Open Science as a Signaling Device: Evidence from Firm Publications*

Sharon Belenzon
Duke University and NBER

Andrea Patacconi
University of Oxford

May 19, 2008

Abstract

This paper argues that private firms may contribute to the creation and dissemination of scientific knowledge in order to signal the quality of their research to the capital market. A simple model of financing with adverse selection is presented where academic publications convey information about a firm's ability to carry out a project. We find that young firms with a short track record have stronger incentives to publish than mature firms. This difference is amplified in industries where performance varies substantially across firms and in highly developed financial markets. Using a novel and comprehensive database on academic publications, we test these predictions. Our results suggest that signaling is an important reason for private firms to publish in academic journals. We find little support for the idea that patents may perform an analogous function.

Keywords: Open Science, Certification, Publications, Patents, Capital Markets

JEL Classification: O31, O32, O16

*We would like to thank Tomer Berkovitz, Ian Cockburn, Clare Leaver, Josh Lerner, Rocco Machiavello, Mark Schankerman, Scott Stern and John Van Reenen for very helpful discussions and remarks. Financial support from the British Academy and the Fell Fund is gratefully acknowledged. All remaining errors are ours.

1. Introduction

In the last forty years economists have increasingly appreciated the importance of institutions in mitigating problems of quality uncertainty. Akerlof himself devoted part of his path-breaking 1970 article to illustrating the relevance and ubiquitousness of guarantees, brand names and certification. He noted for instance that "[m]ost skilled labor carries some certification indicating the attainment of certain levels of proficiency. The high school diploma, the baccalaureate degree, the Ph.D., even the Nobel Prize, to some degree, serve this function of certification. And education and labor markets themselves have their own "brand names"" (p.500).

This paper empirically investigates the role of certification in one particular but important setting: the production and dissemination of scientific knowledge. The key idea we put forward is that private firms may encourage their employees to publish in academic journals (thus engaging in 'open science') in order to signal the quality of their research to prospective clients and investors.[1]

To take a concrete example consider MediGene, a German publicly quoted biotechnology company whose aim is to integrate "all core areas of a modern biopharmaceuticals company from research to drug development and commercialization".[2] It was recently awarded a research grant of approximately 600,000 Euros from the Federal Ministry of Education and Research to engage in innovation "in the field of top-level research" and is currently looking for external financing for its mTCR Program. Its workforce is highly skilled and research active.[3] Our conjecture is that for firms like MediGene, publishing in academic journals is an important way to signal the quality of their research to private investors and public funding bodies.[4]

[1] In this paper the terms signaling and certification will be used interchangeably.

[2] More information about MediGene can be found at http://www.medigene.de/englisch/index_e.php.

[3] For an example of a MediGene publication, which is included in our dataset, see Figure A1.

[4] Another example is HR Wallingford, an Oxford-based consulting firm specializing in civil engineering and environmental hydraulics. Consulting firms often employ respected academics to certify the quality of their recommendations. In the case of Wallingford, its employees have published extensively in academic

Our argument crucially relies on academic publications being an useful signal of the quality of a firm's R&D. There are several reasons why this should be the case. First and foremost, when a firm's employee publishes in a scientific journal, the academic community *certifies* (through the peer review process) that the research in question conforms to scientific standards. This is likely to enhance its credibility, for instance, if it reports that a new drug was useful in treating a certain pathology. Academic publications also convey good news about the quality of a firm's research unit. A recent publication record, in particular, indicates that its owner is in touch with the latest developments of his or her discipline. And last but not least, compared to alternatives such as patents, publications may be a fast and cheap way for firms to signal competence to outsiders.[5]

In this paper we make two important contributions. The first one is to document the importance of private corporations in the advancement of basic scientific knowledge. To this end we assembled a novel and comprehensive database on academic publications. We matched European firms by name (using authors' affiliations) to the complete Thomson's ISI Web of Science, which covers about 20 million publications in thousands of international journals in 'hard' sciences, such as physics and biochemistry. Financial information is from Amadeus, a dataset of about 8 million European firms. For each publication we have information on the number of times it has been cited as well as on the quality of the journal in which the article was published, which we use to control for the quality of publications.

We find that private firms contribute substantially to the creation of new scientific knowledge. We matched more than 230 thousand publications to firms over the period 1970-2004. The trend is clearly upward, with the number of firm publications per year more than quadrupling over the last 20 years in most technology areas. Qualitatively, however, firm publications compare somewhat unfavorably with non-firm publications.

journals. For an example, see Soulsby and Damgaard (2005) (both authors are Wallingford's employees).

[5]This is especially true for publications in 'hard' sciences (on which this paper focuses), as there the time elapsing from submission to publication tends to be short.

The average number of citations that a firm publication receives, for instance, is about 7.4. The corresponding figure for publications that we did not match to Amadeus is 10.1.

The second contribution of this paper is to provide systematic evidence in support of a signaling explanation of open science. A simple model of certification is presented to motivate our empirical analysis. In the model all firms need external funding to undertake a project but differ in terms of their ability to carry out innovative research. Thus, because of adverse selection, innovative firms may find it difficult to raise funds. This gives them an incentive to signal their ability by publishing in academic journals.

In this setting, three main predictions are derived. Our starting point is the observation that since mature firms have longer track records and/or more internal funds than young firms, they are less likely to be affected by adverse selection. Thus, under the signaling hypothesis, we should expect young firms to have the strongest incentives to publish. We then explore how the propensity to publish changes as industry and country conditions change. We find that young firms have stronger incentives to publish, compared to mature firms, in more heterogeneous industries and in countries with more developed financial markets. The intuition is that, when performance varies substantially across firms, the value of being ranked among the best tends to be large. Publications should also be more valuable in highly developed financial markets because "financial development should be related to the variety of intermediaries and markets available, [and] the efficiency with which they perform the evaluation, certification, communication and distribution functions" (Rajan and Zingales (1998), p.569). Thus we might expect academic publications to be more thoroughly evaluated (and hence to be more relevant) in highly developed financial markets.

Our econometric analysis supports these predictions. Consistent with our theoretical model, we find that controlling for size young firms publish more than mature firms. Furthermore, young firms tend to publish disproportionately more in more heterogeneous industries and in more developed financial markets. These results are robust to a variety

of controls and different empirical measures of our theoretical concepts. The same broad patterns of results emerge, for instance, when we control for publication quality or when we restrict attention to firms that recently raised external finance by selling equity stakes.

We also use firm-level data on patents (from both the European and the US Patent Offices) to test for the possibility that patents may also perform a signaling function. Interestingly, our interaction coefficients are then insignificant. Thus, for instance, young firms do not appear to patent disproportionately more than mature firms in more heterogeneous industries. We find this reassuring since, although it has been suggested that aggressive patenting might sometimes be used to attract investments from venture capitalists (Hall and Ziedonis (2001)), the primary function of the patent system is arguably to protect intellectual assets.[6]

We haste to stress that, beside signaling, many other factors may influence a firm's decision to engage in open science. Firms, for instance, may encourage their researchers to publish in order to 'plug in' to research conducted elsewhere (Cohen and Levinthal (1989), Cockburn and Henderson (1998)). Publications may be a by-product of R&D investments, especially those directed toward the basic-science end of the spectrum (Murray (2002), Murray and Stern (2006)). In general, we view these explanations as complementary. While in fact many scientific discoveries are made during R&D, a pro-publication culture may in part be due to a need to certify progress. Our finding that relatively large, young firms are the most likely to publish, in particular, suggests that both initial investments in R&D and signaling considerations (to raise even more funds) could be important.[7]

To the best of our knowledge this paper is the first to suggest that firms may use academic publications as a signaling device. Related ideas, however, have appeared elsewhere. Lichtenberg (1986,1988), for instance, has argued that the procurement method known as 'by design and technical competition' is deliberately used by the US government

[6]See Cohen et al. (2000) for an insightful survey of why (or why not) firms patent.

[7]The relationship between innovation and firm size is explored in greater depth in Belenzon et al. (2008).

to encourage private investments in military R&D. Private firms, in particular, appear to use elaborate technical proposals to signal their ability to perform government contracts. Another relevant paper is Hall and Ziedonis (2001). They report interviews with practitioners in the US semiconductor industry suggesting that, especially for young specialized design firms, aggressive patenting may be a way to attract venture capital funds. Their econometric analysis, however, highlights mainly the role of patents in reducing concerns about being held up by external patent holders.

This paper is also related to a growing literature emphasizing the many different ways through which firms can signal favorable information to the capital market. Carter and Manaster (1990), for instance, test the view that low risk firms attempt to reveal their type to the market by selecting underwriters with high prestige. Megginson and Weiss (1991) argue that venture capitalists may perform a certification function in initial public offerings (IPOs). Consistent with their hypothesis, they find that the level of underpricing and the amount of compensation to underwriters is lower for venture capital (VC) backed firms than for non-VC backed firms. Gompers (1996) finds that young VC firms bring companies public earlier than older VC firms in an attempt to signal their ability to potential investors. Finally, Hirtle (2006) investigates whether the 2002 SEC order to certify the accuracy of the financial statements of publicly traded firms had a significant effect on the stock market valuation of bank holding companies. She finds that these companies experienced large abnormal returns on the day of certification, especially those characterized by greater opaqueness.

The remainder of the paper is organized as follows. Section 2 presents the theoretical framework motivating the empirical analysis. Sections 3 and 4 describe the data and provide descriptive statistics. Section 5 discusses our econometric specification, while Section 6 reports the results. Section 7 concludes.

2. Theoretical Considerations

This section studies a simple model of corporate financing under adverse selection where firms can signal their competence to investors by publishing in academic journals.

We consider an economy populated by a continuum of firms (or entrepreneurs), each seeking to finance a project costing I. All agents are risk neutral and the interest rate in the economy is normalized to 0. Firms are either innovative ('good') or traditional ('bad'), with equal probability. If the project is financed, profits are $\alpha_G V(I)$ for innovative firms and $\alpha_B V(I)$ for traditional firms, where $\alpha_G = \alpha_B + \Delta\alpha$. If the project is not financed, profits are $\alpha_i V(0)$ for $i = G, B$. Thus $\Delta\alpha > 0$ is a measure of industry heterogeneity. We assume that $\alpha_G(V(I) - V(0)) > I$ (or otherwise it would never be efficient to invest) and that $V(0) \geq 0$, since entrepreneurs may already have some assets in place. For simplicity α_B is set equal to (approximately) zero. Thus investors never want to finance bad firms.

We distinguish between mature and young firms. We assume that because mature firms have a long track record, investors can tell them apart, Thus, the relationship between mature firms and investors is characterized by symmetric information. By contrast, it is difficult for investors to tell young firms apart. We consider a situation where

$$\frac{1}{2}\Delta\alpha V(I) < I \tag{1}$$

so that, on the basis of prior information only, young firms are not be financed. However, by publishing in academic journals, innovative firms may be able to (imperfectly) signal their type to investors. Publications are not perfectly informative because we want to allow for the possibility that investors may vary in their ability to screen borrowers. This is consistent with Cressy et al.'s (2007) argument that specialized private equity firms "possess a deeper knowledge of the competitive environment of acquired companies and their companies' strengths and weaknesses. They are able therefore both to select potentially superior performers and also to provide more effective monitoring and advice once

an investment has been made" (p. 648).[8]

To model differences in investor competence, we consider the following setup. At time 0, innovative firms decide whether or not to publish research related to their core business. If they publish, they incur a cost k that may reflect not only the time and resources that firm scientists must put into writing papers, presenting at conferences, etc., but also the increased risk of expropriation that is associated with disclosure of relevant information. To create a non trivial signal extraction problem, we assume that with probability $\frac{1}{2}$, traditional firms are endowed with 'irrelevant' publications. These publications are not directly related to their core business; one may think of them as research being carried out by firm employees for reasons such as intrinsic motivation, the desire to keep contacts with academic coauthors, etc.

At time 1, firms are randomly matched to investors and can, if they wish, disclose their publications. We parametrize investor competence by $\varkappa$, the probability that an investor is able to discriminate between relevant and irrelevant publications. Investor competence is common knowledge and firms have all the bargaining power when negotiating the terms of the contract with financiers.[9] Throughout, attention is restricted to equity financing since in the empirical part we will focus especially on firms that sold some of their equity (either directly or via IPO).

At time 2 parties agree on the terms of the contract and payoffs accrue.

We can now study the signaling incentives to publish. Our first observation is that mature firms will never engage in open science for signaling reasons. Indeed, since investors can simply use track records to tell firms apart, innovative firms always get financed. Thus engaging in open science would simply reduce their profits from $\Delta\alpha V(I) - I$ to $\Delta\alpha V(I) - I - k$.

[8] Recent evidence also suggests that VCs might have superior knowledge or incentives to evaluate patent porfolios (Mann and Sager (2007)), patent scope (Lerner (1994)) and their technological content (Munari and Toschi (2008)).

