Firm Size and Innovation: Evidence from European Panel Data

Sharon Belenzon
University of Oxford and NBER

Andrea Patacconi
University of Oxford

June 6, 2008

Abstract

Using novel firm-level panel data from European countries, this paper empirically investigates how the performance-innovation relationship varies with firm size. We distinguish between firm-level measures of applied research using patents (from both the European and US Patent Offices) and measures of basic research (using academic publication in "hard science" journals). We look at (total factor) productivity and growth as measures of firm performance. Our results indicate that the correlation of performance with applied research (patents) is stronger for small firms than for large firms. By contrast, the correlation of performance with basic research (academic publications) is stronger for large firms than small firms. A number of possible theoretical explanations for our findings are discussed.

Keywords: firm size, innovation, patents and publication
JEL Classification: O16, O32, G32 and G34

VERY PRELIMINARY AND INCOMPLETE

1. Introduction

Whether firm size affects the amount and composition of corporate R&D is an important question that has been hotly debated among economists and policymakers. Several theoretical arguments have been put forward in favor or against large size. According to Galbraith (1952), for instance, big firms may find it easier to internally generate the funds that are necessary to run large R&D programs. Large, diversified firms may also have an advantage when it comes to find useful applications for the uncertain outcomes of R&D, especially if their research is close to the basic-science end of the spectrum [Nelson (1959)]. On the negative side, bureaucracy and red tape could stifle entrepreneurship and creativity in large firms. Incumbents may even delay the development of new technologies to avoid cannibalizing the streams of rents from existing technologies [Arrow (1962), Reinganum (1983)]. Thus, in the end, whether small or large firms are more conducive to innovation is an empirical matter.

Unfortunately, the existing empirical literature on the relationship between firm size and innovation has produced rather fragile results.[1] One major problem has been the paucity of data, especially at the lower end of the firm-size distribution. As Cohen and Levin (1989) have stressed, in fact, most of the literature has restricted attention to very large firms (typically the 500 or 1000 largest firms in the manufacturing sector), which may not be representative of the whole distribution. Furthermore, with few notable exceptions [e.g., Mansfield (1981), Griliches (1986)], little is known about the composition of corporate R&D and the relationship between basic research, performance, and firm size.

We contribute to this literature by empirically investigating the relationship between

[1] For excellent surveys of this large body of research, see Cohen and Levin (1989) and Kamien and Schwartz (1975). More recently, Blundell, Griffith and Van Reenen (1999) have investigated the related question of how innovation and market share interact. In a panel of British firms, they find a robust, positive effect of market share on different measures of innovative output (headcounts of innovations and patents). Furthermore, the payoff from innovation is larger for high market share companies. They interpret these results as evidence of the importance of the 'efficiency effect' [see Gilbert and Newbery (1982)]; however, due to the necessity of having stock market data, they can only analyze a sample of relatively large firms.

firm size and innovation using a novel and comprehensive database on patents and academic publications. Patent data from the European Patent Office (EPO) and from the United States Patent and Trademark Office (USPTO) is systematically matched to all European firms. This dataset covers a wide distribution of firm size. About 10 percent of the innovating firms in our sample have less than 7 employees and less than $1 million in annual sales. Our dataset also contains new data on firm publications in academic journals. Publications are especially important since they may capture a type of research which is more basic (or science-based) than that captured by patents [e.g., Cockburn and Henderson (1998)]. Patents are in fact required by law to be very specific and tied down to well defined commercial applications. Academic publications, by contrast, are not judged upon their commercial application, but rather on their novelty and applicability to a wide range of scientific problems.

Simply put, our conjecture is that firms that conduct more basic research are more likely to publish than firms that have a stronger focus on applied research. The discovery of the transistor effect by a team of scientists working at the Bell Telephone Laboratories provides a nice illustration. After World War II, Bell started devoting substantial resources to basic research in semiconductors. This research resulted in a number of important academic publications and valuable patents.[2] Such cases do not appear to be isolated events. As Murray (2002) has argued, in fact, a given piece of knowledge can encompass both pure scientific content that is published in academic journals, and commercial applications that are patented. Murray (2002) and Murray and Stern (2006) identify scientific ideas that are initially published in academic journals and subsequently patented. Based on this work, one can interpret publications data as proxing for the scientific base associated with the patented inventions of firms. In other words, publications could be a useful proxy for

[2] The key article here is probably Bardeen J. and W.H. Brattain, published in the Physical Review in 1949 under the title "Physical Principles Involved in Transistor Action". See also the July 1949 issue of The Bell System Technical Journal devoted entirely to the discussion of the transistor and semiconductor devices. Two successful patents resulting from this research were filed in 1948: the one for the point contact transistor and the one for the junction transistor.

basic research conducted within firms.

We construct our measures of applied and basic research as follows. All granted patent applications from the EPO and the USPTO are used to generate a firm-level measure of applied research. To measure more basic research, all patenting firms are matched 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. 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 the publication. Finally, we systematically match publications data to firms and link this information to performance measures. By using multiple indicators of innovative activity, we thus hope to provide a richer and more nuanced view of how firm size and innovation interact.

Our main findings can easily be summarized. Private firms contribute substantially to the advancement of basic scientific knowledge. We matched about 200 thousand publications to firms over the period 1970-2004. These publications appear to be of high quality as indicated both by the number of citations they receive and by the impact factor of the journals in which they were published. We then investigate how the performance-innovation relationship varies with firm size, looking at total factor productivity and growth as measures of firm performance. Our results indicate that the correlation of performance with applied research (patents) is stronger for small firms than for large firms. By contrast, the correlation of performance with basic research (academic publications) is stronger for large firms than small firms. These correlations are robust to controlling for the quality of patents (as measured by patent citations) as well as the quality of academic publications (as measured by forward citations and the impact factor of the journal in which the article was published). Restricting attention to different subsets of the size distribution also does not alter the general picture.

Our results for the productivity-publications relationship relate to an earlier literature studying the *basic research premium*. Griliches (1986) adopted a production function approach to estimate the returns to basic research, where basic research was measured using survey data for large US firms.[3] His main finding was that this type of research was associated with a higher productivity premium than other types of R&D. In this paper, we are able to replicate his finding only for big firms. In our sample, for big firms the relationship between productivity and publications is more than three-fold larger than the relationship between productivity and patents. Yet, for small firms we do not observe an important premium for basic research. Importantly, the fact that the basic research premium varies with firm size mitigates the concern that this premium is driven solely by unobserved firm heterogeneity (as is the case, for instance, if academic publications capture unobserved labor-force skills) and allows us to identify specific channels through which this premium is generated (e.g., efficient internal capital markets).

