Paper to be presented at the DRUID 2011

on

INNOVATION, STRATEGY, and STRUCTURE -
Organizations, Institutions, Systems and Regions
at
Copenhagen Business School, Denmark, June 15-17, 2011

What makes companies contribute to Open Science?

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

Julio D. Raffo
EPFL
CDM - MTEI
julio.raffo@epfl.ch

Abstract

This paper sheds light on companies which contribute actively to Open Science via publishing in scientific journals and conference proceedings. We examine the influences of the internal dimension of the firm, interactions with academic knowledge stakeholder and overall sector conditions that shape the firms? cost-benefit calculation about disclosure in a scientific format. A structural model is proposed and estimated using data from the fourth edition of the French Community Innovation survey (CIS4) with 2.302 R&D performing firms from manufacturing sectors and matched scientific publication data. The analysis provides evidence that a need for academic knowledge sources leads to Open Science contributions by firms, indicating that the access to these knowledge holders imposes the adoption of academic principles in terms of disclosure. Besides, mixed evidence is found that overall industry conditions, including the protection mechanism, influence the cost-benefit rationale of firms in respect to disclosure

Jelcodes:O32,O31

What makes companies contribute to Open Science?

ABSTRACT

Whereas recent scholarly research provides many insights about universities engaging in commercial activities, there is still limited empirical evidence about the reverse phenomenon of companies disseminating scientific knowledge. Our paper fills into this gap and explores the motivations of firms which disclose research outcomes in a scientific format. Besides considering an internal firm dimension, we focus particularly on knowledge sourcing from academic institutions and overall sector conditions using a cost-benefit framework. A structural model is proposed and estimated econometrically with data from the fourth edition of the French Community Innovation survey (CIS4). The sample contains 2.302 R&D performing firms from all manufacturing sectors and matched scientific publication data. Our analysis provides evidence that a need for academic knowledge sources leads to scientific contributions by firms, indicating that the access to these knowledge holders imposes the adoption of disclosure practices. Besides, our results strongly suggest that overall industry conditions are influential in shaping the cost-benefit rationale of firms in respect to scientific disclosure.

1 Introduction

The boundaries between the traditionally distinctive worlds of science and technology become more and more porous. On the one hand, universities and public research institutions are not only contributing to the scientific discourse but engage themselves also in the commercialization of their research results, leading to increasing numbers of university patents (Azoulay et al., 2007; Geuna and Nesta, 2006). The inventions originating from university research are used to generate additional funding and are regarded as an indicator of successful technology transfer. On the other hand, many profit orientated companies are increasingly contributing to the open scientific literature through publishing in peer-reviewed journals (Hicks, 1995; Stephan, 1996). Yet, generic research and scientific publications by firms are counterintuitive since competitors might benefit from the disclosed knowledge (Rosenberg, 1990).

With a few notable exceptions (Stern, 2004; Ding, 2010), there is still limited empirical evidence about the actual motivations and drivers behind contributions of firms to Open Science (thereinafter: OS). Moreover, there is also a strong focus on the pharmaceutical and biotechnology sectors in the wider literature examining science based firms, while other R&D performing industries are much less analyzed (e.g. Gittelman and Kogut, 2003; Cockburn and Henderson, 1998; Liebeskind et al., 1996). In this paper, we intend to provide new empirical insights and model the motivations of firms to publish not only with relation to their internal research activities but also consider explicitly external drivers that shape the firms decision. For this purpose, all manufacturing sectors are considered in the analysis. As descriptive statistics show, the rate of firms publishing is increasing over time for the majority of sectors, supporting the relevance of the question why firms are engaging in this format of knowledge disclosure.

More specifically, the analysis highlights three major dimensions of scientific contributions by firms in a cost-benefit framework. The first dimension captures the firm's internal relation to science. This refers mainly to the degree a firm allocates resources (i.e. time, employees or money) to science but also to scientific firm cultures. The second dimension accounts for interactions with academic institutions and scientists that might induce the firm to publish. The access to academic knowledge can require the acceptance of disclosure clauses in contractual agreements or might even require an own reputation in the scientific community, including reciprocal behavior by the firm. The last dimension accounts for sector conditions favoring or inhibiting scientific disclosure. This refers particularly to the effectiveness of protection instruments that influence the expected spillovers and the scientific openness of firms in a sector.

An econometric analysis of the firm's contribution to OS is conducted using firm-level information from the 4th edition of the French Community Innovation survey (CIS4) and matched scientific publications from Elsevier's Scopus database. The results confirm the majority of the developed hypothesis. First, the more firms are inclined to perform internal research, the more they will contribute to OS. This reflects not only the existence of scientific relevant outcomes deriving from R&D activities but it points also to scientific cultures in firms originating in the employment of scientists and engineers. Second, evidence is found for concessions and reciprocal behavior, which is more related to the firm's appreciation of scientific knowledge inflows rather than the pure existence of formal agreements. Thus, accessing valuable knowledge seems to require disclosure of scientific outcomes whereas collaborations itself do not have a clear influence. Last, results strongly suggest that overall sector conditions influence the scientific openness of firms. In industries where legal protection means are more often used than informal ones, firms are more likely to disclose in a scientific format since the expected spillover effects are lower. Moreover, in sectors where many firms disclosed previously in a scientific format, companies are more often pursuing an OS strategy in the following period. This finding indicates the existence of positive adoption incentives if OS practices are widespread in a sector, e.g. to keep visibility among stakeholders.

The remainder is organized as follows. Section 2 discusses and defines the concepts related to OS in the context of this paper and provides a literature review about scientific publications originating from companies. The framework of analysis and derived hypotheses are presented in Section 3. In Section 4, data sources and sample composition are described, whereas variables and econometric specification are explained in Section 5. In Section 6, results from the multivariate analysis are displayed and discussed before conclusions, limitations and opportunities for future research are presented in Section 7.

2 Scientific publications of companies

Within this paper, determinants behind firms contributing to the scientific literature are examined. The term “Open Science” refers to a set of norms which enhances the diffusion of knowledge. In the academic world, the formal dissemination of research results is crucial to establish priority and to allocate reputation (Merton, 1973; Stephan, 1996). After dissemination, peers can evaluate findings and allocate credit to the authors through citations and the deriving recognition and reputation is then the basement for further career advancement (Dasgupta and David, 1994). Due to this emphasis on disclosure, scientific publications are not only a proxy for OS but can be regarded as a part of its definition. It is important to stress that OS is viewed here from a knowledge diffusion perspective, which does not consider any issues in respect to the free use of the knowledge[1]. Therefore, the terminus Open Science does not exclude here the possibility that companies are seeking for formal IP-instruments for the same knowledge in parallel to the scientific publication[2].

Given the mission of companies in a market economy to generate returns from their R&D investments, scientific publications and the corresponding voluntary knowledge disclosure appear counterintuitive. Besides the costly conduction of generic research with its uncertain outcomes and imperfect appropriability (Nelson, 1959; Rosenberg, 1990), intentional disclosure makes knowledge directly available to competitors. Moreover, the disclosure process itself contains opportunity costs since the researchers have to prepare their publications to meet the respective requirements of the target journals (Penin, 2007; Kinney, 2004). Regarding the organization of R&D in a firm, a compliance of firm scientists to academic norms may lead to agency conflicts between managers and scientists if these receive incentives to publish. The scientists might commit too much effort to generic research and corresponding disclosure activities but do not engage sufficiently anymore in the translation of the results into applied outcomes that can be commercialized (Cockburn et al., 1999; Gittelman and Kogut, 2003).

On the other hand, the potential benefits of OS contributions by firms are various and refer both to knowledge creation and diffusion. The seminal paper of Hicks (1995) highlights several potential benefits and serves therefore as an important starting point for analyzing motivations. A first reason lies in the possible desire of firms to be very closely connected to academic scientists and to be part of the wider scientific community. By maintaining personal contacts to academic scientists, firms might be able to get the latest knowledge which is not

[1] This interpretation is in line with the one applied by Ding (2010) in a very similar context.

[2] For a more general discussion of disclosure regimes with an explicit consideration of access issues, see Gans et al. (2010). However, even in the case of a parallel patent protection, competitors might learn additionally about the general research direction of a firm.

even published yet. In order to build up credibility among academic scientists, publishing firms can gather an entrance ticket to the respective scientific communities (Hicks, 1995; Rosenberg, 1990; Liebeskind et al., 1996). In respect to socialization processes and firm culture which favor OS contributions, Ding (2010) shows that strong scientific backgrounds of biotech start-up founders have a positive effect to the adoption of an OS strategy. Moreover, companies can use scientific publications as an instrument to hire researcher and to motivate them (Stern, 2004; Hicks, 1995). Even though there are some differences in the characteristics between academic scientists and their counterparts in firms, scientists in industry share similar values and are concerned about their reputation (Sauermann and Stephan, 2010). Particularly in the first time after employment changes from academia to industry, scientists often strongly value autonomy in determining research agendas or the right to publish research results in scientific journals. In this respect, Stern (2004) shows that scientists being on the job market are willing to accept lower wages in order to receive the right to publish and to have some freedom in selecting research agendas. This indicates that firms offering their own researchers the concession to publish should have benefits in hiring, motivating and retaining them.