[9] Our qualitative results would be the same if firms had only some bargaining power.

Things are considerably more complicated for young firms. Suppose that at time 0 innovative firms decide not to publish. Then at time 1 no traditional firm would disclose its publications to investors (else it would reveal itself to be bad). Investors thus keep their priors and, if (2) holds, innovative firms get

$$\Pi_G^{NS} = \Delta\alpha V(0). \tag{2}$$

Now suppose that at time 0 innovative firms publish. Suppose further that both innovative and traditional firms disclose their publications to investors.[10] The payoff an innovative firm gets depends on the type of investor it is matched with. With probability $\varkappa$, the firm is matched with a competent investor that correctly infers its type and therefore demands an equity stake of (at least) $I/(\Delta\alpha V(I) - k)$. The firm's payoff in this case is therefore $\Delta\alpha V(I) - I - k$. With probability $1-\varkappa$, the investor is unable to evaluate the publication. However, by just observing that the firm published, it infers that the likelihood that the firm is innovative is 2/3. To invest in the firm it will therefore ask for an equity stake of at least s, where s solves

$$\frac{2}{3}s(\Delta\alpha V(I) - k) = I. \tag{3}$$

Obviously this offer will be accepted by the firm only if yields a payoff that is at least as large than what it would get by not investing.[11] Two cases must thus be considered, depending on whether or not an equity stake can be found that is both acceptable to the firm and allows the investor to break even in expectation. Since both cases yield the same qualitative results, we focus on the simpler case where no such equity stake exists. In that case investors do not lend money to young firms and an innovative firm's payoff is just $\Delta\alpha V(0) - k$. Thus, by signaling an innovative firm gets in expectation

$$\Pi_G^S = \varkappa[\Delta\alpha V(I) - I - k] + (1-\varkappa)[\Delta\alpha V(0) - k]. \tag{4}$$

[10] Disclosure is a weakly dominant strategy for traditional firms if they want to invest whenever innovative firms also want to invest.

[11] Thus s must be lower than $\hat{s}$, where $\hat{s}$ solves $(1-\hat{s})[\Delta\alpha V(I) - k] = \Delta\alpha V(0) - k$.

Innovative firms publish if $\Pi_G^S > \Pi_G^{NS}$.[12] This condition can be rewritten as

$$\varkappa[\Delta\alpha(V(I) - V(0)) - I] > k. \tag{5}$$

Condition (5) implies that young firms will be more likely to publish when $\Delta\alpha$ and $\varkappa$ are large and when k is small.[13] Empirically, we will proxy investor competence $\varkappa$ with measures of financial development. The key empirical implications of the model can thus be summarized as follows.

H.1 Under the signaling hypothesis, young firms have stronger incentives to publish in academic journals than mature firms.

H.2 Furthermore, the difference in the propensity to publish between young and mature firms should be more pronounced

(i) in more heterogeneous industries ($\Delta\alpha$ high) and

(ii) in countries with more developed financial markets ($\varkappa$ high).

In this model young firms have a stronger incentive to publish than mature firms because they face a more severe adverse selection problem. We stress, however, that the same pattern of results would emerge in a model where there is always private information but mature firms are not cash constrained. In that setting, in fact, mature firms could use their internal funds to finance their investments. Again, this would leave only the young firms with the need to signal.

The model also suggests that, across industries, young firms should have the greatest incentives to publish when performance varies substantially across firms ($\Delta\alpha$ high) because

[12]This condition assumes that if innovative firms deviate and do not publish, then investors maintain their priors. However, disclosure can be supported for a broader set of parameters if investors have 'pessimistic' expectations. For instance, investors may believe that if a firm does not publish, then it must be bad. This would obviously give firms a stronger incentive to publish. We restrict attention to the characterization in the text because it Pareto-dominates these alternative characterizations. At any rate, the qualitative results are the same.

[13]The relevant condition when an equity stake can be found that is is both acceptable to the firm and allows the investor to break even (so that (3) holds) is that $\Delta\alpha(V(I) - V(0)) - I - \frac{1}{2}(1-\varkappa)I > k$. The comparative statics is thus the same as in the main text.

there the value of being ranked among the best is very large.[14] Moreover, publications should be more valuable in highly developed financial markets where firms can easily find competent investors. The theory does not provide compelling predictions for how the incentives to signal should be related to product market competition and the need for external funds. Note in fact that while product market competition is likely to increase the threat of expropriation and hence k, it could also affect $V(I) - V(0)$. The 'escape-competition' effect (Aghion et al. (2005)), in particular, suggests that the difference $V(I) - V(0)$ should be larger when competition is fierce. Similarly, the right-hand side of (5) may either increase or decrease with external finance dependence (as parametrized by I), depending on the specific functional form of $V(\cdot)$.

We conclude this section with two brief remarks. First, it should be noted that although our discussion emphasizes the role of academic publications, the scope of certification is much wider. In particular, as mentioned earlier, aggressive patenting could be used to attract potential investors. In the empirical analysis we will therefore also consider that possibility. Secondly, our discussion completely abstracts from issues related to the quality of signals. A careful analysis, however, would probably generate ambiguous predictions. In particular, we would expect firms to face a tradeoff between sending stronger signals (such as publishing in a top journal) and the risks associated with such a strategy (acceptance rates are typically lower in top journals). For these reasons we remain agnostic about the implications of the signaling hypothesis for publication quality.

3. Data

This paper combines data from three main sources: (i) academic publications from the Web of Knowledge database, (ii) patents from the EPO and USPTO and (iii) financial

[14] This result would also be true in a setting where firms wish to signal competence to prospective *clients* rather than to prospective investors. In that setting we would in fact expect the incentives to signal to be greater in industries where heterogeneity in quality of the services provided (as parametrized by $\Delta\alpha$) is substantial.

information from Amadeus and Compustat. In this section, we explain our methodology for constructing these data and describe our sample.

3.1. Academic Publications

The goal of this paper is to better understand why private firms may contribute to the creation and dissemination of scientific knowledge. To this end, we constructed a unique dataset on firm publications. The world's largest source of information on academic publications is the Thomson's ISI Web of Knowledge (WOK), which includes publication records on thousands of international journals in 'hard' sciences (such as natural or physical sciences). Each publication has an address field which contains the authors' affiliation. We match all patenting firms by name to the complete ISI database. European research institutions can be incorporated, thus, they appear in Amadeus as potential firms to be matched. To screen out such firms, we follow two steps. First, as for patent matching, we drop Amadeus names that include strings that are associated with research institutions. Second, we manually examine the websites of firms that have a large number of publications but appear as small firms in terms of their sales and number of patents. For these firms, we check whether their primary activity is research. In case the primary activity is research, we exclude them from our matched sample. Almost 30 percent of the organizations matched to the WOK database were identified as research or non-for-profit institutions. Finally, to control for the quality of publication two pieces of information are used. First, we control for the importance of the journal in which the article was published by using the impact factor from the Journal Citations Report. Secondly, we use information about forward citations at the publication level, where a publication is assumed to be of higher quality if it receives more forward citations.

3.2. Patents

Another instrument that firms might use to signal their ability to perform R&D is the number of patents they hold. To test this conjecture, we constructed a unique dataset of European firm patents by matching all granted patent applications from the EPO and the USPTO to the complete list of Amadeus firms (about 8 million firm names) for the period 1979-2004.

3.3. Accounting

Accounting information is taken from Amadeus - a comprehensive pan-European database by Bureau van Dijk Electronic Publishing (BvDEP), which covers both private and public firms. To ensure we examine firms with substantial economic activity and to harmonize firm coverage across countries, we restrict our sample to include only firms that have at least 10 employees and $1 million. The accounting dataset also provides information on the age of the firm. We refer to the firm age as the number of years elapsed since the date of incorporation. However, the date of incorporation has an important limitation. Legal changes, such as mergers and acquisitions, may in fact affect it. This may cause us, for example, to treat a merger of two mature firms as the formation of a new young company. To mitigate this concern, we follow two steps. First, we manually check the age of the largest 500 publishing firms in our sample,using the firm websites and other public sources. We drop firms from the analysis if their age does not correspond to the date of incorporation from Amadeus. Second, we examine all M&A deals during 1997-2004 and manually check the age of about 1,000 firms that participated in these deals and also have at least one academic publication or at least one patent.

4. Descriptive Statistics

The first contribution of this paper is to systematically show that firms publish substantially in academic journals. We matched more than 230 thousand articles in “hard” science

journals to European firms in the period 1970-2004. Figure 1 plots the distribution of firm publications across main technology areas. Most firm publications (31 percent) are concentrated in Biology and Chemistry, 22 percent in Engineering and 21 percent in Health and Medicine. The trend is clearly upward. As Figure 2 shows, in the last 20 years the number of firm publications per year has more than quadrupled in most technology areas. For instance, during the period 1990-2004 the number of firm publications per year rose from less than 1000 to more than 4000 in Health and Medicine, and from about 200 to approximately 1000 in Computer Science.

Table A1 provides information on the quality of firm publications. On average, an article receives more than 7 citations, but this figure varies substantially across fields, from a minimum of about 2.5 citations in Computer science to a maximum of 11 citations in Biology and Chemistry. As a comparison, the average of citations received by the non-firm publications (that is, publications we have not matched to Amadeus) is 10.1 (a median of 2). Firm publications also compare somewhat unfavorably with non-firm publications when we look at the impact factor of the journals in which the articles were published. The figures here are 2.14 on average for firm publications and 2.5 for all other publications. The fact that firm publications appear to be of lower quality compared to non-firm publications is not surprising in our view. Firm publications are in fact likely to be directed toward applied topics. Furthermore, researchers employed by firms may get less exposure (and thus less citations) than their academic counterparts.[15]

[**FIGURES 1 AND 2 ABOUT HERE**]

Our theory suggests that firms with a short track record may want to use publications

[15] Table A2 examines publications for firms in our estimation sample (see table 1). We split the publications of these firms according to the median firm age. Publications by mature firms appear to receive and make more citations compared to publications by young firms. For example, in Biology and Chemistry, an average firm publication receives about 12 citations, while for young firms the average is of approximately 9. Table A3 reports similar statistics when we restrict attention to publications in leading journal only. (Leading journals are defined as those in the highest quartile of the journal impact factor, as indicated by the Journal Citations Report index.) A similar pattern emerges, with publications by mature firms receiving more citations than publications by young firms.

to certify and enhance their reputation. We use firm age as a measure of the length of the firm track record. Figure 3 plots the relationship between firm age and the share of the firms in our sample that publish at least one article out of all innovating firms (where innovating firms include all firms that either patent or publish). We divide firms into quartiles according to their age. The share of firms that publish in the first quartile is substantially higher than the share of firms that publish in the forth quartile (0.32 versus 0.20). This pattern is consistent with the conjecture that publications are especially valuable to firms without an established track record. Note that a priori one should expect mature firms to be more likely to publish because they are observed over a longer period. Finding the opposite furthers supports the signaling hypothesis.

Figure 4 examines the relationship between firm age and publishing and patenting intensities. The pattern of results suggests that young firms have a higher ratio of both publication and patent stocks to sales. This pattern is again consistent with the signaling hypothesis.[16]

[FIGURES 3 AND 4 ABOUT HERE]

Table 1 reports summary statistics for firms in our estimation sample. To be included in our estimation sample a firm has to satisfy three conditions: (i) report sales and employment, (ii) have at least one patent or at least one academic publication in the period 1978-2004 and (iii) have at least 10 employees and $1 million in annual sales. About 9 thousand firms have at least one patent or one academic publication between 1978 and 2004. About 3 thousand firms have at least one academic publication, while more than 7 thousand firms have at least one patent. Our sample covers a wide distribution of firm size, especially in the lower tail. The median firm in our sample generates about $30 million in annual sales and has 165 employees. 10 percent of the firms in our sample have

[16] An alternative interpretation of this figure is that there is some non-linearity in the relationship between firm size and innovation. In case the relation between size and innovation is convex we would expect a similar pattern as in figure 4 if young firms are smaller than mature firms. In the econometric estimation we also control for non-linearity in firm size to mitigate this concern.

less than 26 employees and less than $4 million in annual sales.[17] The median firm age is 22 years and the average age is 26. 10 percent of the firms are 6 years of age or younger.

Table 2 reports summary statistics, separately for patenting and publishing firms. Panel A includes only firms that have at least one patent. The average firm has 1,481 employees with a median of 175. On average, these firms have close to 1 patent a year and a stock of patents (assuming a 15 percent annual depreciation rate) of 7 patents. The median age of a patenting firm is 23, with an average age of 29. Panel B of table 2 examines only firms with at least one academic publication. The average publishing firm is larger than the average patenting firm. It has 2,861 employees with a median of 208 and. These firms publish about 1 article a year and have a stock of 4 publications. Publishing firms are slightly younger than patenting firms. The median age of a publishing firm is 19, with an average age of 28.