In the second part of the paper we discuss possible theoretical explanations for our findings. As mentioned above, there are several arguments pointing to a causal relationship between firm size and innovation. Because of financial constraints, for instance, returns to innovation might increase with firm size, while the displacement effect suggests that the opposite may be true. We build on these ideas to illustrate why basic and applied research may exhibit the markedly different patterns that we see in the data. Three possible explanations are explored.

The first one is that a key advantage of internal capital markets – namely headquarters' superior ability to pick "winners", relative to external investors [Stein (1997)] – may be especially relevant in the case of basic research. A second explanation hinges on the relative severity of financial constraints for basic and applied research. Basic research is often more risky than applied research and its returns accrue very far in the future. Thus it is plausible that financial constraints may be more stringent for basic research. We

[3] A first attempt to measure basic research was made by Mansfield (1981). He used information on the composition of company-financed R&D expenditures in 1977 provided by 108 firms.

develop a simple model to show that this fact could generate patterns of returns to basic and applied research that are consistent with our observations. Finally, we use a variant of that model to formalize Nelson's (1959) diversification hypothesis, stating that a broad technological base may necessary to find use for the uncertain outcome of an R&D project, especially when research is close to the basic-science end of the spectrum. Throughout we discuss possible ways to empirically disentangle these different channels, but a careful empirical analysis of these issues is left for future work.

The remainder of the paper is organized as follows. Sections 2 and 3 describe the data and provide descriptive statistics. Section 4 reports the econometric results. Section 5 discusses possible explanations for our results. Section 6 concludes.

2. Data

This paper combines data from three main sources: (i) patents from the EPO and USPTO, (ii) academic publications from the Web of Knowledge database and (iii) financial information from Amadeus. In this section, we explain our methodology for constructing these data and describe our sample.

2.1. Patents

In order to generate a firm-level measure of applied innovation, we look at patent based measures which capture technological advances by firms [Griliches (1990) and Trajtenberg (1990)]. 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. In addition to patenting information, we also use patent citations data to measure the quality of patents. Patent quality is highly skewed and only few patents have significant economic value. A common method to proxy for the quality of patents is by counting the number of citations they receive [Trajtenberg (1990) and Hall et al. (2005)]. Another issue is that European firms typically have corporate controlling shareholder. Our matching process is at the firm level and is not aggregated to the highest shareholder level (as is the case, for example, for Compustat firms). We also test the robustness of our results to aggregating patents to the highest corporate level, yet such aggregation is likely to hamper the main advantage of our database: examining the effect of innovation on the performance of small firms.

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. 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. Having updated the USPTO patent database, we follow the same matching procedure as for the EPO to create the matched USPTO patent data for the Amadeus firms[4].

2.2. Academic Publications

Another measure of innovation is publication in academic journals. We develop systematic data on firm publications to proxy for science-based inventive activity by firms. 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

[4]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 (this includes patents that are registered in all three main patents offices: the EPO, JPO, and USPTO). To avoid double counting of inventions, information on patent families is needed. We collect this information from the OECD Triadic database on patent families. 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.

name to the complete ISI database. For each publication, we also have information on the number of citations, which we use to control for the quality of the publication. 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, because our main objective is using academic publications as a proxy for basic research, we want to control for publications that are likely to be linked to applied research and less to pure scientific advance. Such publications are most likely to appear in professional journals. To mitigate this concern and control for the quality of publication we follow two steps. First, 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. Second, we control for the importance of the journal in which the article was published by using the impact factor from the Journal Citations Report.

2.3. Accounting

Accounting information is taken from Amadeus. The source of the accounting information is the Company Register House in each of the twelve countries included in our sample. The key advantage of these data is their large coverage of firms and unique accounting information on private firms with a wide size distribution. Yet, the accounting data has some limitations. First, countries differ in reporting requirements. For example, very small firms (fewer than 10 employees) in Great Britain are not obliged to disclose accounting information including number of employees, sales, or total assets. On the other hand, French firms must provide such information regardless of their size. For very small firms (below 10 employees) we include in our estimation sample only those that report financials. In order to avoid selection bias in reporting financials (i.e., only the best small innovating firms voluntary disclose their financials), we test the robustness of our results by including only firms with more than 50 employees (which is well above the voluntary threshold for financial reporting).

3. Descriptive Statistics

Table 1 reports summary statistics for firms in our sample. About 13 thousand firms have at least one patent between 1979 and 2004. On average, these firms have about 7 patents (with a median of 1). Our sample covers a wide distribution of firm size, especially in the lower tail. The median firm in our sample generates about $20 million in annual sales and has 126 employees. 10 percent of the innovating firms in our sample have less than 7 employees and less than $1 million in annual sales[5]. About 6.5 thousand firms have at least one academic publication over the same period, with an average of 4 publications.

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 1,495 employees with a median of 150 employees. Out of the 13 thousand patenting firms, 1,613 firms also publish. These publishing firms have on average about 8 publications. Panel B of table 2 reports the same summary statistics for firms with at least one academic publication. European firms published about 200 thousand articles in academic journals between 1979 and 2004. The average publishing firm has 2,414 employees with a median of 113 employees. 25 percent of the publishing firms also patent. On average, these firms have about 4 patents. Firms in the publishing sample are, on average, about 20 percent more productive and about a third more capital intensive than firms in the patenting

[5] 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].

sample. This may indicate that publications are the result of capital intensive research, which may generate large productivity gains.

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. 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). The quality of firm publications is also high according to the Journal Citations Report indicator, which averages 3.8 for publications by firms and 2.5 for all other publications. Table A2 examines publications only for firms that report financials. We split the sample of publications according to the median number of employees. Publications by larger firms appear to receive and make more citations compared to publications by small firms. For example, in Biology and Chemistry, an average firm publication receives about 12 citations, while for small firms the average is of approximately 9 citations. 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 large firms receiving more citations than publications by small firms.

Figures 2 and 3 describe the relationship between firm labor productivity, innovation and size. We are especially interested in differences between patenting and publications, following the conjecture that patenting proxy for applied research, whereas publishing proxy for basic research. Figure 2 plots the relationship between labor productivity and patenting for firms of different size class. We split firms into high and low patenting categories according to whether their number of patents is above or below the median.

The bars then reflect the percentage difference in labor productivity between the high and low innovating firms. For small firms, labor productivity is much higher for high innovators than for low innovators. This means that conditional on being small, there are large productivity gains associated with patenting. By contrast, for big firms the difference in labor productivity between high and low innovators is much smaller. Thus conditional on being large, the productivity gains associated with patenting are lower compared to the case of small firms. This pattern of results could indicate that the displacement effect is important: the productivity gains associated with an incremental improvement are mitigated by the obsolescence of previous inventions, which is likely to be costly especially for large firms. Alternatively, it could be that small firms produce, on average, better inventions than large firms because they face more severe constraints. Such constraints would lead to small firm selection bias: only the best small firms that were able to overcome the constraints and come up with the invention enter our sample. For instance, small firms could face higher unit costs than large firms when filing a patent, or the cost of external finance could be higher for them [Hall (1989, 1992), Mayer (1992), Himmelberg and Peterson (1994)]. Later in the paper we test whether this selection bias is likely to drive our results by examining a sample of small and large non-innovating firms.

