81	0.39	0.86	0.35
External R&D upstream	0.01	0.04	0.01	0.03	0.02	0.05
External R&D downstream	0.07	0.13	0.06	0.13	0.10	0.15
Public funding upstream	0.01	0.04	0.00	0.04	0.01	0.04
Public funding downstream	0.05	0.19	0.04	0.18	0.08	0.22
Technicians	1.11	1.49	1.18	1.58	0.79	0.98
Administration	0.04	0.13	0.03	0.12	0.08	0.17
Geography	1.11	0.44	1.07	0.35	1.28	0.68
Research Centre	0.69	0.46	0.71	0.45	0.60	0.49
R&D intensity	12.78	16.36	10.73	13.82	21.84	22.52
Employees	731.17	6140.93	396.75	1386.41	2213.69	13931.54
National Group	0.50	0.50	0.49	0.50	0.52	0.50
Foreign Group	0.32	0.47	0.32	0.47	0.30	0.46

Table 2 – Regression outputs

	(1) Tobit	(2a) Heckm (S)	(2b) Heckm (C)	(3) NBREG	(4a) BiTobit (Pub)	(4b) BiTobit (Pat)
PhD	0.008***	1.327***	3.433**	2.628***	0.053***	-0.006
	(0.001)	(0.180)	(1.442)	(0.328)	(0.009)	(0.016)
MedDoc	0.006**	0.677**	2.677**	2.414***	0.037**	-0.040
	(0.002)	(0.321)	(1.121)	(0.568)	(0.016)	(0.032)
International	0.003	0.295	1.288	1.036**	0.009	0.010
	(0.002)	(0.244)	(0.821)	(0.472)	(0.013)	(0.019)
Young	0.004***	0.381**	1.079*	1.397***	0.023***	0.018
	(0.001)	(0.157)	(0.641)	(0.396)	(0.009)	(0.012)
Female	0.006***	0.828***	1.684*	0.928**	0.036***	-0.020
	(0.002)	(0.193)	(1.015)	(0.427)	(0.011)	(0.016)
Upstream research	0.001	0.024	0.193	0.114	-0.000	0.009
	(0.001)	(0.105)	(0.297)	(0.224)	(0.006)	(0.008)
External R&D upstream	0.030***	3.040***	6.803*	7.462***	0.101**	0.055
	(0.006)	(0.924)	(3.542)	(2.105)	(0.043)	(0.067)
External R&D downstream	-0.002	-0.372	-0.807	-1.477***	-0.030*	-0.019
	(0.002)	(0.283)	(0.903)	(0.490)	(0.016)	(0.021)
Public funding upstream	0.013**	1.488**	5.116*	6.567**	0.092**	-0.102
	(0.006)	(0.709)	(2.828)	(2.917)	(0.046)	(0.141)
Public funding downstream	0.006***	0.602***	1.623**	1.269*	0.026***	0.008
	(0.001)	(0.165)	(0.717)	(0.651)	(0.008)	(0.015)
Technicians	-0.000	-0.107***	-0.301*	-0.239***	-0.007***	-0.005**
	(0.000)	(0.036)	(0.155)	(0.084)	(0.002)	(0.002)
Administration	-0.001	-0.284	-1.058	-0.563	-0.004	0.048***
	(0.002)	(0.283)	(0.758)	(0.482)	(0.013)	(0.018)
Geography	0.000	0.026	0.112	0.145	0.002	-0.001
	(0.001)	(0.078)	(0.193)	(0.124)	(0.003)	(0.005)
Research Centre	-0.000	-0.065	-0.231	-0.054	-0.006	-0.005
	(0.001)	(0.083)	(0.237)	(0.151)	(0.004)	(0.006)
R&D intensity	0.002***	0.404***	-0.078	0.934***	0.014***	0.017***
	(0.000)	(0.040)	(0.410)	(0.078)	(0.002)	(0.003)
Employees	0.002***	0.366***	0.097	0.917***	0.009***	0.016***
	(0.000)	(0.034)	(0.355)	(0.058)	(0.002)	(0.002)
National Group	-0.000	0.008	-0.345	-0.623**	-0.009*	0.002
	(0.001)	(0.109)	(0.297)	(0.262)	(0.005)	(0.009)
Foreign Group	-0.001	-0.136	-0.608*	-0.762**	-0.014**	0.006
	(0.001)	(0.123)	(0.356)	(0.297)	(0.006)	(0.009)
Sector controls	YES	YES	YES	YES	YES	YES
Constant	-0.021***	-3.482***	-11.169**	-7.697***	-0.130***	-0.160***
	(0.003)	(0.326)	(4.636)	(0.541)	(0.036)	(0.045)
Observations	2,070	2,070	381	2,070	1,767	
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	
Pseudo R2						
Log-Likelihood	883.99			-1718.59	152.26	
Mills-Ratio / Rho		0.063			0.807	