[TABLES 1 AND 2 ABOUT HERE]

5. Econometric Modeling

We use the Negative Binomial model to analyze our publication count data. The key advantage of this model is that it relaxes the Poisson distributional assumption of equality of the conditional variance and mean by allowing for overdispersion.[18] In our baseline specification we estimate a conditional expectation of the form

$$E(P_{it}|X_{it}) = \exp\{\beta_1 Age_{it} + \beta_2 \ln Sales_{it-1} + Z'_{it}\beta_4 + \varphi_j + \tau_t\}$$

where P_{it} is a count of the number of publications by firm i at time t and X_{it} is a vector of regressors. Age_{it}, defined as the number of year elapsed since that year of incorporation, is the main variable of interest. $Sales_{it-1}$ is used to control for firm size (we use lagged value to mitigate transitory shocks that can affect both the incentive to innovate and sales).

[17] As a comparison, Compustat patenting firms have on average $3 billion in annual sales with a median of $500 million (Bloom, Schankerman and Van Reenen (2005)).

[18] See Wooldridge (2002) and Hausman et al. (1984) for discussions of count data models of innovation.

Z_{it} is a vector of controls (such as a complete set of country dummies) and φ_j and τ_t are complete sets of three-digit industry SIC and year dummies. The signaling hypothesis implies that young firms with a shorter track record should be more likely to publish for any given firm size. Thus, we expect $\beta_1 < 0$. We also estimate equivalent specifications where the dependent variable is the number of patents and probit specifications where the dependent variable is an indicator to whether the firm has ever published.

To further explore the signaling hypothesis, we examine how the effect of firm age on publication intensity varies across exogenous industry conditions. We focus mainly on three industry characteristics: the degree of industry heterogeneity in firm performance, dependence on external funds, and the Lerner competition index. To measure industry heterogeneity we follow Acemoglu et. al. (2007) and compute Productivity Growth Dispersion, defined as the difference in the three-year average productivity growth between the 90th and 10th percentiles in a specific industry (we also compute this measure as the difference between the 95th and 5th percentiles and the standard deviation of the industry productivity growth). A high dispersion means that firm performance is affected more by idiosyncratic factors and less by aggregate shocks and systematic industry factors. Thus it should hard for potential lenders to learn about the quality of an innovating firm by examining aggregate information about the industry in which the firm operates. We compute Productivity Growth Dispersion from the complete set of Amadeus firms for each industry.

To measure dependence on external funds at the industry level we use External Finance Dependence. This is defined as the ratio between capital expenditures minus cash flow from operations and capital expenditures. To compute it, we follow Rajan and Zingales (1998) and rank industries according to their dependence on external funds. These rankings are based on data on US Compustat firms.[19] To be applicable to our European firms,

[19] As discussed by Rajan and Zingales (2000), using US firms has important advantages: (i) Since the US market is one of the most advanced capital markets in the world, large publicly-traded firms face the least frictions in accessing finance. This means that the amount of external finance used by these

two assumptions are needed. First, technological differences should explain why some industries rely on external funds more than others. Secondly, these differences should persist across countries. Note also that we face a practical limitation in computing measures of external dependence from Amadeus since we have no information on capital expenditures for European firms.

The Lerner index is a commonly used (inverse) proxy for product market competition which is computed as the industry average ratio between profits and revenues.

Our empirical specification becomes

$$\begin{aligned} E(P_{it}|X_{it}) &= \exp\{\beta_1 Age_{it} + \beta_2 Age_{it} \times Industry\ Measure_j \\ &\quad + \beta_2 \ln Sales_{it-1} + Z'_{it}\beta_4 + \varphi_j + \tau_t\} \end{aligned}$$

where our industry measures are productivity growth dispersion, external finance dependence and the Lerner index. Our main interest is in the coefficient β_2. The most direct implication of the model is that young firms should publish disproportionately more in heterogeneous industries where it is hard for outsiders to evaluate firm quality. Thus we expect β_2 to be negative when age is interacted with productivity growth dispersion.

Finally, we exploit the cross-country variation in our sample by looking at financial development. To measure financial development, we refer to the world-bank indices developed by Beck, Demirgüç-Kunt and Levine (2000, 2007). For each country we examine two measures: stock market and banking system. For the stock market development, we use the ratio of stock market total value to GDP and for the banking system, we use the ratio of private credit by deposit money banks and other financial institutions to GDP, where bank deposits are the demand, time, and saving deposits in deposit money banks.[20]

companies is likely to be a pure measure of their demand for external finance. (ii) Disclosure requirements imply that data on external financing are comprehensive. (iii) While using US industry data is rather exogenous to European firms, it is likely that an industry's dependence on external funds in the US is a good measure of its dependence in European countries.

[20] Table A6 presents measures of financial development for the countries in our sample. Note that although these countries are all members of the European Union and have a relatively high per capita income, they differ significantly in the development of their stock market and banking system. For

Our specification becomes

$$E(P_{it}|X_{it}) = \exp\{\beta_1 Age_{it} + \beta_2 Age_{it} \times FinDev_c + \beta_2 \ln Sales_{it-1} + Z'_{it}\beta_4 + \varphi_j + \mu_c + \tau_t\}.$$

The model predicts that young firms should have stronger incentives to publish in countries with more developed financial markets. Thus, we expect $\beta_2 < 0$.

6. Results

6.1. The Effect of Firm Age

The most basic prediction of our model is that young firms should, ceteris paribus, publish more than mature firms. And if aggressive patenting is also used as a signaling device, the same should be true for patents. Table 3 summarizes the results for the relationship between firm age, publications and patenting. Columns 1-6 examine the effect of firm age, measured by years from date of incorporation, on the firm-year number of publications, controlling for firm size (where size is proxied by lagged sales). In columns 1-3 we include the log of firm age, which is significantly negative. This means that as age increases, publications intensity drops. The result is robust to controlling for different proxies of firm size (sales in millions versus log of sales) as well as for nonlinearities. In column 4 firm age is broken down to quartiles, where the first quartile (which includes the youngest firms) is the baseline. The results support the negative relationship between age and publication intensity, where firms in the lower age quartile are more publication intensive. For example, the coefficient on the third quartile of firm age is -0.644 (with a standard error of 0.120), while the coefficient on the forth is -0.911 (with a standard error of 0.117). Columns 5 and 6 examine different sub-samples of firms and give similar results.

In columns 7-10 we conduct the same analysis but now the dependent variable is the firm-year number of patents. The pattern of results also suggests a negative relationship

instance, the ratio of stock market value to GDP is 1.9 in Great Britain and 0.23 in Greece. This is important as it allows for identification.

between firm age and patenting intensity. In column 11-13 we examine the relationship between the likelihood of publishing and firm's age. The dependent variable is an indicator that receives the value of one if a firm published at least one article and zero otherwise. The sample is cross-sectional for the last year a firm appears in our sample. The estimation results support the pattern in figure 3, where the share of publishing firms in the highest in the lowest quartile of firm age.

[TABLE 3 ABOUT HERE]

Table 4 examines the robustness of the negative relation between firm age and publications, as found in table 3, when publication quality is taken into account. We use two measures to proxy for the quality of a publication. In columns 1-4 the dependent variable is the firm-year number of citations received. Thus, instead of counting the number of publications a firm has in a given year, we aggregate the number of citations from other publications to the firm-year level. The pattern of results remains the same, where for a given firm size, young firms, on average; receive more citations than more mature firms. In columns 5-10 we repeat the analysis separately for publications in low and high impact journals, which yields similar pattern of results.

[TABLE 4 ABOUT HERE]

6.2. Industry Characteristics

Table 5 summarizes the results of Negative Binomial regressions examining the relationship between publication and patent intensities and industry characteristics. Ultimately we are interested in the interactions between these industry characteristics and firm age, as they provide a much cleaner test for the signaling hypothesis. Nevertheless, finding a strong propensity to publish in more heterogeneous industries would clearly lend some support to the model.

In columns 1-2 we examine the relationship between productivity growth dispersion (measuring the industry spread in firm labor productivity growth) and publication intensity. Consistent with our expectations, we find a positive and highly significant coefficient on the dispersion variables. In columns 3-4 we examine the relationship between the external finance dependence and publication intensity. The relationship is positive and highly significant, suggesting stronger publication intensity in industries with higher demand for external funds. In column 5 we add the Lerner index of competition (higher values mean less competition) and the industry R&D intensity. Their coefficients are insignificant but the pattern of results for dispersion and external dependence remains the same. In columns 6-10, the dependent variable is the firm-year number of patents. Interestingly, the relation between dispersion and patents is no longer significant, implying there is no evidence suggesting stronger patenting intensity in more heterogeneous industries. This result is consistent with the argument that publications are a better proxy for signaling than patents, either because patents are not well-suited as a signaling device, or (more likely) because the primary function of patents is different.

Columns 11-15 examine the relationship between industry characteristics and the likelihood of publishing. Similar to columns 1-5, we find a higher likelihood of publishing in more heterogeneous industries and in industries that rely more on external finance.

[TABLE 5 ABOUT HERE]

6.3. Firm Age and Industry Dispersion

Our results so far suggest that firms with a shorter track record and firms that operate in industries where there is greater dispersion are more publication intensive. To strengthen the link between these results and our signaling hypothesis, Table 6 examines whether the interaction between industry dispersion and firm age helps explain publication intensity. According to the theory, young firm should publish disproportionately more in industries where performance varies substantially across firms. Thus, if the signaling hypothesis is

correct, we should expect a negative coefficient on the interaction of firm age with industry dispersion. In column 1 we include the interaction between firm age and dispersion. The coefficient on this interaction is negative and significant, supporting the signaling theory. In columns 2-9, we separately estimate the marginal effect of firm age on publication intensity across quartiles of industry dispersion. In the first quartile of dispersion (columns 1-2) the marginal effect of firm age is -0.028 (with a standard error of 0.012). Yet, in the forth quartile (where the heterogeneity in firm performance is the highest), the marginal effect of firm age is -0.192 (with a standard error of 0.032).

Columns 10-18 carry out the same exercise in the patenting equation. We find negative interactions here as well; however, they are smaller in absolute value compared to interactions from the publications equation. For example, the marginal effect of firm age in the forth quartile of industry dispersion in the patenting equation is -0.043 (with a standard error of -0.013), where it is -0.192 in the publications equation. This is consistent with the idea that publications might be more frequently used as a certification device than patents.

Finally, Table 7 summarizes the estimation results for our three main industry measures. Columns 1-9 report the results when the dependent variable is the number of publications, while columns 10-18 report equivalent specifications when the dependent variable is a dummy for publishing. Consistent with the signaling hypothesis, we find that the interaction between firm age and industry dispersion is also highly significant in the probit equation. By contrast, the coefficients on the interaction between firm age and external finance dependence and between firm age and the Lerner index are always insignificant. We also attempted to control for publication quality by splitting the sample between articles that were published in high quality journals and those that were published in low quality journals. However, the same broad pattern of results emerges.

[TABLES 6-7 ABOUT HERE]

6.4. Firm Age and Financial Development

The signaling hypothesis suggests that young firms operating in more developed financial markets should have stronger incentives to publish. Intuitively, financial development should be correlated with the variety and efficiency of financial intermediaries (Rajan and Zingales (1998)). Thus, we might expect publications to be more thoroughly evaluated in countries where financial development is higher.

Table 8 summarizes the results for the interaction between country financial development and firm age. Our measures of financial development are the size of the stock market and the level of private credit over GDP. Consistent with our expectations, we find a negative and significant interaction between financial development and firm age in the publications equations (columns 1-12), but no significant relation in the patents equation (columns 13-14). Thus in countries with high financial development young firms are more publication intensive (relative to mature firms) than in countries with low financial development, but the same is not true for patents.

In unreported regressions, we also experimented with another measure of financial development often used in empirical work: accounting standards.[21] Since accounting standards measure the transparency of firms' annual reports across countries, we did not expect them to be a particularly good proxy for the competence of the financial sector. And indeed the coefficient on the interaction between age and accounting standards turns out to be insignificant (-0.049 with a standard error of 0.049) in a regression analogous to the ones in columns 1 and 4.

Finally, an important caveat is that a large fraction of the publications of our sample come from only three countries: the UK, Germany and France (see Table A4). Thus, although we find these results suggestive, more evidence from a larger group of countries is probably needed before drawing strong conclusions.

[21] This information is from Rajan and Zingales (1998). See also Table A6 in the Appendix.

[TABLE 8 ABOUT HERE]

6.5. Evidence from M&A Deals

This section tests the signaling hypothesis by examining a sub-sample of firms that participated in M&A deals. Since these firms sold some of their equity directly or via IPO, this sample includes only firms that actually raised external finance (by selling equity stakes). The signaling hypothesis suggests that firms with a shorter track record should have strong incentives to publish prior the deal to certify their quality. To test this conjecture we consider the age of a firm in the year the deal was announced (the rumor year). We then test whether younger firms (whose age is calculated at the announcement year) are more likely to have published. We use Zephyr for M&A information during the period 1997-2004 and match this information to our sample of innovating firms. 1,109 innovating firms in the sample sold part of their equity during the estimation period. Out of this number, 482 firms published at least one article prior to the year the deal was announced.