Figure 3 provides a similar description of the relationship between productivity gains to publishing and firm size. Here we observe the opposite pattern than for patenting. Productivity gains associated with publishing are much higher when firms are large. Since the cost of submitting a paper for publication is likely to be negligible, this result suggests that large firms may have a comparative advantage in developing commercial applications of scientific breakthroughs. In Section 5 we explore this idea in greater depth.

4. Econometric Results

Our econometric analysis focuses on identifying robust correlations between firm size and the private gains from patenting and publishing. More specifically, we examine the effects of the stocks of patents and publications on firm performance (total factor productivity and sales growth) and analyze how these correlations vary across firms of different size. Our analysis differs from previous "classical" productivity estimations [e.g., Griliches (1986)] in that we do not directly observe R&D expenditures. This implies that we cannot compute net returns to innovation, as the costs associated with the innovative output are not observed. We thus focus on differences in the private gains associated with innovation (both basic and applied) by small and large firms.

4.1. Firm Size and Patents Stock

Table 3 reports the relation between patents stock and (total factor) productivity and examines how this relation varies with firm size. Columns 1 to 4 include all innovation firms, that is, firms that have at least one patent or academic publication between 1979 and 2004. Column 1 reports the effect of lagged patents stock on sales, controlling for employment, capital and complete sets of three-digit SIC, country and year dummies. The coefficient on the stock of patents is positive and highly significant (an elasticity of 0.055). In column 2 we interact patents stock with the lagged number of employees. The coefficient on this interaction is negative and highly significant (-0.012 with a standard error of 0.003). Column 3 adds firm fixed-effects to control for firm unobserved time-invariant heterogeneity (which may be correlated with patenting). The coefficient on the patents stock actually rises and remains highly significant (0.073 with a standard error of 0.014). The same pattern of results also holds when including an interaction between firm number of employees and patents stock (column 4). This means that the productivity gains associated with a given level of patents stock fall with the size of the innovating firm. Or in other words, an additional patent is associated with higher productivity gains, in percentage terms, for small firms than for large firms.[6]

To test the robustness of our results we also examine the relation between patenting

[6] We also interacted lagged employment with lagged capital stock because of the high correlation patent and capital stocks in our sample. The results are robust for adding this interaction.

and productivity for different sub-samples of firm sizes. In columns 5 and 6 we split observations according to the number of lagged employees. Columns 5 includes only observations where the number of employees is in the lowest quartile (25 employees). The elasticity of sales with respect to patents stock is 0.126 (with a standard error of 0.044). This elasticity falls as we examine sub-sets of larger firms. For example, for firms in the highest employment quartile (column 8), the elasticity of sales with respect to patents stock falls to 0.038 (with a standard error of 0.010). Columns 9 to 12 exclude firms that never patent to mitigate the concern of including non-for-profit firms in our sample. The same pattern of results holds, meaning that for small firms, patenting is associated with higher productivity gains as compared to large firms.

4.2. Firm Size and Publications Stock

Table 4 reports the estimation results when the publications stock is added. The estimation sample includes all innovating firms (patenting and publishing). Column 1 includes linearly patents and publications stocks. The coefficient on patents stock remains almost unchanged, while the coefficient on publications stock is positive and significant (0.030 with a standard error of 0.015). The same pattern emerges when the interaction between patents stock and size is added. This result is interesting because under the assumption that publications measure the quality of patented knowledge, we would expect the coefficient on patents stock to fall when controlling for publications. Thus the stability of the patents stock coefficient suggests that publications capture additional information about a firm's inventive activities, not just unobserved patents quality.

Column 4 includes an additional interaction between the lagged number of employees and publications stock. The coefficient on this interaction is positive and significant (0.012 with a standard error of 0.004). This means that the productivity gains associated with a given level of publications stock rise with the size of the innovating firm. This result is also inconsistent with the hypothesis that publications measure unobserved patent quality.

There is no clear reason to suspect that publications would measure patent quality better for large firms than for small firms.

Columns 5 to 8 examine the robustness of this result by splitting the size of the sample according to firm size. Column 5 includes only observations where the lagged number of employees is below the median. The elasticity of patents stock is positive and highly significant (0.085 with a standard error of 0.025), where the elasticity of publications stock is negative and not significant (-0.016 with a standard error of 0.034). Column 6 includes only observations where the lagged number of employees is above the median. For this sub-sample, the elasticity of publications stock is positive and highly-significant (0.048 with a standard error of 0.015), where the elasticity of patents stock is still positive and significant, but lower than for the smaller firms. In columns 7 and 8 we include only observations where the number of employees is above 435 (75th percentile) and 1,600 (90th percentile), respectively. The elasticity of publications stock rises to 0.083, where the elasticity of patents stock falls to 0.029 (column 8).

Our results about the relationship between firm productivity and publications relate to an earlier literature on the returns to basic research. Griliches (1986) also adopted a production function approach to measure the returns to basic research, where basic research was measured using survey data for large US firms. He found that a small number of typically very large firms were responsible for a substantial fraction of total and basic R&D expenditures and that basic research was associated with a much high premium than other types of research.[7] In our sample, this high premium for basic research only exists for big firms. Indeed, for firms with more than 1200 employees the relationship between productivity and publications is more than three-fold larger than that between productivity and patents. By contrast, for small firms, we do not observe an important premium for basic research, as measured by academic publications. Note that the fact

[7] Specifically, he found that the productivity gains associated with basic research were eight times larger than the productivity gains associated with other types of research.

that our basic research 'premium' varies with firm size helps mitigate the concern that it is driven solely by unobserved firm heterogeneity. If, in fact, academic publications captured just unobserved labor-force skills, we would expect this premium to be roughly constant as firm size varies.

As reported in table A7 in the appendix, the same pattern of results holds when we screen firm publications according to quality measures, such as the number of forward citations a publication receives and the impact factor of the journal in which the article was published.

Table 5 reports similar estimations for firm growth (measured separately by employment and sales growth). The pattern is similar to the one found in the productivity estimation. The contribution of patents stock to firm growth is much more prominent for small firms. By contrast, the correlation of firm growth with publications is stronger for large firms than for small firms.