Muller and Penin (2006) argue with a theoretical model that the disclosure of knowledge can be useful for establishing credibility among other firms and for signaling competences to have a better access to potential collaboration partner. Moreover, signaling strategies can also refer to the audience groups of public funding authorities or professional customers. With disclosure in scientific journals, companies might have access to public funding since firms signal their capability to use the R&D funding effectively (Penin, 2007). On the other hand, the reverse logic may apply correspondingly since some R&D subsidies are bounded to certain disclosure requirements, yet not necessarily in a scientific format (see EC, 2003).

Similarly, scientific publications can signal the companies' innovativeness to professional costumer groups (Hicks, 1995; Penin, 2007). For instance in the pharmaceutical sector where regulatory authorities demand for information and where physicians have to be convinced of the value-add of a drug, scientific publications could be a useful instrument. In addition, scientific articles and conference proceedings can be used to set standards when companies are competing for different alternatives in network industries (Penin, 2007). In relation to intellectual property protection, publications in journals can be used to keep the freedom to operate in very competitive environments where firms are engaging in patent races (Parchomovsky, 2000).

As this review suggests, the decision of firms to publish in scientific journals may depend on a variety of reasons. On the other hand, publishing in scientific journals means in a first instance the revelation of knowledge which let competitors potentially benefit. In the following, a new perspective is proposed which emphasizes knowledge sourcing from the public science base and overall sector openness as important drivers of the firms' decision to participate in OS.

3 Cost-benefit considerations behind Open Science contributions

In this section we present and discuss framework and hypotheses. Scientific publications by firms can be seen as the outcome of a process evaluating the benefits and costs of this particular type of disclosure. The central cost component lies in spill-overs which allow competing firms to exploit the disclosed knowledge at least partly for free whereas the potential benefit considerations are more various. Therefore, the discussion is structured around factors that influence the cost-benefit threshold of publishing.

A precondition for disclosing knowledge is the underlying creation of it and the conduction of R&D can therefore be seen as a reference point. However it is to consider that many firms are integrating external knowledge sources for their R&D process. We believe that this knowledge sourcing component might be a driver for scientific firm publications due to concessions and reciprocal behavior imposed by the academic partner. Moreover, the openness of an entire industry in respect to disclosure and the used protection instruments for inventions - which are potentially moderating the threshold through the relative spill-over costs - are considered. In this respect, the protection instruments should directly influencing the cost of disclosure. These elements represent the basic framework of the analysis and are discussed in-depth within this section.

3.1 Internal research and firm culture

The production of relevant knowledge can be seen as a precondition for disclosing results in scientific journals and subsequently for realizing potential benefits from this strategy. The more resources a firm commits to internal R&D and particular to the research component, the more outcomes are likely to be eligible for scientific publishing. In this respect, not only pure basic research might be adequate for publishing but also outcomes of a more applied nature, which potentially contain fundamental insights (Stokes, 1997; Murray, 2002).

In addition, internal R&D represents implicitly considerations of managers with respect to their firm scientists. The production of internal knowledge is associated with the employment of a certain number of highly qualified scientists and engineers. Scientists in firms often share norms and values which they have obtained due to long stays in academic institutions. As mentioned in section 2, many industry scientists are concerned about academic values as reputation and autonomy and show similar motivations than their academic counterparts

(Sauermann and Stephan, 2010; Stern, 2004). Even though the adoption of scientific principles may have different expressions, e.g. the ways of problem-solving for knowledge creation, the most obvious can be seen in the outcome of scientific publications since these are the central instrument for allocating rewards in OS (Dasgupta and David, 1994). In this respect, scientific practices are adopted quite often in firms where scientists are key employees, as recent evidence suggests. Using scientist-level data from the US "Sestat" survey, Sauermann and Stephan (2010) found that around 71% of all corporate scientists in the life sciences and 54% in the physical sciences have published at least one article in peer-reviewed journals.

From a management perspective, the right for firm scientists to publish reflects the benefit consideration of having motivated scientists (who might otherwise leave the firm) which outweighs the potential spillover costs. It is to stress that the firm's benefits regarding disclosure are likely to go beyond the aspect of sustaining motivation and hiring possibilities, for instance marketing effects towards professional customer groups or defensive publishing in order to keep the freedom to operate. However, also these motivations are to some extent reflected in the commitment of the firm to internal knowledge creation since the potential use for defensive publishing is dependent on inventions deriving from the R&D activities. Nevertheless, these potential usages of scientific publications are not in the scope of this analysis and are therefore not further discussed.

Beyond the internal perspective of R&D, it is widely acknowledged in scholarly discussion that internal R&D efforts are necessary for the ability to absorb external knowledge, making it an incentive to conduct R&D (Cohen and Levinthal, 1989). Within the following subsection, drivers of scientific disclosure originating from interactions with academic institutions and scientists are discussed. Summarizing this subsection, the following hypothesis is stated:

H1: Firms committing more intensively to internal knowledge creation are more prone to contribute to OS.

3.2 Knowledge sourcing from Academia

Many innovating firms are not only relying on internal knowledge creation but integrate also external sources for their innovation process. Suppliers, costumers but also the public science base can serve as important information sources for project initiation and conduction (Chesbrough, 2003; Laursen and Salter, 2006).

In respect to the channels between the public science base and firms, different kinds are evaluated as important, ranging from passive sources as scientific literature to personal interactions with academics (Cohen et al., 2002; Arundel and Geuna, 2004). It is obvious that accessing these channels impose different efforts and costs to the firm. The exploitation of passive knowledge sources like scientific journals requires only a public library access or the subscription of some particular relevant journals. It can be expected that the access is not similarly straightforward if companies intend to interact with academic scientists aiming to reach their tacit knowledge. Generally speaking, academic laboratories may have a low interest to enter collaborations with firms if the opportunity costs of pursuing a common research project are too high (Carayol, 2003; Fontana et al., 2006). As a consequence, firms have to offer topics that are interesting from a scientific point of view, particularly if well reputed scientists or labs are targeted. These should be able to select their collaboration partner more carefully due to a better access to alternative funding channels.

Furthermore, the governance of the relationship is an important question since the institutional missions and incentive systems in academia and industry are distinctive

(Dasgupta and David, 1994; Bruneel et al., 2010). One particular point of potential controversy concerns the disclosure of project outcomes. Whereas the incentive systems of the academic sciences always favor disclosure, firms tend to keep project outcomes secret in order to avoid spillover effects to competitors. In this respect, it is often argued that inter-organizational collaborations may impose restrictions to the academic system where the norm of disclosure is threatened, e.g. through secrecy and delay of publication (Blumenthal et al., 1996; Czarnitzki et al., 2011). However, as universities are not necessarily the weaker negotiating party, it can also be argued that collaborations might impose a greater openness to firms (Tijssen, 2004). In other words, firms might concede the disclosure of project outcomes in order to convince the academic partner to collaborate.

In the context of informal interactions, slightly different mechanisms are expected to apply. Academic scientists share the norm of communalism that facilitates knowledge exchanges in the academic domain but it should exclude researcher which are not contributing to the community (see Merton, 1973; Hicks, 1995). This implies that scientists in universities may not have much incentive to interact with (firm) researchers if there is only one direction of knowledge flows. In contrast to formal agreements, financial resource provisions are less likely the basement of informal interactions, which should increase the importance of contributing to the community in order to access academic scientists. Moreover, concerns of academic scientists about the exclusive appropriation of shared knowledge by firm researcher can impose barriers for interacting. Supporting this line of argumentation, Häussler (2010) suggests that academic researcher differentiate in their information sharing behavior between inquirers from academia and industry, where industry researcher face a higher likelihood of getting their request rejected. Therefore, firms actively participating in knowledge sharing are likely to have an advantage in maintaining informal contacts. In respect to scientific publications as means of openness it is self-evident that these are not restricted to any interested party. However, firms and their researchers show credibility and can build up trust via providing knowledge to the community (Hicks, 1995). This appears particularly plausible if one considers that the academic community is not only associating a firm publication with the company but with the authoring firm scientist in person.

Bringing these arguments back to the firm level, it is proposed that the importance and formality of the interactions sought with academia influences the considerations of R&D managers about costs and benefits of disclosure.

H2a: Firms evaluating universities as important knowledge sources are more prone to contribute to OS.

H2b: Firms relying on informal interactions with academic institutions rather than contract-based collaborations are more prone to contribute to OS.