Table 9 summarizes the estimation results. The dependent variable is a dummy that receives the value of 1 for firms that published prior the announcement year and 0 for firms that did not publish. Columns 1-2 examine the marginal effect of firm age on the likelihood of publishing. The results support the signaling hypothesis since firms that have a shorter track record at the year of the deal are more likely to have published at a prior date. In columns 3-8, we also examine the interaction between firm age and industry characteristics. The signaling hypothesis implies stronger signaling by young firms in industries with higher performance heterogeneity. We indeed find a highly significant negative interaction between firm age and industry heterogeneity (column 3). By contrast, the interaction between firm age and the Lerner competition index is insignificant (column 6). Also consistent with the signaling hypothesis, we find that young firms tend to publish disproportionately more in countries where financial markets are highly developed (columns 9-14). All these results appear to be robust to controlling for publication quality.

[TABLE 9 ABOUT HERE]

7. Conclusions

What are the incentives for private firms to engage in open science? This paper argues that firms may encourage their researchers to publish in academic journals in order to signal the quality of their research to prospective clients and investors. Using a novel and comprehensive database on academic publications, we demonstrate that European firms contribute substantially to the advancement of basic scientific knowledge. Consistent with our signaling view of open science, we also find that, controlling for size, young firms publish more than mature firms, and disproportionately so in more heterogeneous industries and in countries where financial markets are highly developed.

These results could be extended in several directions. An interesting finding is that firms that publish are on average larger than those that patent. In ongoing work (Belenzon et al. (2008)) we take a closer look at the relationship between firm size and the amount and composition of innovative activity (patents and academic publications). We find that the correlation between publications and measures of firm performance is stronger for large firms than for small firms, suggesting that basic research may benefit disproportionately more larger firms. Interestingly, the opposite pattern emerges when we look at the correlation between patents and performance.

Another direction for future research would be to consider alternatives to the signaling hypothesis. Absorptive capacity, defensive publications, financial constraints are all frequently mentioned reasons why firms may or may not engage in open science, but their relative importance has yet to be assessed. Clearly, much work remains to be done, both theoretically and empirically. We hope that this paper may spur further research (and academic publications) in this exciting area.

References

Acemoglu, D., P. Aghion, C. Lelarge, J. Van Reenen and F. Zilibotti, (2007), 'Technology, Information, and the Decentralization of the Firm', *Quarterly Journal of Economics*, Vol. 122, pp. 1759-1799.

Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt, (2005) 'Competition and Innovation: An Inverted U Relationship', *Quarterly Journal of Economics*, Vol. 120, pp. 701-728.

Akerlof, G.A., (1970), 'The Market for "Lemons": Quality Uncertainty and the Market Mechanism', *Quarterly Journal of Economics*, Vol. 84, pp. 488-500.

Beck, T., A. Demirgüç-Kunt and R. Levine, (2000), 'A New Database on Financial Development and Structure', *World Bank Economic Review*, Vol. 14, pp. 597-605 (updated in 2007).

Belenzon, S. and T. Berkovitz, (2007), 'Innovation in Business Groups', Working Paper.

Belenzon, S. and T. Berkovitz, (2008), 'Business Group Affiliation, Financial Development and Market Structure: Evidence from Europe', Working Paper.

Belenzon, S., A. Patacconi and J. Van Reenen, (2008), 'Firm Size and Innovation: Evidence from European Panel Data', Working Paper.

Bloom, N., Schankerman, M. and Van Reenen, J. (2005), "Identifying technology spillovers and product market rivalry", *CEPR Discussion Papers 4912.*

Carter, R. and S. Manaster, (1990), 'Initial Public Offerings and Underwriter Reputation', *Journal of Finance*, Vol. 45, pp. 1045-1067.

Cockburn, I.M. and R. Henderson, (1998), 'Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery', *Journal of Industrial Economics*, Vol. 46, pp.157-182.

Cohen, W.M. and D.A. Levinthal, (1990), 'Innovation and Learning: The Two Faces of R&D', *Economic Journal*, Vol. 99, pp. 569-596.

Cohen, W.M., R.R. Nelson and J.P. Walsh, (2000), 'Protecting their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)', NBER Working Paper No. 7552.

Cressy, R., F. Munari and A. Malipiero, (2007), 'Playing to Their Strengths? Evidence that Specialization in the Private Equity Industry Confers Competitive Advantage', forthcoming *Journal of Corporate Finance*.

Gompers, P., (1996), 'Grandstanding in the Venture Capital Industry', *Journal of Financial Economics*, Vol. 42, pp. 133-156.

Hall, B.H., R.H. Ziedonis, (2001), 'The Patent Paradox Revisited: An Empirical Study of Patenting in the U.S. Semiconductor Industry, 1979-1995', *Rand Journal of Economics*, Vol. 32, pp. 101-128.

Hausman, J., B.H. Hall and Z. Griliches, (1984), 'Econometric Models of Count Data with an Application to the Patents-R&D Relationship' *Econometrica*, Vol. 52, pp. 909-938.

Hirtle, B., (2006), 'Stock Market Reaction to Financial Statement Certification by Bank Holding Company CEOs', *Journal of Money, Credit and Banking*, Vol. 38, pp. 1263-91.

La Porta, R., F. Lopez-de-Silanes, A. Shleifer and R.W. Vishny, (1997), 'Legal Determinants of External Finance', *Journal of Finance*, Vol. 52, pp. 1131-1150.

Lerner, J. (1994), 'The Importance of Patent Scope: an Empirical Analysis', *Rand Journal of Economics*, Vol. 25, pp. 319-333.

Lichtenberg, F.R., (1986), 'Private Investment in R&D to Signal Ability to Perform Government Contracts', NBER Working Paper No. 1974.

Lichtenberg, F.R., (1988), 'The Private R&D Investment Response to Federal Design and

Technical Competitions', *American Economic Review*, Vol. 78, pp. 550-559.

Mann, R.J. and T.W. Sager, (2007), 'Patents, Venture Capital, and Software Start-Ups' *Research Policy*, Vol. 36, pp. 193-208.

Megginson, W.L. and K.A. Weiss, (1991), 'Venture Capitalist Certification in Initial Public Offerings', *Journal of Finance*, Vol. 46, pp. 879-903.

Munari, F. and L. Toschi (2008), 'The relationship between patent portfolios and VC investments: the case of nanotechnology', working paper.

Murray, F., (2002), 'Innovation as Co-Evolution of Scientific and Technological Networks: exploring Tissue Engineering', *Research Policy*, Vol. 31, pp. 1389-1403.

Murray, F. and S. Stern, (2006), 'Do formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? An Empirical Test of the anti-Commons Hypothesis', NBER Working Paper 11465. Forthcoming, *Journal of Economic Behavior and Organization*.

Rajan, R. and L. Zingales, (1998), 'Financial Dependence and Growth', *American Economic Review*, Vol. 88, pp. 559-586.

Soulsby, R.L. and J.S. Damgaard, (2005), 'Bedload sediment transport in coastal waters', *Coastal Engineering*, Vol. 52, pp. 673-689.

Wooldridge, J.M., (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, Ma, MIT Press.

A. Appendix

A.1. Matching academic publications

The largest database on academic publications is the ISI Web of Knowledge (WoK) by Thomson. This includes millions of records on publications in nearly 9,000 leading academic journals. The data is divided to three main categories based on the publication type: hard sciences, social sciences, and arts and humanities. Because we are interested in capturing investment in scientific research, we focus only on the hard sciences section of WoK. This section includes about 20 million publication records over the period 1970-2004. The address field on each record indicates the affiliation of the authors of the publication. For example, the following is a record in our database. This affiliation is typically either a research institution or a firm. We use the name appearing in this field and match it to the complete list of Amadeus firms. "HIGH-CAPACITY DIGITAL RADIO WITH TRELLIS CODING", BACCETTI B , TAVERNA M , BELLINI S , SALVINI G, EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS, NOV-DEC 1993. Address: BACCETTI B (reprint author), SIEMENS TELECOMUN SPA, I-20060 CASSINA DE PECCHI, ITALY. The record would be matched to SIEMENS TELECOMUN SPA, which is a firm in Amadeus. We follow the same matching procedure as described above for the EPO and USPTO patent matching. Articles may have more than one author (the median number of authors per article is 2). In this case, the address field would include multiple affiliations. We assign an academic publication to a specific firm if the name of this firm appears at least once in the address field of the article. This procedure means that a single article can be assigned to more than one firm, but a firm cannot be assigned more than once to the same article. For each article, we also extract information on the number of times it was cited, the journal in which it was published, and the year of publication. Information about the importance of journals is taken from the Journal Citations Report index (JCR). The Web Of Science often uses abbreviations. For example, “Chemicals”, “Chemische" (chemical in German) and Chemistry appear as “Chem”. Such standardization is important for name matching, because the name of the same company can appear differently in Amadeus and on the address field of the article (the country origin of each author is also listed for each publication, which ease the translation to English).

Finally, European research institutions can be incorporated, thus, they appear in Amadeus as potential firms to be matched. To screen out such firms, we follow two steps. First, as for patent matching, we drop Amadeus names that include strings that are associated with research institutions (such as, UNIVERSITY, RESEARCH, INSTITUTION, etc.) or government organizations (endings such as, NCR for Italy, CEA for France, etc.). Second, we manually examine the websites of firms that have a large number of publications but appear as small firms in terms of their sales and number of patents. For these firms, we check whether their primary activity is research. In case the primary activity is research, we exclude them from our matched sample. At the end of this procedure we are left with 234,864 publications that are matched to 21,052 Amadeus firms. Because our aim is to examine the effect of publications on firms performance, we match to the publishing firms accounting information. Firms that never report accounting information are dropped from our sample. After dropping firms with no financial information, we are left with 163,833 firm publications between 1970 and 2004. Over the estimation period,

1995-2004, our sample of firms publish 87,671 articles. Figure B7 plots the total number of firm publications over time. Starting at 1990 there has been a sharp increase in the number of firm publications, especially Biology and Chemistry, Health and Medicine and Engineering. A similar pattern holds when we include only firm publications in leading journal (journals with above median impact factor).

A.2. Matching patent data

A.2.1. European Patent Office (EPO)

The matching between EPO patent applicants and Amadeus firms has been a collaborative project with the Institute for Fiscal Studies (IFS) and the Centre for Economic Performance (CEP).[22] This section is a brief summary of the matching procedure described in the CEP/IFS AmaPat document and is included here for completeness. See also Belenzon and Berkovitz (2007).

Our main information source on patents is the April 2004 publication of the PATSTAT database, which is the standard source for European patent data. This database contains all bibliographic data (including citations) on all European patent applications and granted patents, from the beginning of the EPO system in 1979 to 2004.

We match the name of each EPO applicant listed on the patent document to the full name of a firm listed in Amadeus (about 8 million names). Since we are interested only in matching patent applicants to firms, we exclude applicant names that fall into the following categories: government agencies, universities, and individuals. We identify government agencies and universities by searching for a set of identifying strings in their name. We identify individuals as patents where the assignee and the inventor name strings are identical.

The matching procedure follows two main steps. (i) Standardizing names of patent applicants. This involves replacing commonly used strings which symbolize the same thing, for example "Ltd." and "Limited" in the UK.[23] We remove spaces between characters and transform all letters to capital letters. As an example, the name "British Nuclear Fuels Public Limited Company" becomes "BRITISHNUCLEARFUELSPLC". (ii) Name matching: match the standard names of the patent applicants with Amadeus firms. If there is no match, then try to match to the old firm name available in Amadeus. We need to confront a number of issues. First, in any given year, the Amadeus database excludes the names of firms that have not filed financial reports for four consecutive years (e.g. M&A, default). We deal with this issue in several ways. First, we use information from historical versions of the Amadeus database (1995-2003) on names and name changes. Second, even though Amadeus contains a unique firm identifier (BVD ID number), there are cases in which firms with identical names have different BVD numbers. In these cases, we use other variables for identification, for example: address (ZIP code), Date of

[22] We extend our gratitude to the tremendous work done by Rachel Griffith and the IFS team, especially Gareth Macartney in developing and implementing the patent matching. More information about the matching is available at: "AmaPat: Accounting, Ownership and Patents for European Firms" (CEP/IFS AmaPat document).

[23] The complete list of strings is available in the CEP/IFS AmaPat document.

incorporation (whether consistent with the patent application date), and more. Finally, we manually match most of the remaining corporate patents to the list of Amadeus firms.