4.3. Compustat versus AmaPat

The main advantage of our new European dataset, AmaPat, is that it captures a wide range of firm size. To demonstrate this advantage, we estimate similar productivity-innovation specifications for Compustat firms and large European firms. The sample of Compustat firms includes all patenting firms, where patents data is taken from the NBER archive. We follow the same matching procedure as for the European firms to assign publications to Compustat firms. The average US firm in our sample, which covers the period 1980-2001, has 14,843 employees with a median of 2,625 employees (as compared to AmaPat, where the average firm has 1,370 employees with a median of 126). Table 6 reports the estimation results of the variation of the productivity-innovation relationship across firm size. Column 1 reports the estimation results of the firm productivity against patent and publication stocks. The coefficient on the patent stock is negative and not significant while the coefficient on the publications stock is positive and significant (0.022 with a standard error of 0.008). This result is somewhat consistent with our result that, for large firms, the correlation between productivity and publications should be strong while the correlation between productivity and patenting should be weak. In columns 2 and 3 we add interactions between a firm's patents and publications stocks and employment. None of the interactions is significant. In columns 4-9 we perform similar estimations for large European firms. In none of these specifications we are able to replicate the results on firm size and the productivity gains associated with patenting and publishing.

4.4. Innovation Intensity and Firm Size

Table 7 investigates Schumpeter's hypothesis that size should be positively associated with innovation intensity in our dataset. Two measures of innovation intensity are considered: patents stock to sales ('patents intensity') and publications stock to sales ('publications intensity'). The sample is cross-sectional for 2003, the year for which we have most firms. Our proxy for firm size is employment in 2000.

Column 1 gives the results of our simplest specification. Patents intensity appears to be strongly negatively associated with firm size, thus contradicting Schumpeter's hypothesis. However, by adding past employment squared in column 2, a significant nonlinearity emerges. Patents intensity appears to be largest for both very large and very small firms, a finding that resonates well with the results by Bound et al. (1984). The inflection point is at about 170 employees, which is well within our firm size range. In column 3 and 4 we check the robustness of our results by excluding publishing firms that do not patent. The qualitative results are the same but the magnitude of the coefficients nearly doubles. The location of inflection point remains essentially unchanged.

Columns 5-8 look at publications intensity. The same pattern of results emerges. The main difference is that the coefficient on past employment in the linear regression is only marginally significant when the sample includes both publishing and patenting firms. In the nonlinear regression, however, all coefficients are highly significant. Again, publications intensity appears to be largest for both very large and very small firms. The inflection point is at about 160 employees, essentially the same as for patents. Similar qualitative results obtain when we restrict attention to publishing firms only.

4.5. Selection into Innovation

Our serious concern is that our results about patents could be driven at least in part by a small firm selection bias. Suppose in fact that small firms face greater difficulties in developing and implementing their ideas than large firms. Then many of their projects would fail to reach development stages – typically the low quality ones. As a result, the average quality of inventions by small firms should be higher than that of large firms.

To test this idea we use data on non-innovating firms (that is, firms that never patent or publish). We randomly sample 10 percent of all Amadeus firms that report financial information and that have not been matched to the patents or publications data. The selection hypothesis implies that the effect of innovating on productivity should be larger for small firms than for large firms.

Table 8 estimates the effect an innovation indicator on productivity and examines the extent to which this effect varies with firm size. The innovation indicator is defined as a dummy variable that receives the value of one for firms that innovate (patent or publish) and zero for firms that never innovate. All regressions are cross-sectional for the year 2004. Column 1 includes a linear dummy for innovating. The coefficient on this dummy is positive and significant, indicating a 6 percent average productivity premium for innovation. Column 2 adds an interaction term between the innovation dummy and the size of the firm (as previously, measured by lag employment). The coefficient on this interaction is positive, but not significant. This means that we cannot reject the hypothesis that selection into innovation is identical across firm size. Columns 3 to 6 report the estimation of the effect of innovating for different sub-samples of firm size. The pattern of results suggests that as the size of firms increase, the effect of innovating on productivity rises. Clearly this pattern is the opposite of what we would expect under the selection hypothesis.

5. Basic and Applied Research: Theoretical Considerations

In this section we discuss possible explanations for our empirical findings. Our focus will be on differences between basic and applied research that may help rationalize the markedly different patterns that we observe in the data.

5.1. Internal Capital Markets

Large firms frequently reallocate scarce resources across projects through internal capital markets. The costs and benefits of such markets (relative to external finance) have been studied extensively. On the positive side, corporate headquarters may have informational advantages over external investors that could be conducive to better financing decisions [Stein (1997), Guedj and Scharfstein (2005)]. Furthermore, while external investors do not have residual rights over the assets they invest in (only in case the firm defaults), headquarters owns and controls the project in which it invests and therefore has stronger incentives to improve its quality [Gertner, Scharfstein and Stein (1994)]. On the negative side, internal capital markets may be prone to agency problems and influence activities, which could lead to biased decision-making [Meyer, Milgrom and Roberts (1992), Scharfstein and Stein (1998)].

A plausible conjecture is that the benefits of having an active internal capital market (which is typically related to the size of the firm) may be particularly large in the case of basic research. A straightforward extension of Stein's (1997) 'winner-picking' model suffices to make this point.

In Stein's model there are two types of firms: large firms running several projects and small firms running only one project. Because of moral hazard, all firms are cash constrained. In that setting the headquarters of a large firm can perform a useful role by reallocating resources from projects with low NPV to projects with high NPV. Stein makes the simplifying assumption that headquarters perfectly observes whether returns are low or high, whereas external investors observe nothing. It is not hard to relax this assumption. In general, the value of having an internal capital market rises with the amount of private information that the headquarters has (in principle, this value could be negative because supervision by headquarters generates an 'effort-dilution' effect). Thus Stein's model suggests that large firms should perform better than small firms when projects cannot easily be evaluated by outsiders (i.e., when the amount of asymmetric information between headquarters and external investors is large), and should perform poorly when the opposite is true. To the extent that it is harder for outsiders to evaluate basic research than applied research (relative to insiders), this yields precisely the pattern of returns that we observe in the data.[8]

5.2. Financial Constraints and the Displacement Effect

Financial constraints may also help explain our empirical findings if they are particularly stringent in basic research. We illustrate this idea in the context of a simple model of R&D where firms can choose among different types of research.

The economy is populated by a continuum of firms. Each firm is characterized by a productivity parameter θ and initially operates in a single, 'primary' market. Both the short-run profits in the primary market, $\pi_1(\theta)$, and the optimal number of workers employed by the firm, $L^*(\theta)$, are assumed to be increasing in θ.[9] Thus θ is a measure of firm size.

A firm can engage in R&D to expand its business into different markets. There are two periods. In the first period, short-run profits accrue and investment decisions are made,

[8] Empirically, one could try to use cross-industry variation in our data to provide exogenous variation for the importance of internal capital markets. Industry measures such as Productivity Growth Dispersion and average Tobin's Q might be good starting points.