Appendix A.1: Variables

Variable	Description
DV	
PUBint_0809	Publication intensity (Number of publications from the years 2008 & 2009 /R&D expenditures); used in Tobit and Heckman models
Patint_2008	Patent intensity (Declared number of EU patents in 2008 / Employees), used in bivariate Tobit model
Pubint_08	Publication intensity (Number of publications of years 2008/R&D expenditures), used in bivariate Tobit model
Pubcount0809	Total numbers of publications in the years 2008+2009. Used in count model
Demographics $_i$	
Doctorate	Share of researchers with Doctoral degree
- PhD	Share of researchers with doctoral degree in non-health related field
- MedDoc	Share of researchers with doctoral degree in health related field (e.g. medicine)
Young	Share of researchers younger than 35 years
International	Share of researchers with non-french nationality
Female	Share of female researchers and engineers
Scope of research	
Upstream	Firms is spending some (internal) R&D in fundamental research (dummy)
External R&D upstream	Intensity of ext. R&D spending to academic institutions in relation to total R&D expenditures
External R&D downstream	Intensity of ext. R&D spending not focusing on upstream activities in relation to total R&D expenditures
Public Funding upstream	Intensity of public R&D funding obtained from institutions that target upstream research in relation to total R&D expenditures
Public Funding downstream	Intensity of public R&D funding obtained from institutions that is not targeting upstream research in relation to total R&D expenditures
Modes of knowledge creation	
Technician	Proxy for use of modern equipment, computed as ratio (R&D Technicians/Scientists)
Geography	Herfindhal index that measures geographical concentration of research (based on department in France)
Administration	Proxy for research formalization, computed as ratio (R&D administrative employees / Researchers)
Controls	
R&D intensity	R&D budget in relation to number of employees. (in log)
Employees (in log)	Number of employees (measured in log)
Research center	Firm runs a specialied organizational unit for R&D (dummy)
FR Group	Firm belongs to a business group with headquarter in France (dummy)
FOR Group	Firm belongs to a business group with headquarter outside France (dummy)
Sector controls	Industry dummies based on Nace2 codes

Appendix A.2: Correlations

Variable		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
(1)	Pubcount0809	1																		
(2)	PhD	0.03	1																	
(3)	MedDoc	0.03	-0.05	1																
(4)	International	0.01	0.01	-0.01	1															
(5)	Young	-0.03	-0.07	-0.05	-0.13	1														
(6)	Female	0.03	0.16	0.23	0.04	0.13	1													
(7)	Upstream research	0.03	0.08	-0.03	0.02	-0.01	-0.01	1												
(8)	External R&D upstream	0.05	0.05	0.12	-0.02	0.01	0.08	0.02	1											
(9)	External R&D downstream	0.07	0.03	0.17	-0.03	0.00	0.05	-0.01	0.00	1										
(10)	Public funding upstream	0.02	0.13	0.00	0.01	0.02	0.00	0.03	0.00	-0.02	1									
(11)	Public funding downstream	0.03	0.01	-0.02	0.00	0.06	-0.03	0.00	0.06	-0.01	0.06	1								
(12)	Technicians	-0.04	0.05	-0.01	-0.01	-0.12	-0.02	0.00	-0.02	-0.04	-0.05	-0.03	1							
(13)	Administration	0.08	0.07	0.18	-0.02	-0.06	0.09	0.01	0.01	0.09	0.01	-0.03	0.01	1						
(14)	Geography	0.21	0.04	0.03	-0.01	-0.06	0.03	0.02	-0.02	0.08	0.01	-0.02	-0.05	0.20	1					
(15)	Research Centre	-0.08	0.01	-0.03	0.02	0.06	0.01	0.06	-0.01	-0.02	-0.06	-0.05	0.07	-0.11	-0.09	1				
(16)	R&D intensity	0.11	0.03	0.11	0.01	-0.01	0.02	0.05	0.00	0.24	0.03	0.05	-0.13	0.17	0.14	-0.09	1			
(17)	Employees	0.37	0.00	-0.01	0.00	-0.01	0.00	0.02	0.00	0.05	0.01	-0.01	-0.02	0.09	0.30	-0.02	-0.02	1		
(18)	National Group	0.01	-0.02	-0.03	-0.01	0.01	0.00	0.03	0.04	0.02	-0.03	0.02	0.04	-0.01	0.04	-0.01	-0.02	0.05	1	
(19)	Foreign Group	-0.03	0.02	0.06	-0.02	-0.05	0.03	-0.01	-0.03	0.01	-0.06	-0.13	0.03	0.07	0.01	0.04	-0.07	-0.02	-0.68	1