A.2.2. United States Patents and Trademarks Office (USPTO)

The procedure described above matches European firms to patents registered with the EPO. Yet, some European firms register patents only with the USPTO, without applying to the EPO. In order to identify the European firms that only apply to the USPTO, we match the complete set of Amadeus firms to the name of the patent applicants from the USPTO. The most updated patent database for the USPTO is the 2002 version of the NBER patents and citations data archive.[24] Because this database covers patent information only up to 2002 and our accounting data go up to 2004, we updated the patent data file by extracting all information about patents granted between 2002 and 2004 directly from the USPTO website.[25] Having updated the USPTO patent database, we follow the matching procedure described above to create the matched USPTO patent data for the Amadeus firms.

Firms can apply for patents for the same invention with both the EPO and the USPTO. Patents protecting the same invention across different organizations are called a patent family. To avoid double counting of inventions, information on patent families is needed. We collect this information from the OECD Triadic database on patent families.[26] Having identified inventions that belong to the same family, we exclude patents granted by the USPTO that belong to the same family of patents granted by the EPO.

A.3. Accounting database

The accounting information is taken from Amadeus. The database contains financial information on about 8 million firms from 34 countries, including all the European Union countries and Eastern Europe. The accounts of each firm are followed for up to ten years. The information source for Amadeus is about 50 country vendors (generally the office of register of Companies). The main advantage of Amadeus over other data sources is its coverage of small and medium size firms.

The accounting database includes items from the balance sheet (22 items) and income statement (22 items). No information is available from the changes in cash flow report (i.e., investment data is not available). The accounting data is harmonized by BvD to enhance comparison across countries. This comparison becomes easier over time due to the improvement in the European Union harmonization is accounting standards. In addition to accounting data items, Amadeus provides a description of firms including their product market activity. The main descriptive items are legal form (public versus private), date of incorporation, types of accounts (consolidated versus unconsolidated), country, US SIC and NAIC for the product market activity of the firm (primary and non-primary). The industry location information includes up to eight different six-digit NAIC codes per firm (note that the sales of the firm are not broken-up across the different product markets).

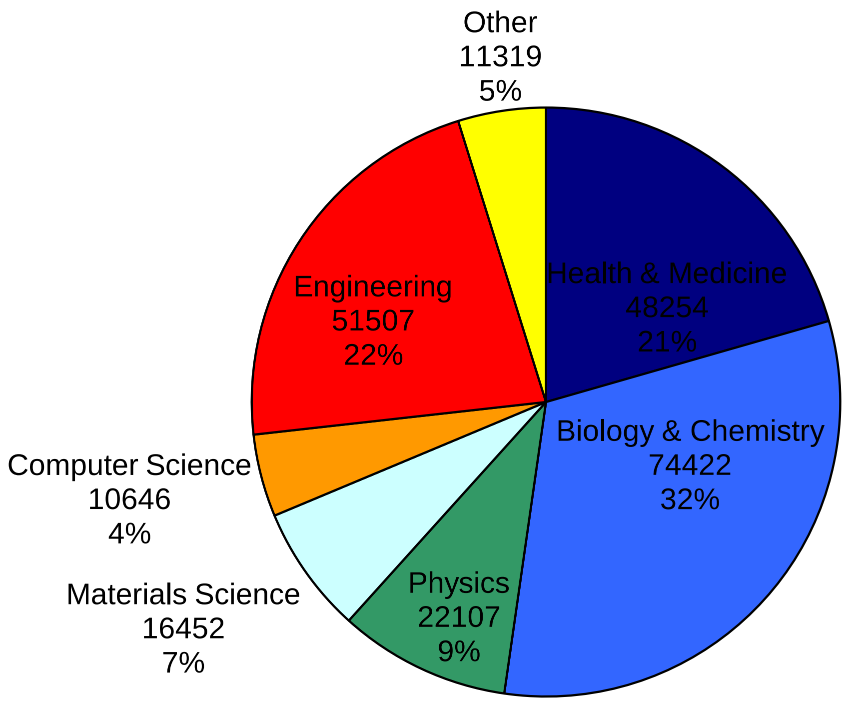
[24] http://elsa.berkeley.edu/~bhhall/bhdata.html

[25] http://patft.uspto.gov/netahtml/PTO/srchnum.htm

[26] This includes patents that are registered in all three main patents offices: the EPO, JPO, and USPTO.

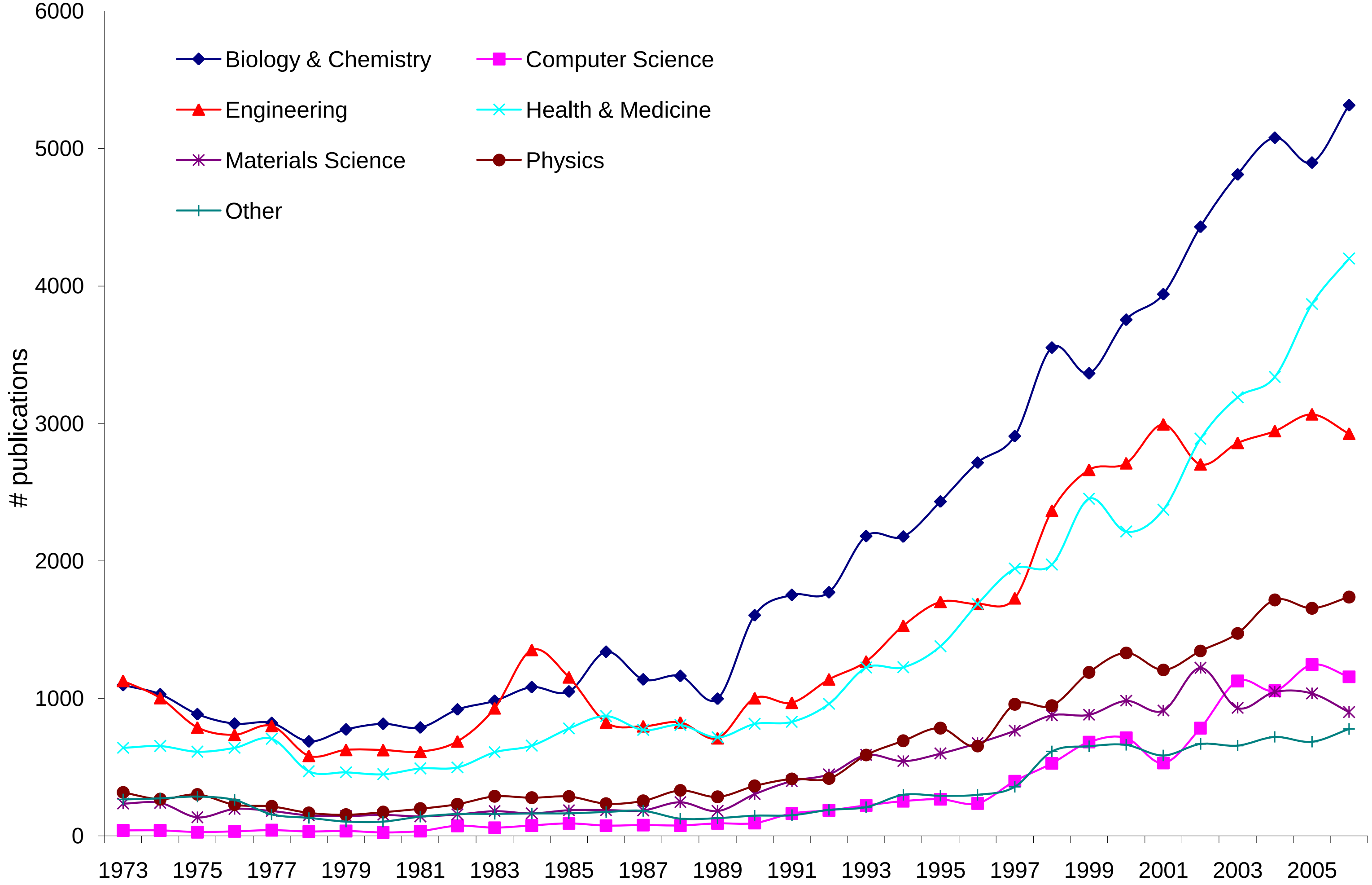
An important feature of the data is the criteria for dropping firms from the sample over time. As long as a firm continues to file its financial statements, it continues to appear in Amadeus. In case a firm becomes inactive, it stops filing its financial statement (alternatively, a firm can be late in filing its financial statement). This firm will be kept in the sample for four extra years since the last year financial statements were reported (thus, in the fifth year the firm will be removed from the sample). For example, a firm that becomes inactive and stops filing its reports in 1995 (i.e., 1994 is the last year when a financial statement was reported) will remain in the database until 1998 (including) and in 1999, it will be dropped from the sample (all observations of the specific firm will be taken out from the Amadeus database in the 1999 update). In order to mitigate the problem of losing dead firms, we purchased old Amadeus disks that allow tracking firms that exit the sample in previous years. For example, the firm that exits in 1995 will appear in the 1998 Amadeus disk, but not in the 1999 disk. By using both 1998 and 1999 disks, we mitigate the selection bias of dropping inactive firms after 4 years of missing data.

FIGURE 1: FIRM PUBLICATIONS ACROSS MAIN TECHNOLOGY AREAS (234,707 ARTICLES BETWEEN 1970 AND 2004)



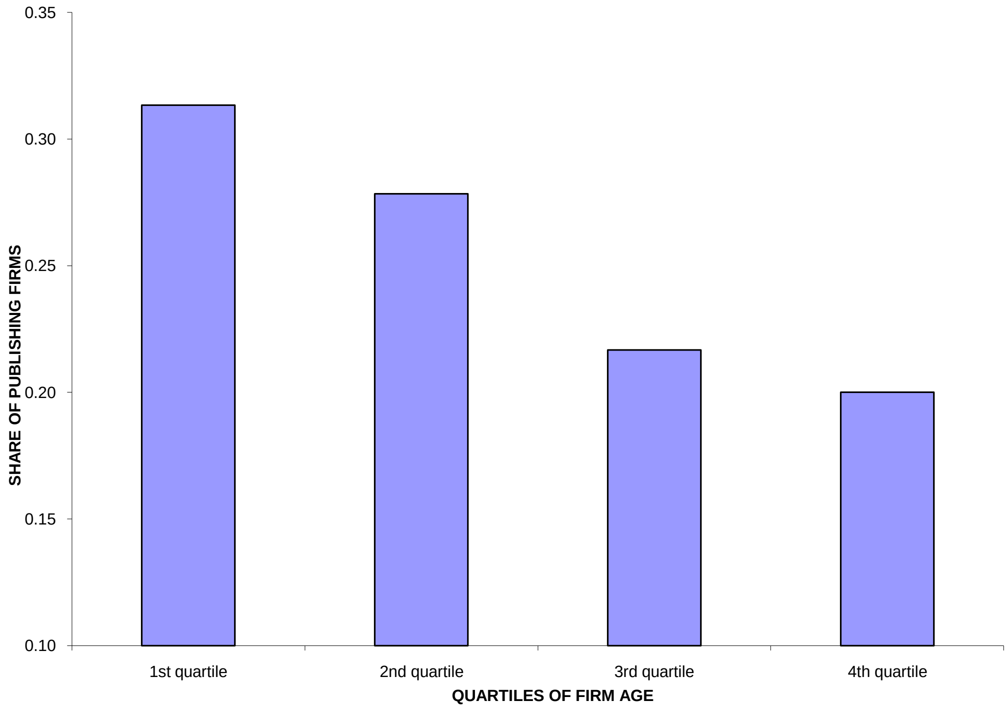
Notes: This figure plots the distribution of firm publications across main technology areas. We include all publication that were matched to about 8m European firms from Amadeus over the period 1970-2004. The academic publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database.

FIGURE 2: FIRM PUBLICATIONS OVER TIME AND ACROSS MAIN TECHNOLOGY AREAS



Notes: This figure plots the distribution of firm publications over time and across main technology areas. We include all publication that were matched to about 8m European firms from Amadeus over the period 1970-2004. The academic publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database.

FIGURE 3: SHARE OF PUBLISHING FIRMS ACROSS FIRM AGE QUARTILES



Notes: This figure plots the share of publishing firms across firm age quartiles. A firm is assumed not to publish if there is no single match between the name of the firms and the ISI Web of Science in the period 1970-2006. Firm age is years from date of incorporation, computed at the last year where financials are reported. The sample includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2004. Firms in our sample are required to have more than 10 employees and $1 in annual sales in the last year they report accounts.