[9] This is the case if, for instance, L^* solves $\max_L \{\theta R(L) - wL\}$, where R is a revenue function (increasing in L) and w is the wage.

subject to financial constraints. In the second period, the firm may introduce a new product. Long-run profits are $\pi_2(\theta)+\Delta\pi_2(\theta)$ if the new product is introduced, and $\pi_2(\theta)$ if the new product is not introduced. We assume that $\Delta\pi_2(\theta)$ is (weakly) decreasing in firm size θ. This captures the idea that large firms may be less inclined to invest in R&D, since the new product may cannibalize the streams of rents from existing technologies (i.e., Arrow's (1962) *displacement effect*). This assumption is satisfied if, for instance, the new product reduces sales in the primary market by a given fraction.

To increase the chances of developing a new product, the firm can engage in basic or applied research (or both). Either type of research costs I and stochastically produces new knowledge.[10] Let $\underline{k}_A$ and $\underline{k}_B$ denote the firm's initial stocks of technical and scientific knowledge, respectively. Engaging in applied research increases the stock of technical knowledge from $\underline{k}_A$ to k_A, where k_A is a random variable defined on $\left[\underline{k}_A, \bar{k}_A\right]$. Similar remarks apply to basic research and the firm's stock of scientific knowledge (denoted by k_B). Thus, due to the uncertainties of the R&D process, firms of the same size may well end up with with different knowledge stocks. The probability that a new product is introduced is given by the separable function $q(k_A, I_A)+r(k_B, I_B)$, where $I_A, I_B \in \{0, I\}$.[11] We normalize $q(\underline{k}_A, 0)$ and $r(\underline{k}_B, 0)$ to zero and define $\alpha \equiv q(k_A, I)$, $\beta \equiv r(k_B, I)$.[12] Of course $\partial\alpha/\partial k_A > 0$ and $\partial\beta/\partial k_B > 0$. The expected values of α and β are denoted by $\bar{\alpha}$ and $\bar{\beta}$, respectively.

We assume that both investments have positive net present value, but that the investment in applied research is ex-ante more profitable:

$$\bar{\alpha}\Delta\pi_2(\theta) - I > \bar{\beta}\Delta\pi_2(\theta) - I > 0.$$

[10] The investment I can of course be interpreted as an additional expenditure, on top of the firm's normal R&D expenditures.

[11] This rules out complementarities or substitutability between basic and applied research. See Aghion and Tirole (1994) for a similar assumption.

[12] More generally, what we need is some form of complementarity between knowledge and investment so that, for instance, $\partial q(k_A, I_H)/\partial k_A > \partial q(k_A, I_L)/\partial k_A$ for $I_H > I_L$. In other words, investments in R&D must have a component of development, not just research.

This is obviously the case if the investment in basic research is very risky: $\bar{\beta} < \bar{\alpha}$.

At the end of the first period, k_A and k_B are realized and, conditional on its stock of knowledge, the firm's expected long-run profits (denoted by Π^{LR}) are $\pi_2(\theta)$ if no investment is made, $\pi_2(\theta) + \alpha\Delta\pi_2(\theta)$ $(\pi(\theta) + \beta\Delta\pi_2(\theta))$ if the firm invests in applied (basic) research, and $\pi_2(\theta) + (\alpha+\beta)\Delta\pi_2(\theta)$ if the firm invests both in applied and basic research. In the second period, profits accrue.

Due to credit market imperfections, the firm may be unable to raise all the money it needs. Specifically, we posit that a firm with short-run cash flow $\pi_1(\theta)$ can raise at most $\mu\pi_1(\theta)$, where $\mu \geq 1$ measures the extent to which the firm can borrow using its short-term profits as collateral.[13] Since investments in applied research yield higher ex-ante returns than investments in basic research, financial constraints are more likely to be binding for basic research. Three cases must thus be considered: $\mu\pi_1(\theta) < I$ (no investment is made), $I \leq \mu\pi_1(\theta) < 2I$ (only applied research is undertaken), and $\mu\pi_1(\theta) \geq 2I$ (both types of research are undertaken).[14]

The interplay between financial constraints and the displacement effect can easily generate the kind of patterns that we observe in the data. To see this, suppose that $\mu\pi_1(\theta) \geq I$, so that applied research is always undertaken. The derivative of the expected long-run profits Π^{LR} with respect to k_A has the same sign as

$$\frac{\partial \Pi^{LR}}{\partial \alpha} = \Delta\pi_2(\theta)$$

which, because of the displacement effect, is decreasing in θ. Thus the returns to technical knowledge (or 'patents') are decreasing in firm size. Now consider basic research. The derivative of Π^{LR} with respect to k_B has the same sign as

$$\frac{\partial \Pi^{LR}}{\partial \beta} = \begin{cases} 0 & \text{if } \mu\pi_1(\theta) < 2I \\ \Delta\pi_2(\theta) & \text{if } \mu\pi_1(\theta) \geq 2I \end{cases}.$$

[13] See Aghion et al. (2007) and Almeida and Campello (2006) for similar assumptions.

[14] Note that in this simple model firm size affects the financial constraints only through the short-run cash-flow $\pi_1(\theta)$. In practice, however, size is related to financial constraints through several channels. For instance, both unit bankruptcy costs and transaction costs for new share or bond issues typically decrease with size. See Schiantarelli (1996) for a careful discussion.

Here, despite the displacement effect, the returns to scientific knowledge (or 'publications') may well increase with firm size, since only large firms have the financial resources to conspicuously invest in R&D.

We can gain further insight by considering an extension of the model where some firms are always constrained while others are not. To introduce variation among firms of the same size, suppose that firms may be hit by an adverse shock and be required to plow in extra cash D to remain afloat (D could be negative – a positive shock). If D is drawn from the cumulative distribution $F(\cdot)$, a fraction $F(\mu\pi_1(\theta) - I)$ of the firms will engage in applied research, and a fraction $F(\mu\pi_1(\theta) - 2I)$ will engage in both basic and applied research. Expected long-run profits are thus

$$\Pi^{LR} = \pi_2(\theta) + \alpha\Delta\pi_2(\theta) F(\mu\pi_1(\theta) - I) + \beta\Delta\pi_2(\theta) F(\mu\pi_1(\theta) - 2I).$$