3.3 Sector openness

Apart from direct firm-level considerations, sector level characteristics are also likely to be influential for the firms' cost-benefit evaluation of scientific disclosure. In this respect, we distinguish between knowledge flows and appropriation conditions for characterizing industry openness. Concerning the former, firms might have different incentives imposed by sector-wide mechanism. The most straightforward example are regulatory requirements in some sectors which force firms to show scientific evidence in drug development (clinical trials) or when stating health claims about food products (Hicks, 1995; EP, 2006). Thus, obtaining the authorities approval and market access constitute the indirect benefits of the scientific disclosure.

However, other drivers originating from sector conditions may also apply. In the pharmaceutical and biotechnology sector, firms frequently publish in high quality journals. One reason is to be closely connected to academic research (Hicks, 1995; Cockburn and Henderson, 1998; Gittelman and Kogut, 2003). In such an environment, there might be important implications for non-publishing firms such as a lower visibility and reputation. Those firms could face disadvantages in entering collaborative research agreements with universities or other firms. Similarly, firms which are not allowing their researchers to publish are likely to face disadvantages in attracting the best university graduates (see Stern, 2004). As a consequence, firms should have incentives to adopt OS practices and to publish in scientific journals if other firms of the same sector show scientific openness[3].

Moreover, the information flows in a sector do not necessarily represent unintended spillovers (Allen, 1983; von Hippel and von Krogh, 2003). The most recent example comes from the Open Source Software sector where profit oriented companies are actively disclosing latest program codes in order to leverage follow-on work. It is imaginable that higher knowledge flows are leading to higher scientific openness since the company benefits from disclosing of competitors, leading to lower perceived costs of own disclosure.

Applying a stricter view on openness, the disclosure of knowledge by a firm does not necessarily imply the access for other companies to it (Murray and O'Mahony, 2007). This is referring to the central cost component, the spillovers deriving from disclosure. The effective protection means of knowledge in a sector and their relationship to disclosure are likely to be influential for the decision to publish or not. In this respect, the sector perspective is justified since the use of different instruments is often considerably varying between industries but is more homogeneous within a sector (Levin et al., 1987; Cohen et al., 2000).

---

[3] See also Ding (2010) for the influence of widespread OS strategies in the biotechnology sector and managerial decision making in bio-tech start-up firms.

Particularly legal instruments and informal means should reveal a different relationship towards disclosure (Gans et al., 2010). For instance, the procedure to obtain patent protection includes a compulsory disclosure after 18 months from the application date. As a consequence, the marginal costs of the additional disclosure in a scientific publication should be lower when assuming a sufficient content overlap of the patent and the publication (cf. Murray, 2002; Stokes, 1997). Differently, the reliance on informal appropriation instruments like secrecy can be regarded in a straight opposite direction towards disclosure. As Harabi (1995) shows, many firms regard already the disclosure in a patent application as problematic and rely more on informal methods in order to avoid any disclosure. In such a case, publishing in a scientific journal is likely to be harmful. It is clear that knowledge contents might be partly separable on the project level and firm strategies concerning knowledge protection may include a diverse set of strategies, meaning a parallel reliance on secrecy and patents (Arundel, 2001). Nevertheless, a high overall importance of informal means in contrast to formal protection should affect the attitude of disclosing in scientific journals since the perceived and actual costs are likely to be different.

H3a: Firms operating in sectors with greater scientific openness in the past are more prone to contribute to OS.

H3b: Firms operating in sectors with a higher use of formal (informal) protection means are more (less) prone to contribute to OS.

4 Data and descriptive statistics

4.1 Data sources

The empirical analysis relies on two main data sources referring to firms and their scientific publications, respectively. The firm level information is extracted from the fourth edition of the French Community Innovation Survey (henceforward CIS4) which gathers information about the innovative behavior of firms during the period 2002-2004[4]. It is important to stress that the sampling procedure of the French CIS4 favors the inclusion of larger firms which might restrict the representativeness in respect to smaller firms. The scientific publications of each firm were retrieved manually from the Elsevier's Scopus online database (Data extraction date: 15.05.2010). The Scopus database has a very broad journal coverage which goes beyond the also commonly used ISI Web of Science database (Falagas et al., 2008) and represents therefore the best available data source[5].

For the sake of the analysis, both data sources have been limited in several aspects. Regarding the CIS4 data, only R&D performing firms from manufacturing sectors have been included as we are interested in scientific publications as outcomes of scientific research and technology development. Moreover, non-innovative firms have been disregarded since a considerable amount of CIS4 questions is only answered by innovative firms (see also Mairesse and Mohnen, 2010). In order to avoid potential bias arising from academic spin-offs and science based start-ups – e.g. the biotechnology industry – a conservative approach is pursued by

---

[4] For further details, see Mairesse and Mohnen (2010).

[5] All firms in our sample are based in France and some firms could target French edited journals only. Since Scopus has an international scope there might be an underrepresentation of French journals. In order to assess this potential bias, we cross-checked its coverage with the PASCAL database from the French institute INIST. Results have shown that Scopus fairly represents French speaking journals since around 45% of the journals covered in PASCAL are included in Scopus. Given its quality standards and greater international coverage, Scopus still represents the best available source, although we cannot completely rule out that our approach could miss some publications.

removing firms with a high R&D turnover ratio (i.e. more than 50%) and also those with less than 20 employees (21 and 237 observations removed respectively).

Regarding Scopus' scientific publication data, only articles and conference proceedings have been considered since only these represent original research outcomes. Besides, publications have been classified by their publication date to three different periods of time: before (1999-2001), during (2002-2004) and after (2005-2007) the CIS4 coverage period.

The retrieval and cleaning procedures related to the scientific publications deserve further explanation. While generally speaking the affiliation field in the Scopus database is from a good quality (e.g. not sensitive to language specific characters), a broad text search was conducted aiming to minimize false negative matches. The search strategy consisted in using both truncation of the firms' names and their known acronyms together with wildcards. The main potential limitation of such procedure, as any search based only in firm names, concerns the risk of a firm changing its name or publishing under the headquarters name instead of its own. Yet, the former is only from a limited risk since the search was performed on a nine years period, which allowed for detecting any abrupt discontinuities. In respect to the latter, the decision of aggregation is pre-determined since CIS4 is a firm survey but not a business group one. Since the different units of a group may pursue rather different innovation strategies, a consolidation to the business group level is not a pertinent approach.

In order to minimize false positives, all retrieved publications were carefully checked both automatically and manually. As a general rule, we decided to apply a conservative approach where the risk of misattributing a scientific publication was minimized. First, automatic examination queries were executed to rule out any academic affiliation of the potential matches that refer in fact not to the firm but to institutions with similar name. This includes also mixed affiliations between public research institutions and companies, which were also excluded[6]. The remaining matches were examined manually. Each potential match was cross-checked using the legal address of the company – including the addresses of all known establishments – and the author's address in the scientific publication. In very few cases, firms with only common French or English terms in their names produced an extremely large amount of false positive matches. The impracticability to check all the publications manually led us to fully exclude those.

4.2 Descriptive statistics

The final sample contains 2'302 R&D performing firms. Within this sample, 339 firms (almost 15 percent) have published at least one scientific article or conference proceeding during the nine years period. The resulting data suggests that the population of publishing firms is increasing over time, although most of the increase occurred between the first and the second period where the number almost doubled[7].

-- Insert Table 1 about here –

As depicted in Table 1, there is considerable variation across industries. Not surprisingly, the sector class covering pharmaceutical, biotechnology and chemical industries (NACE 24) is

---

[6] This refers to public research units which are attached to a private company facility. This is not to confuse with co-authored scientific publications between firms and academia or authors with double affiliations. The most frequent case in our dataset consists of those known as *unité mixte*, where a research unit is shared between the Centre Nationale de la Recherché Scientifique (CNRS) and a private firm.

[7] The fast growing and following slowing down stagnation might be overstated, as the sample does not account for the firms departing in the pre period and those entering and departing in the post period.

the strongest sector in the subsample of publishing firms, although it accounts only for 28 percent of it. Indeed, with the exceptions of Apparel (NACE 18), Leather (19) and Wood products (20), all other sectors contain at least one publishing firm in the 9-years period. Moreover, in half of the sectors there are more than 10 percent of firms having at least one scientific publication. In particular, firms from the industries Communication equipment (NACE 32), Medical and optical instruments (33) and Transport equipment (35) are almost as likely to publish as those from the Pharmaceutical and Chemical sector (24). Therefore, scientific publications are by no means only a phenomenon of the Life Sciences and Chemistry industries.

Summary statistics in Table 2 provide a general description of the sample and some particular traits of the subsample of publishing firms[8]. Even in the setting of R&D active and innovative firms only, the practice of publishing is highly skewed. Around one third of publishing firms has only one publication in the estimated three years post-survey period, while two third have less than five publications.