FIGURE 4A: PUBLICATIONS INTENSITY AND FIRM AGE

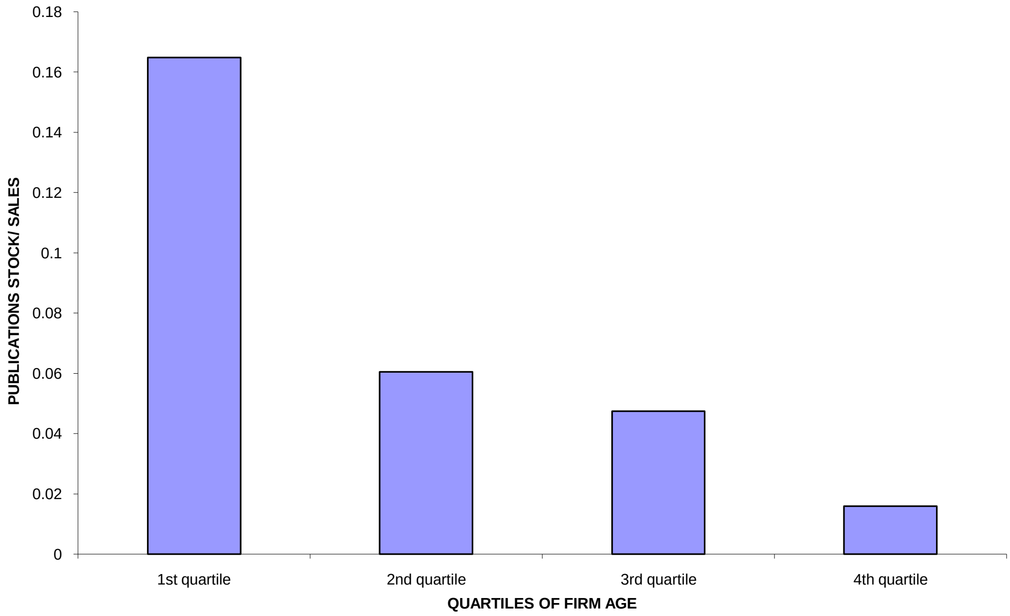
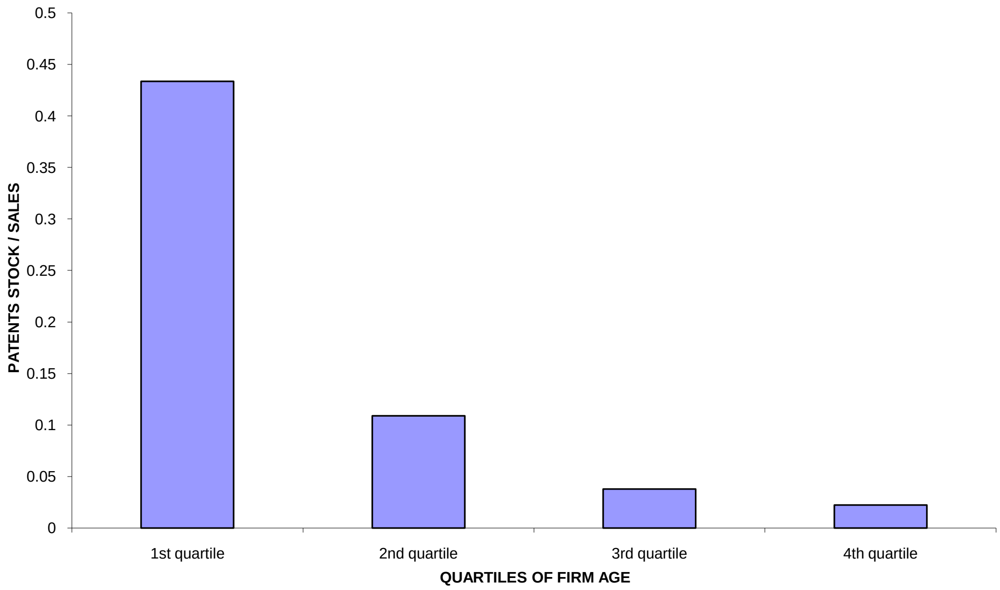


FIGURE 4B: PATENTS INTENSITY AND FIRM AGE



Notes: Figure 4A plots publications intensity (defined as the ratio between a firm's publications stock and its sales) across quartiles of firm age. Figure 4B provides the same information for patents intensity. Firm age is years from date of incorporation, computed at the last year where financials are reported. Sales are in millions of US dollars. The sample includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2004. Firms in our sample are required to have more than 10 employees and $1 in annual sales in the last year they

TABLE 1-

SUMMARY STATISTICS FOR MAIN FIRM CHARACTERISTICS (FIRM-YEAR)

					Distribution		
Variable	# firms	# Obs	Mean	Std. Dev.	10st	50th	90th
# of publications	2,835	12,953	1.1	5.6	0	0	2
# of patents	7,374	33,712	0.8	5.8	0	0	1
Publications stock	2,835	12,953	4.3	18.7	0	0.9	8.0
Patents stock	7,374	33,712	7.1	42.0	0.1	0.9	12.3
Sales (`000)	9,239	41,754	403,913	3,660,317	4,042	29,025	393,230
Employess	9,239	41,754	1,430	10,363	26	165	1,658
Age	9,239	41,754	29	23	6	22	64
Capital (`000)	8,238	37,993	312,862	3,933,511	431	6,632	162,871
Cash flow (`000)	8,021	36,591	48,085	543,923	-727	1,685	34,588

Notes: This table provides summary statistics for firms in our estimation sample over the period 1995-2004. The sample includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2004. Academic publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database. Patents are constructed by matching the name of all Amadeus firms to all EPO patent records for the period 1978-2004. Cash is defined as net income plus depreciation. Capital is defined as fixed-assets. Age is the number of years since the date of incorporation. Patents and publication stocks are computed using the perpetual inventory method using a depreciation rate of 15 percent.

TABLE 2-

COMPARISON OF KEY VARIABLES: PATENTING FIRMS VERSUS PUBLISHING FIRMS

PANEL A: PATENTING

Variable	# firms	# Obs	Mean	Std. Dev.	Distribution		
					10st	50th	90th
# of patents	7,374	33,712	0.8	5.8	0	0	1
Patents stock	7,374	33,712	7.1	42.0	0.1	0.9	12.3
Sales (`000)	7,374	33,712	389,418	3,290,521	4,269	29,611	390,279
Employess	7,374	33,712	1,481	10,871	28	175	1,746
Age	7,374	33,712	29	23	6	23	65
Capital (`000)	6,405	30,067	313,297	4,106,974	554	7,335	162,320
Cash flow (`000)	6,263	29,060	45,623	512,340	-780	1,762	33,856

PANEL B: PUBLISHING

Variable	# firms	# Obs	Mean	Std. Dev.	Distribution		
					10st	50th	90th
# of publications	2,835	12,953	1.1	5.6	0	0	2
Publications Stock	2,835	12,953	4.3	18.7	0	0.9	7.9
Sales (`000)	2,835	12,953	891,975	6,251,996	4,046	43,661	908,252
Employess	2,835	12,953	2,865	17,359	26	208	3,008
Age	2,835	12,953	28	23	5	19	65
Capital (`000)	2,698	12,471	612,573	4,669,980	305	10,462	442,950
Cash flow (`000)	2,593	11,827	103,710	810,476	-891	2,535	83,331

Notes: This table provides summary statistics for firms in our estimation sample over the period 1995-2004. The sample includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2004. Panel A is for firms that have at least one patent between 1978 and 2004 and panel B is for firms that have at least one publication over the same period. Patents are constructed by matching the name of all Amadeus firms to all EPO records for the period 1978-2004. Academic publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database. Cash is defined as net income plus depreciation. Capital is defined as fixed-assets. Age is the number of years since the date of incorporation. Patents and publication stocks are computed using the perpetual inventory method using a depreciation rate of 15 percent.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Dependent variable:													
Firms:	All	All	All	All	Age>1st quartile	Only publishing	All	All	Age>1st quartile	Only patenting	All	All	Age>1st quartile
log(Firm Age)	-03‰(/)6(y)10( )]TETBT1 0 0 1 410MCID 9‰ng (x-none)>BDC BT1 0 0 1 636.34 497.98 Tm[(11))]T263												

TABLE 4-

PUBLICATIONS AND FIRM AGE: CONTROLLING FOR PUBLICATION QUALITY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	# of citations received (Negative Binomial)				# of publications (Negative Binomial)					
Journal Impact Factor	All				Below median			Above median		
Firms:	All	All	Age>1st quartile	Only publishing	All	All	Age>1st quartile	All	All	Age>1st quartile
log(Firm Age)	-0.596***				-0.417***			-0.355***		
	(0.032)				(0.050)			(0.056)		
Firm Age quartiles:										
Dummy for 2nd quartile		-0.620***		-0.528***		-0.507***			-0.496***	
		(0.133)		(0.116)		(0.107)			(0.127)	
Dummy for 3rd quartile		-0.902***	-0.460***	-0.701***		-0.658***	-0.184		-0.562***	-0.162
		(0.149)	(0.153)	(0.148)		(0.128)	(0.120)		(0.157)	(0.154)
Dummy for 4th quartile		-1.350***	-0.887***	-1.089***		-0.974***	-0.480***		-0.771***	-0.375**
		(0.147)	(0.173)	(0.158)		(0.125)	(0.120)		(0.159)	(0.161)
$\log(\text{Sales})_{t-1}$	0.519***	0.516***	0.590***	0.362***	0.494***	0.494***	0.515***	0.449***	0.444***	0.493***
	(0.032)	(0.031)	(0.041)	(0.028)	(0.027)	(0.027)	(0.031)	(0.033)	(0.033)	(0.043)
Three-digit SIC dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	-22,199.3	-22,256.6	-13,959.9	-18,770.4	-10,762.0	-10,796.2	-6,902.4	-8,859	-8,888.9	-5,527.3
# Firms	9,239	9,239	6,667	2,816	8,432	8,432	6,316	7,988	7,988	5,966
# Observations	41,754	41,754	29,518	12,888	38,071	38,071	27,718	35,775	35,775	26,023

Notes: This table reports the results of Negative Binomial regressions that examine the relationship between publication quality and firm age, controlling for the quality of publications. The sample includes firms that have at least one patent or one academic publication in "hard" science journals between 1978 and 2004, have at least 10 employees and at least $1 million in annual sales. Age is calculated from the date of incorporation. Journal Impact Factor is a measure of the importance of a journal and is included in the ISI database. Standard errors (in brackets) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 5-

THE RELATIONSHIP BETWEEN PUBLICATIONS, PATENTS AND INDUSTRY CHARACTERISTICS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Dependent variable:	# of publications (Negative Binomial)					# of patents (Negative Binomial)					Dummy for publishing (Probit)				
Productivity Growth Dispersion (90-10)	1.931***			1.231***	1.335***	0.097			0.027	0.018	0.437***			0.439***	0.432***
	(0.337)			(0.282)	(0.287)	(0.150)			(0.152)	(0.154)	(0.058)			(0.053)	(0.052)
Productivity Growth Dispersion (95-5)		1.298***					0.092					0.299***			
		(0.161)					(0.088)					(0.037)			
External Finance Dependence			0.637***	0.605***	0.523***			0.248***	0.247***	0.265***			0.270***	0.273***	0.299***
			(0.064)	(0.064)	(0.101)			(0.037)	(0.037)	(0.061)			(0.017)	(0.017)	(0.022)
Lerner Competition Index					0.488					-2.017*					2.045***
					(1.335)					(1.129)					(0.558)
Industry mean of R&D / Sales					0.032					-0.003					-0.007
					(0.029)					(0.020)					(0.007)
log(Firm Age)	-0.418***	-0.417***	-0.433***	-0.389***	-0.390***	-0.149***	-0.147***	-0.129***	-0.128***	-0.128***	-0.142***	-0.138***	-0.142***	-0.117	-0.115***
	(0.067)	(0.066)	(0.072)	(0.076)	(0.075)	(0.035)	(0.035)	(0.034)	(0.034)	(0.034)	(0.017)	(0.018)	(0.017)	(0.018)	(0.018)
$\log(\text{Sales})_{t-1}$	0.453***	0.469***	0.396***	0.412***	0.410***	0.517***	0.517***	0.515***	0.513***	0.516***	0.115***	0.115***	0.117***	0.122***	0.124***
	(0.040)	(0.039)	(0.030)	(0.031)	(0.031)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Three-digit SIC dummies	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	-16,992.0	-16,934.7	-16,871.7	-16,639.6	-16,634.1	-23,687.5	-23,685.9	-23,617.7	-23,589.4	-23,584.8	-5,068.4	-5,051.9	-5,059.4	-4,927.3	-4,917.5
# Firms	9,239	9,239	9,239	9,239	9,239	9,239	9,239	9,239	9,239	9,239	9,239	9,239	9,239	9,239	9,239
# Observations	41,754	41,754	41,754	41,754	41,754	41,754	41,754	41,754	41,754	41,754	9,239	9,239	9,239	9,239	9,239