Appendix A.3: Robustness tests

	(5) TOBIT-FRC	(6) NBREG - FRC	(7) TOBIT-IMP	(8) NBREG-IMP	(9) TOBIT-ART	(10) NBREG-ART
PhD	0.004***	2.643***	0.005***	2.386***	0.004***	2.335***
	(0.001)	(0.313)	(0.001)	(0.385)	(0.001)	(0.371)
MedDoc	0.002*	1.893***	0.004	2.270***	0.003	2.168***
	(0.001)	(0.555)	(0.002)	(0.542)	(0.002)	(0.533)
International	0.001	0.915**	0.003**	1.195***	0.002**	1.146***
	(0.001)	(0.411)	(0.001)	(0.458)	(0.001)	(0.441)
Young	0.002**	1.171***	0.003***	1.047***	0.003***	1.035***
	(0.001)	(0.428)	(0.001)	(0.389)	(0.001)	(0.373)
Female	0.003**	0.872**	0.006***	1.147**	0.006***	1.122**
	(0.002)	(0.418)	(0.002)	(0.465)	(0.002)	(0.453)
Upstream research	0.000	0.073	0.001	0.249	0.000	0.240
	(0.000)	(0.236)	(0.001)	(0.239)	(0.001)	(0.233)
External R&D upstream	0.014**	5.659***	0.026	8.321***	0.023	8.067***
	(0.006)	(1.551)	(0.017)	(1.928)	(0.015)	(1.933)
External R&D downstream	-0.001	-1.645***	-0.002	-1.234**	-0.001	-1.135**
	(0.001)	(0.487)	(0.002)	(0.567)	(0.001)	(0.552)
Public funding upstream	0.007*	5.374**	0.009	3.267	0.008	3.471
	(0.004)	(2.523)	(0.006)	(2.204)	(0.005)	(2.174)
Public funding downstream	0.002**	0.784**	0.002*	0.149	0.002*	0.205
	(0.001)	(0.314)	(0.001)	(0.384)	(0.001)	(0.363)
Technicians	-0.000*	-0.292***	-0.001**	-0.326***	-0.001**	-0.317***
	(0.000)	(0.096)	(0.000)	(0.125)	(0.000)	(0.122)
Administration	-0.000	-0.747*	-0.000	-0.425	-0.000	-0.386
	(0.001)	(0.401)	(0.001)	(0.465)	(0.001)	(0.448)
Geography	0.000	0.110	0.000	0.101	0.000	0.104
	(0.000)	(0.116)	(0.000)	(0.142)	(0.000)	(0.135)
Research Centre	-0.000	-0.076	-0.001	-0.043	-0.001	-0.068
	(0.000)	(0.138)	(0.001)	(0.171)	(0.000)	(0.165)
R&D intensity	0.001***	0.905***	0.001***	1.017***	0.001***	0.990***
	(0.000)	(0.072)	(0.000)	(0.081)	(0.000)	(0.078)
Employees	0.001***	0.941***	0.001***	0.906***	0.001***	0.894***
	(0.000)	(0.054)	(0.000)	(0.064)	(0.000)	(0.061)
National Group	-0.000	-0.596**	0.000	-0.461*	0.000	-0.442*
	(0.000)	(0.281)	(0.001)	(0.258)	(0.001)	(0.255)
Foreign Group	-0.001	-0.687**	-0.001	-0.626**	-0.001	-0.603**
	(0.000)	(0.322)	(0.001)	(0.294)	(0.001)	(0.285)
Sector controls	YES	YES	YES	YES	YES	YES
Constant	-0.011***	-8.237***	-0.016***	-7.680***	-0.015***	-7.730***
	(0.003)	(0.505)	(0.003)	(0.552)	(0.003)	(0.534)
Observations	2,070	2,070	2,070	2,070	2,070	2,070
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Log-Likelihood	1156.70	-1270.40	748.08	-1402.41	784.73	-1343.13