-- Insert Table 2 about here --

Without taking into account any multivariate interaction, firms publishing scientific outcomes are expected to be larger and belonging to more technological intensive sectors. These firms are also more often part of a business group, which is more likely to be foreign. Publishing

[8] In the remaining discussion, we focus on the publishing behavior of firm in the three years period immediately after the CIS4 surveyed period. The main reason concerns our intention to explain this ex-post publishing behavior by the ex-ante innovation strategies taken.

firms are more R&D intense in average, being twice as intense in terms of turnover share and almost three times in terms of R&D expenditures per employee.

In general, these firms are closer connected to innovative institutional environments. This refers both to information sourcing and collaboration agreements with any kind of partner. Nevertheless, publishing firms are particularly more relying on academic sources and collaborate more frequently with academic partners than non-publishing firms. Similarly, a higher proportion of publishing firms benefits from public support to R&D and innovative activities – including tax benefits – from French or European agencies. This spread between publishing and non-publishing firms emerges across all institutional levels, but European sources show the larger imparity. Interestingly, publishing firms are not only more likely to use formal appropriation means – such as patents, trademarks, designs or copyrights – but also informal ones like secrecy, conception complexity or technological lead.

5 Econometric specification

5.1 Theoretical model

We propose to frame the firm's contribution to OS as a model containing three main elements. First, the firm has to have eligible outcomes for disclosing in a scientific format. Second, the firm evaluates the benefits and costs of disclosing scientific information or not. Third, if agreeing to disclose, the firm decides about the quantity of disclosed information.

This can be formalized for a firm i in a simple equation as follows:

$$P_i = \tau_i . S_i \qquad (1)$$

Where S_i represents the eligible outcomes arising from the R&D activities, τ_i is the firm's propensity to disclose and P_i the quantity of disclosed scientific information. By explicitly decomposing P_i into S_i and τ_i, we pretend to have a clear-cut account of what drives the firm for the disclosure decision, which is captured by τ_i. It is straightforward from the above equation that a firm will not contribute to open science ($P=0$) if it does not produce at least some scientific outcome ($S=0$) or if it decides for whatever reason not to disclose any scientific outcome ($\tau=0$).

There are two basic difficulties to overcome in order to correctly estimate these three elements. First, the determinants of S_i and τ_i are shared to a great extent. As it has been discussed in the section 3.1, firm scientists may affect not only the rate of produced scientific outcomes but also the attitude of the firms' management towards disclosure since the permission to publish can be used as a Human Resource policy. Second, scientific outcomes are not fully observable as we capture them only when disclosed, i.e. not every outcome eligible for disclosure in a scientific journal is actually revealed. However, the level of the created scientific outcomes can be observed indirectly through the intensity of the internal R&D expenditures[9]. Nonetheless, as just mentioned, these scientific activities are also determinants of τ_i.

Given these considerations and assuming $S = \theta_i \; R_i$ – where R_i represents the scientific activities and θ_i a transformation factor – we can rewrite equation (1) as:

$$P_i = \tau_i(R_i, \theta_i, X_i).\, \theta_i R_i \qquad (2)$$

[9] It can be argued that scientific outcomes can also be reached through the means of external research only. But, in such case, we expect their published version not to be attributed to the acquiring firm.

We have only partially solved both difficulties since θ_i is not directly observable and there are many factors affecting τ_i – summarized in X_i – which may also affect the scientific activities R_i. We propose to solve this latter obstacle by jointly estimating the three elements of interest in a recursive system of three equations. It is to stress that the non-observation of θ_i implies that its effect will be absorbed by other elements in the estimation described below.

5.2 Equations and variables

The first equation of the system accounts for the endogenous nature of R_i by modeling it as:

$$R_i = RDint_i = \alpha_1 Sources_i + \alpha_2 Approp_i + \alpha_3 Z_i^1 + \eta_i \quad (3)$$

Basically, it is possible to consider either the R&D intensity or the level of R&D as dependent variable of this first equation. This choice is influential in respect to the degree of representation of the scientific outcomes and the scientific culture dimension. Whereas the level of R&D captures more the eligibility of outcomes for publication, the R&D intensity should additionally reflect the scientific culture dimension. Another shortcoming of the R&D amount is its strong correlation with firm size. Given these reasons, we have chosen the R&D intensity. Since the aim is to isolate the firm internal dimension of scientific disclosure (see accordingly section 3.1), the R&D intensity refers only to internal R&D expenditures.

Like "typical" R&D estimations as established in literature, we expect the different origins of information inflows in vector $Sources_i$ - i.e. customers, suppliers, competitors, conference, universities or specialized journals - to affect the firm's trade-off between absorbing and creating knowledge (Cohen and Levinthal, 1989). This includes also cooperation with universities. Similarly, the formal and informal appropriation strategies in $Approp_i$ – i.e. patents, secrecy, conception complexity or technological lead – are also expected to affect the internal R&D expenditures. The vector Z_i^1 contains additional variables – such as foreign group, independent firm, size, the technological intensity of the sector and public R&D subsidies – aiming to control for the remaining heterogeneity[10]. The error term η_i is allowed to be correlated with the other two equations in case there are still unobservable factors affecting the R&D intensity and both the decision to disclose scientific information and its amount. We will use this particularity to test the exogeneity of R_i with respect to X_i.

The second equation in the system captures the decision to disclose scientific information, i.e. the likelihood of contributing to OS and is obtained from:

$$\tau_i = E[PUByes_i = 1|\tau_i^*]$$

$$PUByes_i = \begin{cases} 1 \; if \; \tau_i^* > 0 \\ 0 \; if \; \tau_i^* \leq 0 \end{cases}$$

$$\tau_i^* = \beta_1 R_i + \beta_2 Collab_i + \beta_3 SectCond_i + \beta_4 Z_i^2 + \epsilon_i \qquad (4)$$

The firm's propensity to disclose scientific outcomes (τ_i) derives from the firm's evaluation of benefits and costs of disclosing the knowledge. In this respect we only observe if the firm indeed publishes the scientific outcomes but not if otherwise. Thus τ_i is obtained from the expected probabilities of the firm publishing at least one scientific article ($PUByes_i$), which is itself conditioned to an unobservable latent variable (τ_i^*). As mentioned above, it is assumed that the internal scientific activities (R_i) shape this decision. Estimating the scientific activities

[10] A full description of variables used in the system of equations is provided in the Annex Table A.1.

R_i in the first stage of the system allows not only to estimate correctly its impact on τ_i, but also to estimate without bias the effects of the other factors contained in X_i. This is particularly important for separating the effects of academic knowledge sources towards internal R&D and disclosure. In order to account for the time lags between the firm-level decisions and the disclosure of research outcomes, we estimate the scientific publications in the post-survey period (i.e. publications from the years 2005-2007) with the firm level information of the years 2002-2004.

As stated in section 3.2, a core argument of our study is that the disclosure of scientific outcomes is influenced by the type of interactions the firm has established with the scientific community ($Collab_i$). We consider four different interaction types which distinguish between the attributed importance and the formality of these interactions. The first one reflects firms which are (formally) cooperating with academic partners and simultaneously considering them as important information sources for innovation. The second type covers companies evaluating universities as important sources for innovation but do not have any cooperation agreements, indicating informal links to academic scientists. As a further distinction, we separated explicitly firms claiming to collaborate with academia without evaluating them as important sources. This category refers to cases where universities might just provide technical services or where the collaborative project failed. Recent evidence suggests that cooperation failures are not a rare phenomenon in firm-university collaborations (Lhuillery and Pfister, 2009). The remaining category comprises firms that have no link with scientific institutions, i.e. firms that do not collaborate and do not source information from academia. The underlying variables enabling these 4 distinctive types are the importance of universities as a source for innovation (measured with a 4 point Likert scale, aggregated to dummy high/medium use) and the existence of cooperation with universities (dummy variable).

Moreover, we expect the overall sector openness to be influential in explaining Open Science contributions by firms (see Section 3.3) which is represented in the vector $SectCond_i$. This set of variables covers the average use of protection means where we distinguish each sector rate of formally and informally protected firms. These should reveal a different relationship in respect to the relative spillover level of disclosure. Moreover, the sector commitment to OS (i.e. scientific publications in the pre-survey period) and the overall spillover levels among competitors are further measures of sector openness. These sector variables are computed on the NACE2 level[11].

The vector Z_i^2 contains a set of control variables accounting for firm level heterogeneity. The specification is almost identical to the vector Z_i^1 of the first equation. The only exception is that only national and EU level public funding institutions are considered to be relevant for the disclosure decision as opposed to all institutional levels. This control variable deserves particular attention since the disclosure decision might be imposed by certain requirements of the granting institution. For instance, in the 6th European Framework Program, there is a disclosure clause included, although without determination of the particular medium of disclosure (see EC, 2003). Last, ε_i is the error term with some specific properties that are discussed below.