Notes: This table examines the relationship between publications, patents and industry characteristics. Productivity Growth Dispersion (90-10) is the difference in annual growth rate in productivity between the 90th and 10th deciles within an industry and is based on Acemoglu et al. (2007). External Finance Dependence is defined as the ratio between capital expenditures minus cash flow from operations and capital expenditures and is based on Rajan and Zingales (1998). The Lerner competition index is the industry average ratio between profits and sales. The sample includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2004 and have at least 10 employees and $1 million in annual sales. The sample in columns 11-15 is cross-sectional for the last year financials are available for each firm. Standard errors (in brackets) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 6-

FIRM AGE AND PRODUCTIVITY GROWTH DISPERSION (PGD). NEGATIVE BINOMIAL ESTIMATION

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Dependent variable:	# of publications (marginal effects)									# of patents (marginal effects)								
PGD quartile:	All	1st quartile		2nd quartile		3rd quartile		4th quartile		*All*	1st quartile		2nd quartile		3rd quartile		4th quartile	
PGD × log(Firm Age)	-0.233**									-0.035								
	(0.107)									(0.066)								
log(Firm Age)	-0.326***	-0.028**		-0.017***		-0.053***		-0.192***		-0.012	-0.016**		-0.013		-0.023**		-0.043***	
	(0.064)	(0.012)		(0.021)		(0.014)		(0.032)		(0.010)	(0.008)		(0.011)		(0.009)		(0.013)	
Firm Age quartiles:																		
Dummy for 2nd quartile			-0.040**		-0.111***		-0.064***		-0.197***			-0.030**		-0.030		-0.015		-0.060**
			(0.019)		(0.033)		(0.022)		(0.043)			(0.014)		(0.021)		(0.020)		(0.027)
Dummy for 3rd quartile			-0.041*		-0.119***		-0.093***		-0.270***			-0.028*		-0.032		-0.032		-0.071***
			(0.024)		(0.040)		(0.021)		(0.051)			(0.017)		(0.023)		(0.020)		(0.027)
Dummy for 4th quartile			-0.060***		-0.121***		-0.089***		-0.322***			-0.039**		-0.030		-0.046**		-0.091***
			(0.023)		(0.044)		(0.021)		(0.049)			(0.018)		(0.025)		(0.021)		(0.024)
$\log(\text{Sales})_{t-1}$	0.479***	0.035***	0.036***	0.134***	0.128***	0.057***	0.055***	0.078***	0.079***	0.083***	0.082***	0.082***	0.113***	0.114***	0.067	0.068***	0.069***	0.069***
	(0.026)	(0.006)	(0.006)	(0.022)	(0.020)	(0.006)	(0.006)	(0.014)	(0.013)	(0.004)	(0.007)	(0.007)	(0.008)	(0.008)	(0.005)	(0.005)	(0.009)	(0.009)
Three-digit SIC dummies	Yes	No	No	No	No	No	No	No	No	Yes	No	No	No	No	No	No	No	No
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	-15,840.8	-2,492.0	-2,493.9	-4,675.5	-4,693.9	-3,547.1	-3,558.9	-5,955.4	-5,960.5	-23,682.6	-5,874.4	-5,895.6	-6,510.4	-6,553.7	-5,629.2	-5,648.5	-5,454.7	-5,473.2
# Firms	9,239	2,208	2,208	2,137	2,140	2,143	2,143	2,350	2,350	9,239	2,208	2,208	2,137	2,140	2,143	2,143	2,350	2,350
# Observations	41,754	10,348	10,348	9,703	9,703	9,782	9,782	11,921	11,921	41,754	10,348	10,348	9,703	9,703	9,782	9,782	11,921	11,921

Notes: This table examines the relationship between publications, patents, firm age and industry heterogeneity. Firm age is years from date of incorporation. Productivity Growth Dispersion is the difference in annual growth rate in productivity between the 90th and 10th deciles within an industry and is based on Acemoglu et al. (2007). The sample includes firms that have at least one patent or one academic publication in "hard" science journals between 1978 and 2004, and have at least 10 employees and $1 million in annual sales. Standard errors (in brackets) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 7-

THE RELATIONSHIP BETWEEN PUBLICATIONS, FIRM AGE AND INDUSTRY CHARACTERISTICS: INTERACTION EFFECTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Dependent variable:	# of publications (Negative Binomial)									Dummy for publishing (Probit)								
Industry measure:	Productivity Growth Dispersion (90-10)			External Finance Dependence			Lerner index (Profits/Sales)			Productivity Growth Dispersion (90-10)			External Finance Dependence			Lerner index (Profits/Sales)		
Journal Impact factor:	All	Below median	Above median	All	Below median	Above median	All	Below median	Above median	All	Below median	Above median	All	Below median	Above median	All	Below median	Above median
Industry Measure × log(Firm Age)	-0.233**	-0.207*	-0.491***	0.058	-0.040	0.104**	1.074	2.426**	0.104**	-0.139**	-0.131**	-0.132**	-0.011	-0.027	0.023	0.283	0.964*	0.391
	(0.107)	(0.126)	(0.182)	(0.043)	(0.044)	(0.041)	(1.036)	(1.059)	(0.041)	(0.058)	(0.059)	(0.065)	(0.019)	(0.019)	(0.016)	(0.548)	(0.598)	(0.690)
log(Firm Age)	-0.326***	-0.344***	-0.171*	-0.496***	-0.472***	-0.429***	-0.480***	-0.563***	-0.429***	-0.010	0.029	-0.045	-0.051*	-0.058*	-0.146***	-0.083**	-0.081**	-0.121**
	(0.064)	(0.073)	(0.094)	(0.079)	(0.083)	(0.082)	(0.084)	(0.088)	(0.082)	(0.030)	(0.031)	(0.035)	(0.031)	(0.033)	(0.031)	(0.039)	(0.042)	(0.049)
log(Sales)$_{t-1}$	0.479***	0.496***	0.454***	0.475***	0.491***	0.236***	0.477***	0.492***	0.236***	0.123***	0.138***	0.130***	0.120***	0.134***	0.048***	0.119***	0.134***	0.128***
	(0.026)	(0.027)	(0.034)	(0.025)	(0.027)	(0.023)	(0.026)	(0.027)	(0.023)	(0.010)	(0.011)	(0.013)	(0.010)	(0.011)	(0.007)	(0.010)	(0.011)	(0.013)
Three-digit SIC dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	-15,840.8	-10,710.9	-8,769.8	-16,017.1	-10,786.8	-13,877.5	-16,012.7	-10,765.9	-13,877.5	-4,348.5	-3,856.6	-2,755.7	-4,387.9	-3,887.9	-4,364.9	-4,383.9	-3,866.1	-2,775.9
# Firms	9,239	8,496	8,204	9,239	8,496	8,204	9,239	8,496	8,204	9,239	8,496	8,204	9,239	8,496	8,204	9,240	8,496	8,380
# Observations	41,754	39,039	36,534	41,754	39,039	36,534	41,754	39,039	36,534	9,239	8,496	8,204	9,239	8,496	8,204	9,240	8,496	8,380

Notes: This table examines the relationship between publications and industry characteristics across firm age. Firm age is years from date of incorporation. Productivity Growth Dispersion (90-10) is the difference in annual growth rate in productivity between the 90th and 10th deciles within an industry and is based on Acemoglu et al. (2007). External Finance Dependence is defined as the ratio between capital expenditures minus cash flow from operations and capital expenditures and is based on Rajan and Zingales (1998). The Lerner competition index is the industry average ratio between profits and sales. Journal Impact Factor is a measure of the importance of a journal and is included in the ISI database. The sample includes firms that have at least one patent or one academic publication in "hard" science journals between 1978 and 2004, and have at least 10 employees and $1 million in annual sales. The sample in columns 10-18 is cross-sectional for the last year financials are available for each firm. Standard errors (in brackets) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 8-

THE RELATIONSHIP BETWEEN PUBLICATIONS, PATENTS, FIRM AGE AND FINANCIAL DEVELOPMENT: INTERACTION EFFECTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Dependent variable:	# of publications (Negative Binomial)						Dummy for publishing (Probit)						# of patents (Negative Binomial)	
Financial Development:	Stock Market Value Traded / GDP			Private Credit / GDP			Stock Market Value Traded / GDP			Private Credit / GDP			Stock Market Value Traded / GDP	Private Credit / GDP
Journal Impact factor:	All	Below median	Above median	All	Below median	Above median	All	Below median	Above median	All	Below median	Above median	All	All
Financial Development × log(Firm Age)	-0.195**	-0.247***	-0.153	-0.364***	-0.349**	-0.374**	-0.091**	-0.095**	-0.105**	-0.168***	-0.208***	-0.117	-0.081	-0.002
	(0.083)	(0.087)	(0.112)	(0.143)	(0.155)	(0.189)	(0.037)	(0.040)	(0.046)	(0.064)	(0.068)	(0.078)	(0.061)	(0.114)
log(Firm Age)	-0.204**	-0.159	-0.193	0.031	0.004	0.102	0.030	0.080*	0.010	0.140*	0.234***	0.043	-0.044	-0.124
	(0.098)	(0.105)	(0.123)	(0.175)	(0.191)	(0.229)	(0.044)	(0.047)	(0.053)	(0.080)	(0.086)	(0.096)	(0.063)	(0.138)
$\log(\text{Sales})_{t-1}$	0.479***	0.495***	0.451***	0.479***	0.492***	0.450***	0.120***	0.136***	0.129***	0.119***	0.135***	0.129***	0.547***	0.546***
	(0.026)	(0.027)	(0.033)	(0.026)	(0.027)	(0.033)	(0.010)	(0.011)	(0.013)	(0.010)	(0.011)	(0.013)	(0.019)	(0.019)
Three-digit SIC dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	-16,038.4	-10,782.3	-8,881.6	-16,037.9	-10,785.2	-8,879.8	-4,386.7	-3,881.5	-2,778.8	-4,386.3	-3,879.9	-2,780.4	-23,480.2	-23,483
# Firms	9,239	8,496	8,204	9,239	8,496	8,204	9,239	8,496	8,204	9,239	8,496	8,204	9,239	9,239
# Observations	41,754	39,039	36,534	41,754	39,039	36,534	9,239	8,496	36,534	9,239	8,496	8,204	41,754	41,754

Notes: This table examines the relationship between publications, firm age and a country's financial development. Firm age is years from date of incorporation. The measures of financial development are for 2005 and are based on Beck et al. (2000, 2007). Stock Market Value Traded is total shares traded on the stock market exchange. Private Credit is the ratio of private credit by deposit money banks and other financial institutions. Journal Impact Factor is a measure of the importance of a journal and is included in the ISI database. The sample includes firms that have at least one patent or one academic publication in "hard" science journals between 1978 and 2004, and have at least 10 employees and $1 million in annual sales. The sample in columns 9-14 is cross-sectional for the last year financials are available for each firm. Standard errors (in brackets) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE 9-

EVIDENCE FROM MERGERS AND ACQUISITIONS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Industry / Country measure:	-		Productivity Growth Dispersion (90-10)			Lerner index (Profits/Sales)			Stock Market Value Traded / GDP			Private Credit / GDP		
Journal Impact factor:	All		All	Below median	Above median	All	Below median	Above median	All	Below median	Above median	All	Below median	Above median
Industry/Country measure × log(Firm Age)			-0.598*** (0.216)	-0.571** (0.227)	-0.574** (0.230)	1.991 (1.294)	3.409** (1.686)	2.074 (1.404)	-0.372*** (0.092)	-0.503*** (0.097)	-0.275*** (0.106)	-0.611*** (0.152)	-0.646*** (0.169)	-0.390** (0.171)
log(Firm Age)	-0.061*** (0.016)		0.167* (0.095)	0.272*** (0.099)	0.104 (0.103)	-0.219** (0.097)	-0.189* (0.117)	-0.287*** (0.111)	0.256*** (0.099)	0.515*** (0.106)	0.094 (0.110)	0.622*** (0.190)	0.784*** (0.210)	0.230 (0.212)
Firm Age quartiles:														
Dummy for 2nd quartile		-0.016 (0.054)												
Dummy for 3rd quartile		-0.137*** (0.049)												
Dummy for 4th quartile		-0.180*** (0.047)												
$\log(\text{Sales})_{t-1}$	0.044*** (0.001)	0.045*** (0.007)	0.126*** (0.025)	0.165*** (0.028)	0.166*** (0.031)	0.124*** (0.025)	0.160*** (0.027)	0.165*** (0.031)	0.128*** (0.025)	0.167*** (0.027)	0.166*** (0.031)	0.133*** (0.025)	0.169*** (0.027)	0.169*** (0.031)
Three-digit SIC dummies	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Log pseudolikelihood	-722.2	-718.3	-541.9	-489.2	-387.9	-548.7	-494.2	-394.5	-542.2	-485.1	-392.4	-542.1	-489.4	-393.1
# Firms	1,109	1,109	1,123	1,072	1,011	1,123	1,072	1,011	1,123	1,072	1,011	1,123	1,072	1,011

Notes: This table examines the relationship between publications and firm age in a sample of Amadeus firms that were acquired in M&A deals during 1999-2004. Our firms have at least one patent or one academic publication in "hard" science journals between 1978 and 2004, and have at least 10 employees and $1 million in annual sales. The dependent variable is a dummy of whether the aquired firm has ever published. Productivity Growth Dispersion is the difference in annual growth rate in productivity between the 90th and 10th deciles within an industry and are based on Acemoglu et al. (2007). The Lerner competition index is the industry average ratio between profits and sales. The measures of financial development are for 2005 and are based on Beck et al. (2000, 2007). Stock Market Value Traded is total shares traded on the stock market exchange. Private Credit is the ratio of private credit by deposit money banks and other financial institutions. Journal Impact Factor is a measure of the importance of a journal and is included in the ISI database. Standard errors (in brackets) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by firms. * significant at 10%; ** significant at 5%; *** significant at 1%.