We have that

$$\begin{aligned}
\frac{\partial^2\Pi^{LR}}{\partial\alpha\partial\theta} &= \frac{\partial\Delta\pi_2(\theta)}{\partial\theta} F(\mu\pi_1(\theta) - I) + \Delta\pi_2(\theta) f(\mu\pi_1(\theta) - I)\,\mu\frac{\partial\pi_1(\theta)}{\partial\theta} \\
\frac{\partial^2\Pi^{LR}}{\partial\beta\partial\theta} &= \frac{\partial\Delta\pi_2(\theta)}{\partial\theta} F(\mu\pi_1(\theta) - 2I) + \Delta\pi_2(\theta) f(\mu\pi_1(\theta) - 2I)\,\mu\frac{\partial\pi_1(\theta)}{\partial\theta}.
\end{aligned}$$

Two opposite effects emerge. In both equations the first term is negative because of the displacement effect. The second term, however, is positive since knowledge is beneficial only when firms invest in R&D. Note that, in general, the displacement effect will be more important (on average) for applied research than for basic research because more firms engage in applied research: $F(\mu\pi_1(\theta) - I) \geq F(\mu\pi_1(\theta) - 2I)$. In particular, if D is uniformly distributed, one obtains

$$\frac{\partial^2\Pi^{LR}}{\partial\alpha\partial\theta} < \frac{\partial^2\Pi^{LR}}{\partial\beta\partial\theta}$$

which is consistent with our evidence.

A second prediction can be obtained by varying the cost of R&D. Indeed, for $I_H > I_L$ (and assuming again that D is uniformly distributed), we get $\frac{\partial^2\Pi^{LR}}{\partial\alpha\partial\theta}(I_L) < \frac{\partial^2\Pi^{LR}}{\partial\alpha\partial\theta}(I_H)$ and

$\frac{\partial^2 \Pi^{LR}}{\partial\beta\partial\theta}(I_L) < \frac{\partial^2 \Pi^{LR}}{\partial\beta\partial\theta}(I_H)$. The reason is simply that few firms experience displacement if R&D costs are high. Empirically, therefore, we should expect our interaction coefficients to be larger in sectors where firms, for exogenous reasons, depend more on external finance.

5.3. Diversification

Diversification can also help explain our empirical findings. To illustrate this point, we modify the above model to capture the basic idea in Nelson (1959). Consider a firm composed of n divisions, or product lines.[15] Total profits are given by $\pi_1(n)$, where $\pi_1(\cdot)$ is increasing. As before, we assume that the firm can engage in basic or applied research, or both. Throughout, we abstract from financial constraints.

There are two main differences between this model and the previous one. The first is that applied and basic research are now assumed to facilitate the invention of two distinct products, A and B. Thus, by engaging in applied research, the firm can increase the likelihood that product A is introduced, and similarly for basic research. As in the previous model, the additional profits in case of success, $\Delta\pi_A(n)$ and $\Delta\pi_B(n)$, are assumed to be weakly decreasing in firm size (or diversification, as parametrized by n).

The second difference captures the essence of Nelson's (1959) diversification hypothesis. Nelson claimed that while applied research can be tied to the solution of a specific practical problem or the creation of practical object, basic research is characterized by a higher degree of uncertainty. This greater uncertainty makes having a broad technological base important because it "insures that, whatever direction the path of research may take, the results are likely to be of value to the sponsoring firm. It is for this reason that firms which support research toward the basic-science end of the spectrum are firms that have their fingers in many pies" (Nelson (1959), p.302). Note that, implicit in this argument, there is the assumption that "a single-product firm is unable to exploit an invention not

[15] For simplicity, n is here a parameter of the model. However, the optimal n could easily be derived endogenously in a model where managerial ability places restrictions to the number of divisions that a manager can efficiently supervise.

directly linked to its primary product, through licensing to others or developing a new product line" (Kamien and Schwartz (1975), p.15).

We capture these ideas as follows. We posit that applied research can easily be tied to the solution of problems related to the firm's n primary products. Thus the probability of successfully introducing a new product is just, as before, $\alpha = q(k_A, I)$. For basic research, however, things are more difficult. Indeed, due to its greater uncertainty, the link between basic research and primary products can only be established with probability $h(n)$, where $h(\cdot)$ is increasing in n. The probability that a new product can be introduced is therefore $\beta h(n)$, where $\beta = r(k_B, I)$. Conditional on (α, β), expected long-run profits are thus

$$\Pi^{LR} = \pi_2(n) + \alpha \Delta \pi_A(n) + \beta h(n) \Delta \pi_B(n).$$

Note that while the returns to applied research always decrease with firm size, $\partial^2 \Pi^{LR} / \partial \beta \partial \theta$ may well be positive if h grows sufficiently fast as diversification increases.

Empirical investigation of the diversification hypothesis is not straightforward. Only about ten percent of firms in our sample are diversified (operating in more than one two-digit industry segments). Yet, unlike in the US, European firms are likely to belong to diversified business groups [Belenzon and Berkovitz (2007)]. Thus, although a firm itself might operate in only one industry segment, it could be tied via ownership links to other firms that operate in different segments (about 30 percent of firms in our sample belong to business groups). This suggests that diversification in the European context should be studied in a business groups framework, and should consider the complex issues associated with groups structure.[16].

6. Conclusion

This paper empirically investigates how the performance-innovation relationship varies with firm size using a novel and comprehensive firm-level panel data from 12 European

[16] See Khanna and Yafeh (2006) for a survey of the literature on business groups.

countries. Unlike previous contributions, our database covers a wide distribution of firm size, which allows us to establish more robust correlations between the variables of interest. We also distinguish between different types of innovative activity. In particular, we use all granted patent applications from the European Patent Office to generate a firm-level measure of applied research, while basic research is proxied by the number of publications in 'hard' science journals.

We find that private corporations contribute substantially to the advancement of basic scientific knowledge. Furthermore, the correlation of performance with applied research (patents) is stronger for small firms than for large firms. By contrast, the correlation of performance with basic research (academic publications) is stronger for large firms than small firms.. A number of possible theoretical explanations for our findings are also discussed, including the role of internal capital markets, financial constraints, and diversification.

There are several ways in which our findings could be extended. The most important limitation of our analysis is that it only highlights correlations, without showing causality. In particular, we do not try to disentangled the specific channels through which firm size might affect the incentives to innovate. Is the positive correlation between size and publications the result of efficient internal capital markets, or is it more related to diversification and knowledge spillovers? A natural place to start would be the rich cross-industry and cross-country variation in our data, that could provide exogenous variation for the importance, for example, of internal capital markets for innovation [e.g., Belenzon and Berkovitz (2007)].

References

Aghion, P., P. Askenazy, N. Berman, G. Cette and L. Eymard, (2007), 'Credit Constraints and the Cyclicality of R&D Investment: Evidence from France', Working Paper.

Aghion, P. and J. Tirole, (1994), 'The Management of Innovation', *Quartely Journal of Economics*, Vol. , pp.1185-1209..