The last equation in the system measures the OS commitment of firms once the disclosure decision has been accounted for. Given the nature of our publication data, P_i follows a discrete distribution with a higher density in lower values. Therefore, this variable has been

[11] It is to stress that these sector variables reveal relatively high correlations (See Annex A.2). Nonetheless, the corresponding variables can be distinguished from a theoretical perspective and we performed several robustness tests to guarantee the validity of results. These are available upon request.

grouped into three ordinal categories (1, 2-5 and more than 5 publications) in order to consistently estimate the impact of the regressors of interest[12].

In formal terms:

$$PUBQ_i \begin{cases} 1 \quad if\ PUBQ_i^* \leq c_1 \\ 2\ to\ 5 \quad if\ c_1 < PUBQ_i^* \leq c_2 \\ 5+ \quad if\ c_2 < PUBQ_i^* \end{cases} \tag{5}$$

$$PUBQ_i^* = \gamma_1 R_i + \gamma_2 Collab_i + \gamma_3 SectCond_i + \gamma_4 Z_i^2 + \upsilon_i$$

Where $PUBQ_i$ is the categorial variable referring to the amount of scientific publications of each firm i (P_i), $PUBQ_i^*$ is the latent variable and c_1 and c_2 are the thresholds separating the three categories of publishing firms. At this stage, we do not have any solid theoretical case for a distinction between factors affecting the propensity to disclose scientific outcomes (τ_i) and those affecting the quantity of disclosed scientific information (P_i). For this reason, the variables of interest and the controls Z_i^2 are exactly the same ones as in the decision equation. From an econometric point of view, it is adequate to have an identical variable specification among the two equations unless there is strong justification for a different choice[13].

The last specificity of our econometric approach concerns the errors terms η_i, υ_i and ε_i. First, the choice of modeling techniques implies a certain variance structure for each equation. Second, these error terms are allowed to be correlated among themselves. Consequently, the

---

[12] An alternative approach would be to use the original variable in a count data model, such as Poisson or Negative Binomial regressions. However this option would not allow us to control correctly for the endogeneity of R&D.

[13] See for instance Crepon et al. (1998).

following error structure is imposed to ensure consistent estimates of the parameters of interest:

$$\Sigma = \begin{bmatrix} \sigma & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix} \quad (6)$$

Where σ is the error term standard deviation for the R_i equation, the decision to publish and the amount are imposed with a unity standard deviation and ρ_{12}=corr($\eta_{i,}$ ε_i,), ρ_{13}=corr($\eta_{i,}$ $\upsilon_{i,}$) and ρ_{23}=corr($\upsilon_{i,}$ ε_i,).

The above discussion implies in practical terms the joint estimation of the vectors α_k, β_k and γ_k by maximum-likelihood, assuming the joint covariance matrix Σ[14]. In detail, R_i is estimated using an GLS estimator, τ_i with a Probit estimator on $PUByes_i$ and P_i through Ordered Probit applied to the categorized variable $PUBQ_i$. Due to the use of sector level variables, all three equations are estimated with robust standard errors, clustered by two-digit NACE industries.

6 Results

The Table 3 displays joint estimation results of the system described in the previous section for two slightly different specifications of independent variables choice[15]. As expected, the internal scientific activities (H1) – captured by the *R&D intensity* – are associated with a

[14] For its computation we use Stata's CMP procedure (see Roodman, 2009).

[15] Results of the auxiliary first equation can be found in the Annex Table A.3. Moreover, the Table A.4 shows a regression in which the influence of Public R&D funding is displayed more fine-grained using interaction effects with academic sources and cooperation. However, the results remain very stable when contrasted to our main regressions presented in Table 3 and are therefore not further discussed.

higher likelihood and larger amounts of OS contributions. We believe this is reflecting both the existence of more outcomes eligible for scientific publication and the internal commitment to scientific practices, which includes the pressure of firm's scientists to get the right to publish. The latter is expected to influence in a greater extent the decision to allow any scientific publication, while the former is expected to increase the amount of publications once the scientists were allowed to publish.

As imagined, the link to academia (H2) implies a higher affinity to participate in OS. A firm relying on academic knowledge sources (*Source Academia)* for its innovation process will approximately increase six percent its likelihood to contribute to OS. However, a firm collaborating with academic partners (*Coop Academia*) in their innovation projects has no significant impact on likelihood and amount. Willing to understand these differences in greater depth and to control for some overlapping forces between these two aspects, we set a second specification where the four possibilities are taken into account[16]. Results depicted in the columns (2) confirm the dominance of the source importance. Firms considering academia as an important source for innovation – regardless the formality of their interaction – are both likely to contribute to OS and do it with larger amounts than other firms. The difference between the interaction kinds formulated in Hypothesis 2b (i.e. *Academia: Source & Coop*, *Academia: Only Source*; *Only Coop*) is found statistically significant in both the likelihood and the amount equation (p-values of equality of coefficients are 0.000 and 0.0178, respectively). When looking only at the first two, the difference is barely significant in the likelihood equation (p-value 0.0914) and the degree of formality is therefore not influential. Only technical services which are not from a high importance are associated with a lower likelihood of OS contributions in comparison to the other interaction types, leading to a non-confirmation of H2b.

---

[16] Firms not sourcing and not collaborating with academia are left as baseline.

In any case, results are supporting the existence of concessions and reciprocal behavior to academic institutions. In the case of formal collaborations, this result can be interpreted as the outcome of contractual agreements about disclosure which are already fixed before the actual project launch. In the case of informal exchanges, the scientific publications after knowledge sourcing from academia may also be interpreted as reciprocal behavior. Even though we use an indirect measurement strategy – as we predict publications in the post-survey period – the results suggest that companies have to show a reverse openness for fully leveraging academic knowledge sources.

The results concerning industry openness show that the conditions mostly affect the decision to contribute to OS, but less the amount of contributions. This is particularly the case for the sector's use of the formal and informal protection means (H3a) which moderate the relative spillovers and misappropriation risks. Not surprisingly, firms in sectors with an intensive use of formal IP instruments (*Legal Protection – sector*) such as patents or trademarks are more likely to consider OS contributions as beneficial, whereas the opposite is found for those in sectors with a high use of informal IP mechanisms like secrecy *(Informal Protection – sector)*. Interestingly, their opposite impact is not statistical different. This means that firms in sectors with lower intellectual property use will not necessarily differ in their perception of costs and benefits of OS contributions, as far as the balance between firms using formal and informal instruments is the same. The sector's protection intensity – formal or informal – seems to have no further effect on the disclosure amount, once the impact on the decision has been taken into account. This is plausible since once the firm feels protected, other reasons for OS contributions should be more influential for the amount determination.

-- Insert Table 3 about here --

Moreover, firms in sectors where a higher percentage of companies has published in the past *(Publications – sector)*, are more likely to contribute to OS and do it also more extensively (H3b). We interpret these results as follows. First, firms in environments with affinity to OS have not only incentives to disclose scientific information but also to build an academic reputation. This reputation might be helpful in accessing potential collaboration partner or in hiring personnel and a scientific openness of firms can help the firm to be visible among stakeholders. Concerning the intra-sector openness among competing firms, we find mixed results. There is no evidence that higher levels of spillovers among firms in a sector *(Competitor spillovers – sector)* are encouraging or preventing the single firm's attitude with respect to OS contributions. Thus, the possibility that information – scientific or not – flowing among firms creates a larger scientific openness seems to be too optimistic. On the other hand, there are no grounds to the skepticism on OS practices being considered harmful in industries with higher technological competition.

Lastly it is to stress that the results strongly support our model and its specification. First, the internal scientific activities and resources – captured by the R&D intensity – are confirmed endogenous as the correlation between errors η_i and ε_i is found significantly different from zero in both specifications (See Annex Table A. 4). The lack of significance for almost all control variables is an additional sign for the endogeneity of the R&D intensity, though indirect[17]. Second, the statistically significant correlation between errors υ_i and ε_i indicates the amount of disclosed scientific outcomes is not independent from the decision to allow scientific disclosure. In other terms, separating this process into two different stages has been confirmed as an appropriate estimation strategy.

---

[17] In the univariate analysis of section 4, most of these variables are seen statistically different for the average firm publishing from those not publishing. Once the effect on the R&D intensity is controlled for, their direct effect on OS commitment and amount disappears.

Given the entire evidence of this analysis, the results provide indirect but strong support that scientific publications of firms are used as strategic instruments, where firms seem to carefully balance benefit-cost considerations. Therefore, scientific publications can be regarded as more than just expressions of an internal concession in respect to firm scientists but are part of a more complex decision-making process[18].

7 Conclusion and directions for future research

This article discusses the firms' position towards Open Science (OS), extending the simplifying view of science as a simple source for companies by regarding them also as active contributors. We distinguish between three major dimensions which influence the firms' cost-benefit calculations concerning the disclosure in a scientific format and propose a corresponding structural model in order to handle endogeneity and simultaneity issues correctly. The three dimensions refer to an isolated firm perspective, interactions with academic institutions and overall sector openness. For the empirical analysis, a unique and rich dataset is built from firm level data and scientific publications.