FIGURE A1: EXAMPLE OF PUBLICATION PAGE FROM THE ISI WEB OF SCIENCE (THOMSON)

Clinical study to assess the immunogenicity and safety of a recombinant Pseudomonas aeruginosa OprF-Oprl vaccine in burn patients

Full Text | Find it Oxford | Holdings | Go | Print | E-mail | Add to Marked List | Save to EndNote Web | more options

Author(s): Mansouri E, Blome-Eberwein S, Gabelsberger J, Germann G, von Specht BU

Source: FEMS IMMUNOLOGY AND MEDICAL MICROBIOLOGY **Volume:** 37 **Issue:** 2-3 **Pages:** 161-166 **Published:** JUL 15 2003

Times Cited: 6 **References:** 28

Abstract: In a recent clinical trial we evaluated the safety and immunogenicity of a recombinant OprF-Oprl vaccine consisting of the mature outer membrane protein I (Oprl) and amino acids 190-342 of OprF of Pseudomonas aeruginosa in burn patients and compared the elicited antibodies with antibodies against tetanus as response to a simultaneous immunization given on the day of admission. Safety and immunogenicity of the vaccine had been tested before in healthy human volunteers as published in 1999. In this first clinical trial we immunized eight burn patients suffering from second or third degree burns involving between 35% and 55% of the body surface three times with 100 mug of the OprF-Oprl vaccine. The vaccine was found to be very well tolerated. The patients did not show any serious side effects - and in particular no activation of the mediator cascade was observed. None of the subjects showed systemic P. aeruginosa infections during or after the treatment of their burns. The serological tests (ELISA) for detection of antibodies against P. aeruginosa and tetanus toxoid showed seroconversion for seven patients after inoculation. The data indicate that OprF-Oprl can be a useful vaccine in the therapeutic management of burn injuries. (C) 2003 Federation of European Microbiological Societies. Published by Elsevier Science B.V. All rights reserved.

Document Type: Article

Language: English

Author Keywords: Pseudomonas aeruginosa; outer membrane protein; vaccine; burn; infectious disease

KeyWords Plus: MEMBRANE PROTEIN-F; HETEROLOGOUS IMMUNOTYPE STRAINS; B-CELL EPITOPES; PROTECTIVE VACCINE; SYNTHETIC PEPTIDES; PORIN PREPARATION; INFECTION; IMMUNIZATION; ANTIBODIES; MODEL

Addresses: Mansouri, E (reprint author), Surg Univ Hosp, Hugstetterstr 55, D-79106 Freiburg, Germany
Surg Univ Hosp, D-79106 Freiburg, Germany
Burn Ctr, Ludwigshafen, Germany
MediGene AG, Planegg Martinsried, Germany

→ MATCHING ADDRESS FIELD TO AMADEUS

Publisher: ELSEVIER SCIENCE BV, PO BOX 211, 1000 AE AMSTERDAM, NETHERLANDS

Subject Category: Immunology; Infectious Diseases; Microbiology

IDS Number: 699GK

ISSN: 0928-8244

DOI: 10.1016/S0928-8244(03)00072-5

Note: This figure is an example of a publication page from the ISI Web of Science by Thomson. We systematically matched the name of all Amadeus firms (about 8 million names) to the address field in the publication page. See the text for details about the exact matching process. In addition to the name of the publishing firm, we also use information about the number of citations the publication received, the journal in which it was published, the publication year and the technology area.

Table A1

SUMMARY STATISTICS FOR FIRM PUBLICATIONS ACROSS MAIN FIELDS

Main field	# Publications	Mean Citations Received	Mean Citations Made	Mean of Journal Impact Factor	Average # Authors per publication
All fields	234,707	7.37	16.44	2.14	3.76
Biology & Chemistry	74,422	11.09	21.23	3.11	4.31
Engineering	51,507	3.14	10.44	0.82	2.87
Health & Medicine	48,254	7.55	14.11	2.84	4.36
Physics	22,107	8.35	19.49	2.08	3.97
Materials Science	16,452	5.50	13.72	1.03	3.45
Other	11,319	6.72	18.86	1.29	2.73
Computer Science	10,646	2.58	17.23	0.72	2.72

Notes: This table reports summary statistics for firm publications. Firm publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database (which includes about 20 million publications). The sample covers the period 1978-2004 and includes all Amadeus firms with at least one publication in "hard" science journals (here we do not condition on reporting financials). Mean citations received is the average number of forward citations a publication receives and mean citations made is the average number of citations the publications makes. The journal impact factor is based on the JCR index included in the ISI Web of Science database.

Table A2

SUMMARY STATISTICS FOR FIRM PUBLICATIONS ACROSS MAIN FIELDS:
MATURE FIRMS VERSUS YOUNG FIRMS

PANEL A: PUBLICATIONS BY MATURE FIRMS (AGE>21, MEDIAN)

Main Field	Publications	Mean of Citations Received	Mean of Citations Made
All fields	18,575	8.80	16.90
Biology & Chemistry	7,706	13.14	22.68
Computer Science	435	2.85	14.17
Engineering	4,215	2.98	9.27
Health & Medicine	3,428	8.45	15.59
Materials Science	1,536	5.76	12.91
Physics	815	8.44	16.29
Other	440	8.65	16.62

PANEL B: PUBLICATIONS BY YOUNG FIRMS (AGE ≤ 21, MEDIAN)

Main Field	Publications	Mean of Citations Received	Mean of Citations Made
All fields	20,257	7.16	17.62
Biology & Chemistry	6,295	10.45	23.34
Computer Science	1,041	1.52	17.09
Engineering	5,669	4.41	13.20
Health & Medicine	3,686	9.21	16.09
Materials Science	1,681	4.16	13.67
Physics	1,318	7.96	18.33
Other	567	5.09	19.31

Notes: This table reports summary statistics for firm publications, disaggregated according to whether the publishing firm is mature (age above the median) or young (age below the median). The sample covers the period 1978-2005 and includes all Amadeus firms with at least one publication in "hard" science journals that have at least 10 employees and at least $1 million in annual sales. Firm publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database.

Table A3

PUBLICATIONS CHARACTERISTICS ACROSS MAIN FIELDS IN LEADING JOURNAL

PANEL A: PUBLICATIONS BY MATURE FIRMS (AGE>21, MEDIAN)

Main Field	Publications	Mean of Citations Received	Mean of Citations Made
All fields	7,002	11.65	16.47
Biology & Chemistry	3,368	16.55	22.84
Computer Science	67	4.64	9.63
Engineering	1,165	2.05	5.51
Health & Medicine	1,667	9.02	12.03
Materials Science	344	9.52	14.45
Physics	275	10.01	15.98
Other	116	17.74	16.41

PANEL B: PUBLICATIONS BY YOUNG FIRMS (AGE ≤ 21, MEDIAN)

Main Field	Publications	Mean of Citations Received	Mean of Citations Made
All fields	6,394	11.40	18.74
Biology & Chemistry	2,651	14.33	25.04
Computer Science	101	1.09	9.74
Engineering	915	7.31	11.90
Health & Medicine	2,003	10.74	14.30
Materials Science	207	7.25	17.60
Physics	405	10.93	18.01
Other	112	5.89	17.69

Notes: This table reports summary statistics for firm publications in leading academic journals, separately for mature and young firms. Only publications in journals in the top quartile of the JCR index are included in this table. Firms in the sample have at least 10 employees and at least $1 million in annual sales.

TABLE A4-

SUMMARY STATISTICS FOR FIRM PUBLICATIONS ACROSS COUNTRIES

Country	Publications	Mean of Citations Received	Mean of Citations Made	Mean of Impact Rate of Journal	Average Number of Authors
All Counties	234,864	7.37	16.44	2.14	3.76
Belgium	6,012	7.85	17.17	2.26	4.26
Germany	40,282	6.51	17.20	2.16	3.88
Denmark	1,301	13.60	21.07	3.31	5.07
Spain	5,111	5.62	22.73	2.55	4.84
Finland	2,975	8.12	18.13	1.86	3.77
France	49,804	7.13	18.85	2.08	4.44
Great Britain	85,284	7.58	13.17	2.04	2.97
Greece	584	5.66	18.75	2.24	3.99
Italy	21,380	6.12	17.72	2.43	4.83
Netherlands	8,474	8.43	18.91	2.31	3.84
Norway	4,247	7.29	21.81	1.56	3.34
Sweden	9,410	11.97	17.26	2.28	3.46

Notes: This table reports summary statistics for firm publications across countries. The sample covers the period 1995-2004 and includes all Amadeus firms with at least publication in "hard" science journals between 1978 and 2004. Firm publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database.

TABLE A5-

FIRM PUBLICATIONS IN A SAMPLE OF LEADING SCIENTIFIC JOURNALS

Field	Journal	# of Firms	# of Publications	Citations per Publication	Mean of Employees of Firm
Molecular Biology & Genetics	Cell	8	10	186	12,068
	Nature Genetics	8	14	151	2,286
	Genes & Develpoment	3	4	8	12
	Nature Cell Biology	4	10	57	3,591
	Molecular Cell	2	3	98	808
Physical	Physical Review Letters	93	322	23	3,087
	Physical Review	22	50	32	19
	European Physical Journal	3	13	11	9
	Applied Physics Letters	167	595	22	11,129
	Europhysics Letters	63	107	14	3,856
Biology & Biochemistry	Nature Biotechnology	32	57	35	7,753
	Structure	12	22	51	6,334
	Nature Structural Biology	3	3	94	46,173
	Systematic Biology	1	1	7	2
	Biological Chemistry	17	34	6	8,750
Chemistry	Angewandte Chemie International Edition	67	192	22	11,591
	Analytical Chemisty	94	165	23	5,158
	Journal of Medicinal Chemistry	120	577	26	5,955
	Electrophoresis	57	132	14	6,656
	Chemical Revirews	11	15	64	39,789
Clinical Medicine	Nature Medicine	9	20	72	18,203
	Journal of the American Medical Association	11	20	9	2,338
	European Journal of Clinical Investigation	28	35	21	20,402
	Journal of the National Cancer Institute	16	18	28	10,322
	Lancet Neurology	4	7	9	399
Microbiology	Molecular Microbiology	19	40	22	10,817
	Journal of Virology	39	79	42	6,075
	International Journal of Antimicrobial Agents	43	70	2	4,504
	Applied and Environmental Microbiology	69	171	33	6,968
	Virology	29	51	29	8,635
Immunology	Immunity	10	22	108	1,691
	Journal of Immunology	61	159	39	2,541
	AIDS	51	95	11	4,397
	European Journal of Immunology	54	128	39	3,340
	Journal of Infectious Diseases	6	10	16	983

Notes: This table reports summary statistics on firm publications in selected top academic journals The sample covers the period 1995-2004 and includes all Amadeus firms with at least publication in "hard" science journals between 1978 and 2005. Firm publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database (which includes about 20 million publications).

TABLE A6-
MAIN FIRM CHARACTERISTICS ACROSS COUNTRIES

Country	Stock Market Value Traded / GDP	Private Credit / GDP	Accounting standards
Belgium	0.219	0.791	61
Denmark	0.717	1.675	62
Finland	1.288	0.712	77
France	0.717	0.959	69
Germany	0.569	1.229	62
Great Britain	1.898	1.605	78
Greece	0.235	0.788	55
Italy	0.527	0.925	62
Netherland	1.146	1.734	64
Norway	0.591	0.835	74
Spain	1.315	1.255	64
Sweden	1.304	1.117	83
Average	0.877	1.135	68

Notes: This table presents measures of financial developments for the countries in our sample. All measures are for 2005 and are based on Beck et al. (2000, 2007). Private Credit is the ratio of private credit by deposit money banks and other financial institutions. Bank Deposits is the demand, time and saving deposits in deposit money banks. Stock Market Value Traded is total shares traded on the stock market exchange. Stock Market Capitalization is the value of listed shares. All measures all divided by the Gross Domestic Product (GDP). Accounting standards is from Rajan and Zingales (2000).