Almeida, H. and M. Campello, (2006), 'Financial Constraints, Asset Tangibility and Corporate Investment', forthcoming, *Review of Financial Studies.*

Arrow, K., (1962), 'Economic Welfare and the Allocation of Resources of Inventions'. In R.R. Nelson, ed., *The Rate and Direction of Inventive Activity*, Princeton, NJ: Princeton University Press.

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

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

Blundell, R., R. Griffith and J. Van Reenen, (1999), 'Market Share, Market Value and Innovation in a Panel of British Manufacturing Firms', *Review of Economic Studies,* Vol. 66, pp. 529-554.

Bound, J., C. Cummins, Z. Griliches, B. Hall, and J. Jaffe, (1984), 'Who Does R&D and Who Patents?'. In Z. Griliches, ed., *R&D, Patents, and Productivity*, Chicago: University of Chicago Press.

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(2), pp.157-182.

Cohen, W.M. and R.C. Levin, (1989), 'Innovation and Market Structure'. In R. Schmalensee and R.D. Willig, eds., *Handbook of Industrial Organization, Volume II*, Elsevier Science

Publisher.

Galbraith, J.K., (1952), *American Capitalism*, Boston: Houghton-Mifflin.

Gertner, R., D.S. Scharfstein and J. Stein, (1994), 'Internal versus External Capital Markets', *Quarterly Journal of Economics*, Vol. 109, pp. 1211-1230.

Gilbert, R. and D. Newbery, (1982), 'Preemptive Patenting and the Persistence of Monopoly', *American Economic Review*, Vol. 72, pp. 514-526.

Griliches, Z. (1986), "Productivity, R&D, and Basic Research at the Firm Level in the 1970s", *American Economic Review*, Vol. 76 (1), pp. 141-154.

Guedj, I. and D. Scharfstein (2005), 'Organizational Scope and Investment: Evidence from the Drug Development Strategies and Performance of Biopharmaceutical Firms', Working Paper.

Henderson, R. and I.M. Cockburn, (1996), 'Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery', *RAND Journal of Economics*, Vol. 27, pp. 32-59.

Himmelberg, C. and B. Petersen, (1994), 'R&D and Internal Finance: A Panel Study of Small Firms in High-Tech Industries', *Review of Economics and Statistics*, Vol. 76, pp. 38-51.

Kamien, M.I. and N.L. Schwartz, (1975), 'Market Structure and Innovation: A Survey', *Journal of Economic Literature*, Vol. 13, pp. 1-37.

Khanna, Tarun, and Yishay Yafeh. (2006). "Business Groups in Emerging Markets: Paragons or Parasites?" *Journal of Economic Literature* (forthcoming).

Mansfield, E., (1981), 'Composition of R and D Expenditures: Relationship to Size of Firm, Concentration, and Innovative Output', *Review of Economics and Statistics*, Vol. 62, pp. 610-615.

Meyer, M., P. Milgrom and J. Roberts, (1992), 'Organizational Prospects, Influence Costs, and Ownership Changes', *Journal of Economics & Management Strategy*, Vol. 1, pp. 9-35.

Murray, F., (2002), 'Innovation as Co-Evolution of Scientific and Technological Networks: exploring Tissue Engineering', *Research Policy*, Vol. 31, Issues 8-9, 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*.

Nelson, R.R., (1959), 'The Simple Economics of Basic Scientific Research', *Journal of Political Economy*, Vol. 67(3), pp.297-306.

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

Reinganum, J., (1983), 'Uncertain Innovation and the Persistence of Monopoly', *American Economic Review*, Vol. 73, pp.61-66.

Scharfstein, D.S. and J. Stein, (2000), 'The Dark Side of Internal Capital Markets: Divisional Rent-Seeking and Inefficient, Investment', *Journal of Finance*, Vol. , pp. 2537-2564.

Schiantarelli, F. (1996), 'Financial constraints and investment: methodological issues and international evidence', *Oxford Review of Economic Policy*, Vol. 12, pp. 70-89.

Schumpeter, J.A., (1942), *Capitalism, Socialism, and Democracy*, New York: Harper.

Stein, J., (1997), 'Internal Capital Markets and the Competition for Corporate Resources', *Journal of Finance*, Vol. 52, pp.111-133.

A. Appendix

A.1. Matching patent data

A.1.1. 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).[17] 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.[18] 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 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.

[17] 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).

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

A.1.2. 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.[19] 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.[20] 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.[21] 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.2. 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

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

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

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

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).

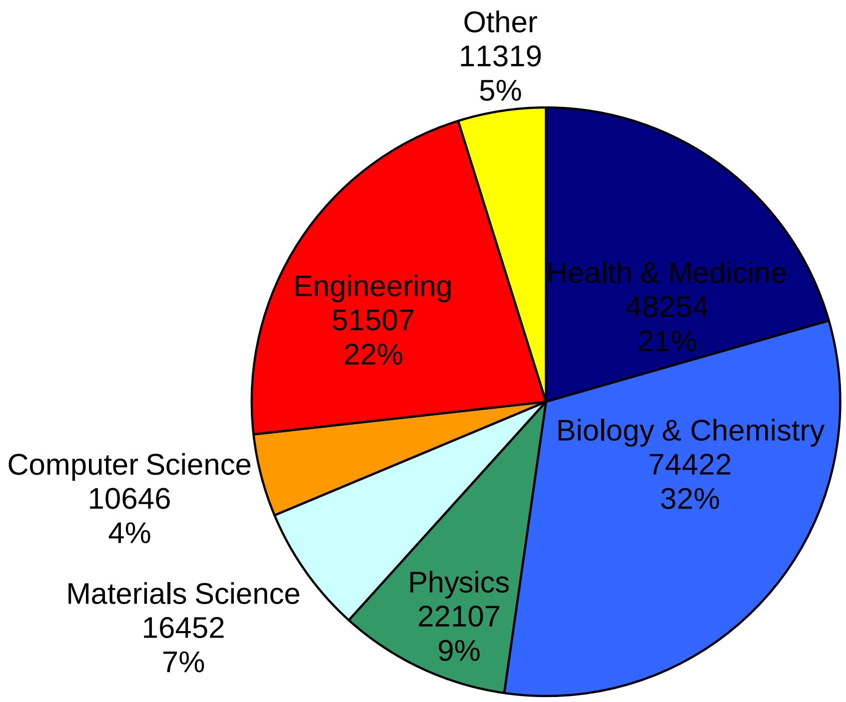
A.3. Accounting database

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).

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.