To begin with, it is to emphasize that almost every sector shows an increasing rate of firms contributing to the scientific literature over time. This underlies the relevance of academic studies covering scientific activities and disclosure of firms. The results of the multivariate analysis provide evidence that the decision to participate in OS is not only an internal decision of the firm but is dependent on interactions with academic partners. Firms which are drawing from the public science base show a stronger openness, indicating that the access to these

[18] In order to support this interpretation, we performed a couple of interviews in large, research intensive firms. Indeed, the interviewees confirmed that the decision and the degree to publish in scientific journals are dependent on several considerations where for instance a general openness in the sector can be an important criterion.

academic scientists and institutions requires an adoption of the academic norm of disclosure. Besides, also sector conditions are influential for the firms' decision to participate in OS. This refers in particular to the role of incentives (visibility) if competing firms publish and the protection means which lowers the threshold to publish.

The findings have some important implications for both scholarly research and practitioners in firms. First, our results indicate that concerns of scholars about threatened academic disclosure in collaborations can be incomplete since interactions with academia are also inducing a greater openness to the corresponding firms. Besides, results are also important for the discussion about Open Innovation since they suggest that the provision of financial incentives for accessing academic knowledge is not necessarily sufficient. This implies also for managers in companies that it can be counterproductive to regulate knowledge outflows, i.e. allowing their scientists to publish or not, too strictly.

Likewise every study, this article is not without limitations which offer simultaneously avenues for further inquiries. Even though we include publication data from a pre- and post-survey period, the firm level data set is from a cross-sectional nature and therefore this study can only provide a very limited view on potential dynamics. It would be desirable to consider firm-level information for more than just one CIS wave, yet the sampling procedure of the CIS surveys lead to a relatively low overlap between the different versions. As a consequence, it is challenging in practical terms to provide such a more dynamic perspective. Further extensions might also cover other reasons for OS contributions. For instance with focus on diffusion related purposes where scientific publications can serve as a medium too. So far, there is no quantification to which extent marketing or standard-setting are meaningful driver for scientific publications of firms.

Besides, we have not distinguished OS contributions in terms of academic quality. Since our sample includes a wide range of sectors as well as both articles and conference proceedings (some sectors focus more on the latter), a consideration of scientific quality might be subject of potential bias. Also the recent timeframe of the analysis (scientific publications until 2007) makes the inclusion of citations not very meaningful. Due to this we were hesitating to extend our analysis to quality pattern. Further work might focus on particular sectors or scientific subject areas with consideration in order limit the potential bias of quality indicators as impact factors or citations counts.

References

Allen, R.C., 1983. Collective Invention. Journal of Economic Behavior and Organization 4, 1-24.

Arundel, A., 2001. The relative effectiveness of patents and secrecy for appropriation. Research Policy 30, 611-624.

Arundel, A., Geuna, A., 2004. Proximity and the use of public science by innovative European firms. Economics of Innovation and New Technology 13 (6), 559 - 580.

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

Bruneel, J., D'Este, P., Salter, A., 2010. Investigating the factors that diminish the barriers to university-industry collaboration. Research Policy 39, 858 - 868.

Carayol, N., 2003. Objectives, agreements and matching in science-industry collaborations: reassembling the pieces of the puzzle. Research Policy 32, 887 - 908.

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., Stern, S., 1999. Balancing Incentives: The Tension Between Basic and Applied Research. NBER Working Paper No. w6882.

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 (22), 157 – 182.

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

Cohen, W.M., Nelson, R.R., Walsh, J.P., 2000. Protecting their intellectual assets: Appropriability conditions and why U.S. manufacturing firms patent (or not). NBER Working Paper w7552.

Cohen, W.M., Nelson, R.R., Walsh, J.P., 2002. Links and Impacts: The Influence of Public Research on Industrial R&D. Management Science 48(1), 1-23.

Cohen, W.M., Sauermann, H., 2010. What makes them tick? Employee Motives and Firm Innovation. Management Science (forthcoming).

Crepon, B., Duguet, E., Mairesse, J., 1998. Research, Innovation and Productivity : An Econometric Analysis at the Firm level. Economics of Innovation and New Technology 7 (2), 115-158.

Czarnitzki, D., Grimpe, C., Toole, A., 2011. Delay and secrecy: Does industry sponsorship jeopardize disclosure of academic research? Working paper, KU Leuven.

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

Ding, W., 2010. The impact of founder professional education background on the adoption of Open Science by for-profit biotechnology firms. Management Science (forthcoming).

EC (European Commission), 2003. 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*

EP (European Parliament), 2006. Regulation (EC) No. 1924/2006 of the European Parliament and of the Council on Nutrition and Health Claims made on foods. WebURL: *http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:404:0009:0025:EN:PDF*

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, 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, 309-323.

Gans, J., Murray, F., Stern, S., 2010. Contracting over the Disclosure of Scientific knowledge: Intellectual Property Protection and Academic Publication. SSRN Working paper 1559871.

Geuna, A., Nesta, L., 2006. University patenting and its effects on academic research: The emerging European evidence. Research Policy 35 (6), 790-807.

Gittelman, M., Kogut, B., 2003. Does Good Science Lead to Valuable Knowledge? Biotechnology firms and the Evolutionary Logic of Citation Patterns. Management Science 49 (4), 366-382

Häussler, C., 2010. Information-sharing in academia and the industry: A comparative study. Research Policy 40 (1), 105-122.

Harabi, N., 1995. Appropriability of technical innovations – An empirical analysis. Research Policy 24, 981-992.

Harhoff, D., 1996. Strategic Spillovers and Incentives for Research and Development. Management Science 42 (6), 907–925.

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.

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

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.

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, 131-150.

Levin, R.C., Klevorik, A.K., Nelson, R.R., Winter, S.G., 1987. Appropriating the Returns from Industrial Research and Development. Brookings Papers on Economic Activity 1987 (3), 783 – 831.

Lhuillery, S., Pfister, E., 2009. R&D cooperation and failures in innovation projects: Empirical evidence from French CIS data. Research Policy 38, 45-57.

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.

Mairesse, J., Mohnen, P., 2010. Using Innovation Surveys for Econometric Analysis. NBER Working paper w15857.

Merton, R., 1973. The Sociology of Science. University of Chicago Press, Chicago.

Muller, P., Penin, J., 2006. Why do firms disclose knowledge and how does it matter? Journal of Evolutionary Economics 16, 85–108.

Murray, F., 2002. Innovation as co-evolution of scientific and technological networks: exploring tissue engineering, Research Policy 32, 1389-1403.

Murray, F., O'Mahony, S., 2007. Exploring the foundations of Cumulative Innovation – Implications for Organization Design. Organization Science 18 (6), 1006-1021.

Nelson, R.R., 1959. The simple economics of basic scientific research. The Journal of Political Economy 67 (3), 297-306.

Parchomovsky, G., 2000. Publish or Perish. Michigan Law Review 98 (4), 926-952.

Penin, J., 2007. Open Knowledge Disclosure: An Overview of the Evidence and Economic motivations. Journal of Economic Surveys 21 (2), 326-347.

Roodman, D., 2009. Estimating Fully Observed Recursive Mixed-Process Models with CMP. Center for Global Development, Working Paper 168.

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

Sauermann, H., Stephan, P.E., 2010. Twins or strangers? Differences and similarities between industrial and academic science, NBER working paper No. 16113

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

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.

Tijssen, R.J.W., 2004. Is the commercialization of scientific research affecting the production of public knowledge? Global trends in the output of corporate research articles. Research Policy 33, 709-733.

Von Hippel, E., Von Krogh, G., 2003. Open Source and the „Private-Collective" Innovation Model: Issues for Organization Science. Organization Science 14 (2), 209-223.

Tables

Table 1: Publishing firms by period and sector

Industry (NACE 2)	Total Firms	Publishing Firms per Period								
		1999-2007			1999-2001		2002-2004		2005-2007	
	Q	Q	%	% of total pub. firms	Q	%	Q	%	Q	%
Food Beverages (15)	360	24	6.7%	7.1%	6	1.7%	13	3.6%	17	4.7%
Textiles (17)	99	6	6.1%	1.8%	2	2.0%	1	1.0%	5	5.1%
Apparel (18)	42	0	0.0%	0.0%	0	0.0%	0	0.0%	0	0.0%
Leather (19)	31	0	0.0%	0.0%	0	0.0%	0	0.0%	0	0.0%
Wood products (20)	49	0	0.0%	0.0%	0	0.0%	0	0.0%	0	0.0%
Pulp & Paper (21)	85	2	2.4%	0.6%	1	1.2%	0	0.0%	1	1.2%
Media (22)	49	2	4.1%	0.6%	0	0.0%	2	4.1%	1	2.0%
Petroleum (23)	18	4	22.2%	1.2%	1	5.6%	1	5.6%	3	16.7%
Chem. & Pharmaceuticals (24)	246	96	39.0%	28.3%	51	20.7%	70	28.5%	78	31.7%
Rubber & Plastics (25)	159	13	8.2%	3.8%	8	5.0%	9	5.7%	6	3.8%
Mineral products (26)	113	14	12.4%	4.1%	6	5.3%	8	7.1%	8	7.1%
Basic metall (27)	69	15	21.7%	4.4%	7	10.1%	7	10.1%	9	13.0%
Metal products (28)	168	12	7.1%	3.5%	5	3.0%	7	4.2%	7	4.2%
Machinery & Equipment (29)	251	28	11.2%	8.3%	7	2.8%	19	7.6%	17	6.8%
Office machines (30)	12	2	16.7%	0.6%	1	8.3%	2	16.7%	0	0.0%
Electrical machinery (31)	129	26	20.2%	7.7%	13	10.1%	19	14.7%	17	13.2%
Comm. Equipment (32)	86	30	34.9%	8.8%	17	19.8%	25	29.1%	28	32.6%
Medical & Optical instruments (33)	125	35	28.0%	10.3%	19	15.2%	22	17.6%	30	24.0%
Motor vehicles (34)	144	12	8.3%	3.5%	3	2.1%	10	6.9%	3	2.1%
Transport equipment (35)	67	18	26.9%	5.3%	9	13.4%	12	17.9%	14	20.9%
Total	**2302**	**339**	**14.7%**	**100%**	**156**	**6.8%**	**227**	**9.9%**	**244**	**10.6%**

Table 2: Summary statistics

	Full sample (2.302 Observations)				Publishing in 2005-2007		
					Yes (244)	No (2058)	
Variables	**Mean**	**Std.Dev.**	**Min**	**Max**	**Mean**	**Mean**	**diff p**
Scientific publications							
PUByes (τi): Publication	0.11	0.31	0.0	1.0	1.00	0.00	
Publication categ: 1 Publication	0.03	0.18	0.0	1.0	0.33	0.00	
2-5 Publications	0.04	0.19	0.0	1.0	0.35	0.00	
>5 Publications	0.03	0.18	0.0	1.0	0.32	0.00	
Pubyes: pre-survey period (99-01)	0.07	0.25	0.0	1.0			
$RDint_i$							
R&D share of turnover	0.03	0.05	0.0	0.5	0.05	0.02	0.000
R&D intensity (per employee)	5.73	17.77	0.0	518.0	13.85	4.77	0.000
$Sources_i$ and $Collab_i$							
Competitors	0.11	0.31	0.0	1.0	0.19	0.10	0.000
Clients	0.36	0.48	0.0	1.0	0.44	0.36	0.008
Supplier	0.17	0.37	0.0	1.0	0.13	0.17	0.099
Conferences	0.08	0.27	0.0	1.0	0.13	0.07	0.002
Journals	0.07	0.25	0.0	1.0	0.14	0.06	0.000
Source Academia	0.06	0.24	0.0	1.0	0.18	0.05	0.000
Coop Academia	0.28	0.45	0.0	1.0	0.68	0.23	0.000
$Approp_i$ and $SectCond_i$							
Patent	0.51	0.50	0.0	1.0	0.86	0.47	0.000
Secrecy	0.43	0.49	0.0	1.0	0.70	0.39	0.000
Complexity	0.36	0.48	0.0	1.0	0.52	0.34	0.000
Tech.advantage	0.48	0.50	0.0	1.0	0.66	0.46	0.000
Trademarks	0.50	0.50	0.0	1.0	0.69	0.48	0.000
Designs	0.33	0.47	0.0	1.0	0.28	0.34	0.075
Copyrights	0.10	0.30	0.0	1.0	0.19	0.09	0.000
Z_i							
Public Funding	0.48	0.50	0.0	1.0	0.73	0.45	0.000
EU/National funding	0.24	0.43	0.0	1.0	0.50	0.21	0.000
Employees	468.95	941.09	20.0	15025.0	1420.01	356.20	0.000
Foreign Group	0.32	0.47	0.0	1.0	0.42	0.31	0.000
National Group	0.48	0.50	0.0	1.0	0.55	0.47	0.025
Independent Firm	0.20	0.40	0.0	1.0	0.03	0.22	0.000
HT	0.14	0.35	0.0	1.0	0.45	0.11	0.000
MH	0.31	0.46	0.0	1.0	0.32	0.31	0.770
MF	0.24	0.42	0.0	1.0	0.14	0.25	0.000
FT	0.31	0.46	0.0	1.0	0.10	0.34	0.000

Table 3: Regression outputs – Contribution to OS

Variables	*PUBYes (τ)*				*PUBQ**	
	(1)		**(2)**		**(1)**	**(2)**
	coef (se)	**mg. eff.**	**coef (se)**	**mg. eff.**	**coef (se)**	**coef (se)**
RDint $_i$						
R&D intensity (in logs)	0.475***	*0.066*	0.468***	*0.062*	0.409***	0.402***
	(0.074)		(0.079)		(0.143)	(0.156)
Collab $_i$						
Source Academia	0.408***	*0.066*			0.503***	
	(0.072)				(0.148)	
Coop Academia	0.035	*0.005*			0.094	
	(0.094)				(0.109)	
Academia: Only Source			0.612***	*0.116*		0.556***
			(0.102)			(0.179)
Academia: Source & Coop			0.447***	*0.073*		0.593***
			(0.107)			(0.195)
Academia: Only Coop			0.183	*0.027*		0.134
			(0.120)			(0.133)
SectCond $_i$						
Legal Protection - sector	2.779***	*0.386*	2.852***	0.376	2.476	2.408
	(1.027)		(1.032)		(1.552)	(1.609)
Informal Protection - sector	-1.869**	*-0.259*	-1.874**	*-0.247*	-1.228	-1.189
	(0.735)		(0.749)		(1.283)	(1.363)
Competitor spillovers - sector	0.533	*0.074*	0.428	*0.056*	0.920	0.846
	(1.214)		(1.210)		(1.767)	(1.832)
Publications - sector	5.055***	*0.701*	4.917**	*0.648*	7.341***	7.493***
	(1.913)		(1.941)		(1.897)	(1.901)
Z_i (Controls)						
EU/National funding	0.259	*0.040*	0.265***	*0.039*	0.281**	0.276*
	(0.085)		(0.086)		(0.141)	(0.147)
Foreign Group	0.038	*0.005*	0.030	*0.004*	-0.044	-0.052
	(0.072)		(0.072)		(0.143)	(0.150)
Independent Firm	-0.336	*-0.040*	-0.351*	*-0.040*	-0.160	-0.148
	(0.204)		(0.210)		(0.325)	(0.331)
Employees (in logs)	0.361	*0.050*	0.340*	*0.045*	0.002	-0.006
	(0.192)		(0.193)		(0.494)	(0.512)
Employees sq. (in logs)	-0.005	*-0.001*	-0.003	*-0.0004*	0.034	0.035
	(0.019)		(0.019)		(0.044)	(0.046)
New Firm	0.135	*0.020*	0.183	*0.027*	0.395	0.426
	(0.473)		(0.482)		(0.555)	(0.552)
HT	-0.233	*-0.029*	-0.206	*-0.024*	-0.178	-0.155
	(0.251)		(0.263)		(0.402)	(0.423)
MH	-0.554	*-0.067*	-0.560***	*-0.064*	-0.685***	-0.674***
	(0.132)		(0.136)		(0.178)	(0.182)
ML	-0.066	*-0.009*	-0.084	*0.011*	-0.255	-0.255
	(0.15)		(0.152)		(0.219)	(0.232)
Constant	-4.049***				-4.040***	
	(0.867)				0.855	
Number of observations	2302		2302		244	244
Number of clusters	20		20		20	20
Log-Likelihood	-4.958,79		-4.956.10		-4.958,79	-4.956.10

note: *** p<0.01, ** p<0.05, * p<0.1

Marginal effects refer to the impact on the probability of *PUByes*=1

Table A.1: Variables

Variable	Description	Type
PUByes (τ_i): Publication	Published a scientific article or conference proceeding in the period 05-07.	Dummy
$PUBQ_i$: Publication categ (1; 2; 3)	Number of publications grouped into 3 categories: 1; 2 to 5; more than 5 publications.	Categorical
$RDint_i$	*Scientific outcomes and culture*	
R&D intensity	Internal R&D expenditures per employee (in logs).	Continuous
$Collab_i$	*Interactions with Academia*	
Source Academia	Academia (universities, PROs) as a source of information for innovation activities is considered of medium or high importance.	Dummy
Coop Academia	Collaborated with an academic partner.	Dummy
Academia: Source & Coop	Collaborated with an academic partner and considered it as important source.	Dummy
Academia: Only Coop	Collaborated with an academic partner without considering it as an important source.	Dummy
Academia: Only Source	Considered academia as an important source without collaborating with it.	Dummy
Academia: No Source & No Coop	Has not considered academia as an important source and has not collaborated with it.	Dummy
$SectCond_i$	*Sector Openness*	
Legal protection - sector	Industry average (NACE2) of firms using legal appropriation means (i.e. patents, utility models, designs, trademarks or copyrights).	Continuous
Informal protection - sector	Industry average (NACE2) of firms using strategic protection means (i.e. secrecy, complexity or technological lead).	Continuous
Competitor spillovers - sector	Industry average (NACE2) of firms considering their competitors as a medium or high important source of information for innovation.	Continuous
Publications - sector	Industry average (NACE2) of firms with at least one scientific publication in the period 1999-2001.	Continuous
Z_i	*Controls*	
Public Funding	Company received public support for its innovation activities (at all institutional levels).	Dummy
EU/National funding	Company received public support for its innovation activities from the EU or National government institutions or agencies.	Dummy
Employees (in log)	Number of employees in 2002 (in logs).	Continuous
Independent Firm	Firm is independent and does not belong to a group.	Dummy
Foreign Group	Firm belongs to group based abroad.	Dummy
New Firm	Firm age < 4 years.	Dummy
HT; MH; ML; LT	Four industry dummies according to their technology intensity (OECD). Include High Tech (HT), Medium High Tech (MH),	Dummies
	Medium Low Tech (ML), Low Tech (LT).	Dummies
$Sources_i$	Set of variables covering different "sources of information for innovation activities" with medium or high importance (based on 4-point scale). Contain Competitors, Clients, Supplier, Conferences, Journals, Universities and PROs. Additionally: Cooperation with universities. For R&D equation.	Dummies
$Approp_i$	Set of variables indicating the use of formal protection instruments (Patents, Trademarks, Utility models, Designs and Copyrights) or informal means (Secrecy, Complexity and Technological lead) on firm level. For R&D equation only.	Dummies

Table A.2: Correlations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
(1) Publication	1.00																	
(2) R&D intensity (in logs)	0.24	1.00																
(3) Source Academia	0.29	0.16	1.00															
(4) Coop Academia	0.32	0.20	0.52	1.00														
(5) Academia Source & Coop	0.30	0.17	0.80	0.72	1.00													
(6) Academia Only Coop	0.05	0.01	0.50	-0.17	-0.12	1.00												
(7) Academia Only Source	0.10	0.08	-0.20	0.58	-0.16	-0.10	1.00											
(8) Academia No Source & No Coop	-0.32	-0.19	-0.76	-0.84	-0.61	-0.38	-0.49	1.00										
(9) Legal protection - sector	0.22	0.26	0.09	0.12	0.09	0.02	0.06	-0.12	1.00									
(10) Informal protection - sector	0.21	0.25	0.12	0.15	0.12	0.02	0.07	-0.15	0.75	1.00								
(11) Competitor spillovers - sector	0.18	0.27	0.08	0.13	0.08	0.01	0.09	-0.13	0.46	0.47	1.00							
(12) Publications - sector	0.33	0.29	0.17	0.18	0.16	0.05	0.06	-0.19	0.61	0.66	0.52	1.00						
(13) Public Funding	0.17	0.20	0.22	0.29	0.23	0.04	0.14	-0.29	0.10	0.14	0.14	0.18	1.00					
(14) EU/National funding	0.21	0.12	0.24	0.31	0.28	0.00	0.11	-0.29	0.05	0.11	0.13	0.17	0.58	1.00				
(15) Employees (in log)	0.32	0.05	0.20	0.33	0.24	-0.01	0.18	-0.30	0.06	0.11	0.06	0.09	0.14	0.08	1.00			
(16) Foreign group	0.07	0.04	-0.01	0.05	-0.01	0.01	0.08	-0.05	0.08	0.13	0.09	0.05	-0.08	-0.11	0.32	1.00		
(17) Independent	-0.15	-0.08	-0.07	-0.14	-0.08	-0.01	-0.11	0.14	-0.08	-0.09	-0.05	-0.09	-0.01	0.06	-0.42	-0.35	1.00	
(18) New firm	-0.06	0.00	-0.05	-0.07	-0.03	-0.04	-0.06	0.09	-0.01	-0.02	0.00	0.00	-0.03	0.02	-0.28	-0.13	0.37	1.00

Table A.3: R&D equations

	Rdint			
	(1)		**(2)**	
	coef (se)		**coef (se)**	
Sources $_i$				
Clients	0.134	(0.084)	0.134	(0.086)
Supplier	-0.158**	(0.069)	-0.154**	(0.070)
Competitor	0.093	(0.103)	0.086	(0.102)
Conferences	-0.335**	(0.162)	-0.338**	(0.165)
Journals	0.170	(0.107)	0.171	(0.108)
Source Academia	0.171	(0.131)	0.179	(0.132)
Coop Academia	0.286***	(0.078)	0.285***	(0.079)
Approp $_i$				
Protection: Patent	0.501***	(0.067)	0.502***	(0.068)
Protection: Secrecy	0.428***	(0.056)	0.428***	(0.055)
Protection: Complexity	0.053	(0.072)	0.055	(0.073)
Protection: Tech. Advantage	0.090*	(0.054)	0.088	(0.055)
Z_i				
Public Funding	0.242**	(0.113)	0.241**	(0.113)
Foreign Group	0.004	(0.103)	0.004	(0.103)
Independent Firm	-0.304***	(0.089)	-0.304***	(0.089)
Employees (in logs)	-0.980***	(0.196)	-0.980***	(0.197)
Employees sq (in logs)	0.076***	(0.018)	0.076***	(0.018)
HT	1.070***	(0.099)	1.070***	(0.099)
MH	0.647***	(0.125)	0.647***	(0.124)
ML	-0.244***	(0.073)	-0.244***	(0.074)
Constant	2.354***	(0.543)	2.352***	(0.543)
Sigma (log)	0.436***	(0.024)	0.436***	(0.024)
Number of observations	2302		2302	
Number of clusters	20		20	
Log-Likelihood	-4'958.79		-4'956.10	

note: *** p<0.01, ** p<0.05, * p<0.1

Table A.4: Error structure matrix

(1)			(2)		
1.547	-0.731***	-0.466	1.547	-0.707***	-0.452
-0.731***	1	1.59***	-0.707***	1	1.512***
-0.466	1.59***	1	-0.452	1.512***	1

Table A.5: Robustness check Public R&D Funding

	PUByes (τ)			*PUBQ**	
	coef	**se**	**mg. eff.**	**coef**	**se**
RDint $_i$					
R&D intensity (in logs)	0.464***	(0.080)	*0.060*	0.389**	(0.180)
Collab $_i$					
Source Academia (only)	0.695***	(0.139)	*0.137*	0.672***	(0.232)
Coop Academia (only)	0.218	(0.169)	*0.032*	0.010	(0.185)
EU/National funding (only)	0.367***	(0.137)	*0.058*	0.125	(0.207)
Academia: Source & Coop (NO pub.funding)	0.487***	(0.125)	*0.084*	0.532***	(0.172)
Source Academia & EU/National funding (NO Coop Academia)	0.778***	(0.185)	*0.165*	0.395	(0.335)
Coop Academia & EU/National funding (NO Source Academia)	0.471**	(0.203)	*0.082*	0.473*	(0.283)
Source Academia & Coop Academia & EU/National funding	0.735***	(0.191)	*0.144*	0.880**	(0.378)
(Baseline: No Source Academia, No Coop Academia, No EU/National funding					
SectCond $_i$					
Legal Protection - sector	2.892***	(1.048)	*0.372*	2.332	(1.621)
Informal Protection - sector	-1.881***	(0.716)	*-0.242*	-1.089	(1.307)
Competitor spillovers - sector	0.459	(1.198)	*0.059*	1.056	(1.977)
Publications - sector	4.950***	(1.896)	*0.636*	7.766***	(1.789)
Z_i *(Controls)*					
Foreign Group	0.033	(0.071)	*0.004*	-0.061	(0.147)
Independent Firm	-0.350*	(0.206)	*-0.039*	-0.170	(0.310)
Employees (in logs)	0.333*	(0.196)	*0.043*	-0.104	(0.608)
Employees sq (in logs)	-0.002	(0.020)	*-0.000*	0.043	(0.054)
New Firm	0.179	(0.493)	*0.026*	0.408	(0.554)
HT	-0.206	(0.265)	*0.024*	-0.134	(0.467)
MH	-0.561***	(0.138)	*-0.062*	-0.649***	(0.203)
ML	-0.086	(0.153)	*0.011*	-0.252	(0.253)
Constant / Cutpoints	-4.092***	(0.867)			
Rho equation 12	-0.692***	(0.222)			
Rho equation 13	-0.421	(0.378)			
Rho equation 23	1.394***	(0.010)			
Number of observations	2302			244	
Number of clusters	20			20	
Log-Likelihood			-4'953.54		

note: *** p<0.01, ** p<0.05, * p<0.1