TABLE 1-

SUMMARY STATISTICS FOR MAIN FIRM CHARACTERISTICS

					Distribution		
Variable	# firms	# Obs	Mean	Std. Dev.	10st	50th	90th
Patents stock	13,087	58,445	6.8	51.5	0.1	0.9	10.9
Publications Stock	6,516	27,657	3.9	28.7	0	0.5	5.7
Sales (`000)	14,251	60,926	435,158	3,933,684	921	21,153	379,236
Employess	17,047	74,321	1,370	10,458	7	126	1,423
Age	16,659	73,876	28	26	5	19	65
Employment growth	17,047	74,321	0.02	0.44	-0.04	0	0.22
Sales Growth	13,741	58,993	0.09	0.53	-0.06	0.09	0.42
Capital (`000)	15,094	64,534	278,911	3,309,300	93	3,822	145,348
Cash flow (`000)	14,341	60,156	43,743	489,349	-523	971	30,371
Capital/Employee (`000)	13,997	61,305	619	30,196	6	37	212
Sales/Employee (`000)	13,088	57,404	263	309	74	176	502

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 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic 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). 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.

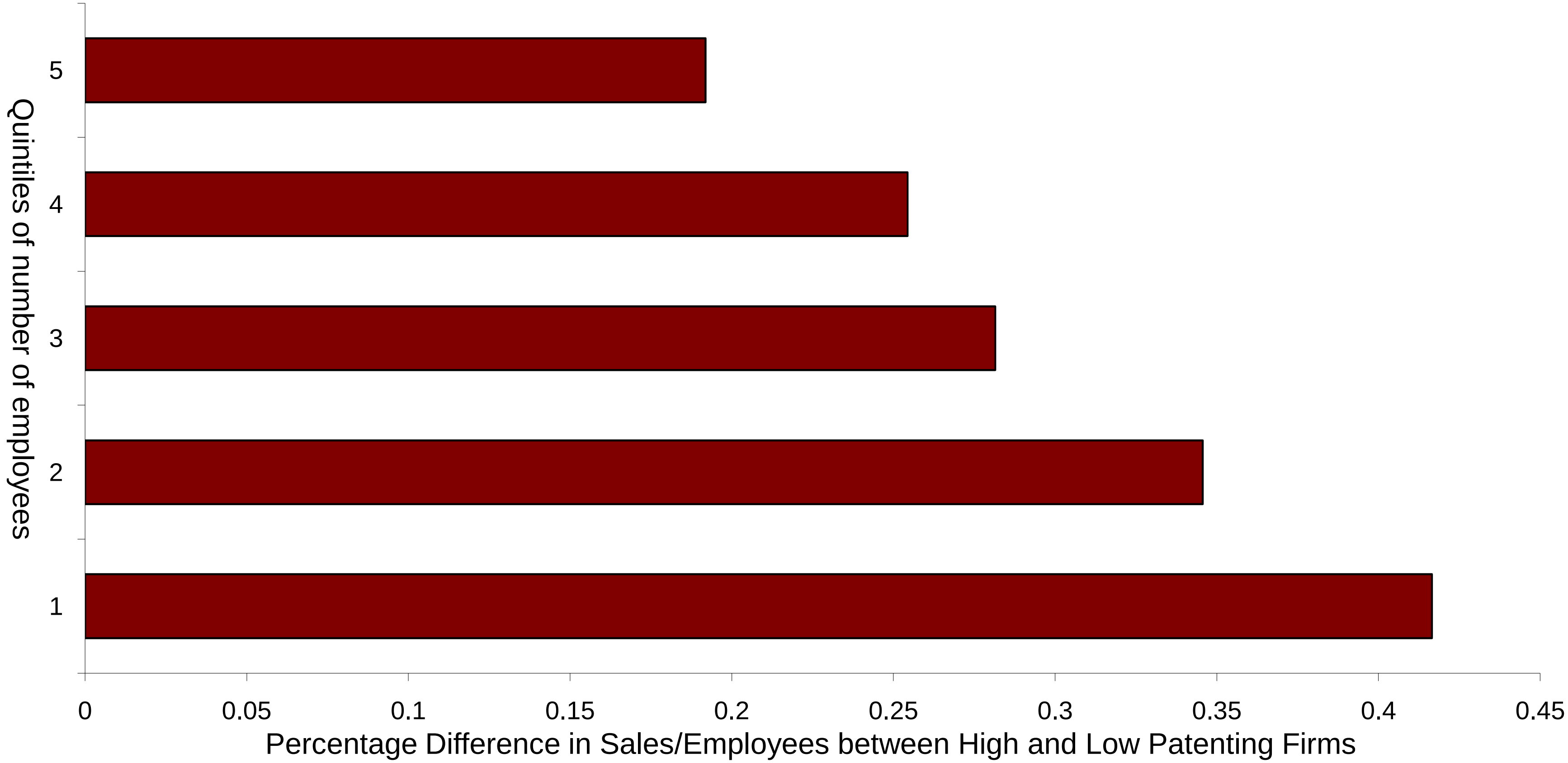
TABLE 2-

COMPARISON OF KEY VARIABLES: PATENTING FIRMS VERSUS PUBLISHING FIRMS

					Distribution		
PANEL A: PATENTING							
Variable	# firms	# Obs	Mean	Std. Dev.	10^{st}	50^{th}	90^{th}
Publications Stock	1,613	8,259	7.9	49.4	0	0.9	12.0
Patents stock	13,087	58,445	6.90	51.50	0	1	10.90
Sales (`000)	9,997	44,384	451,209	3,729,158	1,329	24,890	409,483
Employess	12,558	55,884	1,495	11,353	10	150	1,600
Age	12,245	55,803	29	26	5	21	67
Employment Growth	12,558	55,884	0.02	0.44	-0.17	0	0.21
Sales Growth	9,606	42,801	0.09	0.51	-0.06	0.08	0.40
Capital (`000)	10,428	46,069	293,528	3,580,681	185	4,983	162,882
Cash flow (`000)	9,980	43,205	46,318	496,709	-665	1,233	33,938
Capital/Employee (`000)	9,759	43,769	526	21,309	8	40	201
Sales/Employee (`000)	9,283	41,823	247	272	79	175	449
PANEL B: PUBLISHING							
					Distribution		
Variable	# firms	# Obs	Mean	Std. Dev.	10^{st}	50^{th}	90^{th}
Publications Stock	6,516	27,657	3.90	28.70	0	1	6
Patents stock	1,613	27,657	3.90	28.70	0.00	0.52	5.70
Sales (`000)	5,730	23,995	810,821	5,952,074	695	23,991	776,867
Employess	6,061	26,340	2,414	15,796	5	113	2,211
Age	5,965	26,103	27	27	4	17	67
Employment Growth	6,061	26,340	0.03	0.44	-0.41	0	0.26
Sales Growth	5,571	23,437	0.10	0.56	-0.05	0.09	0.44
Capital (`000)	6,065	25,657	463,153	3,700,162	47	3,787	279,699
Cash flow (`000)	5,702	23,673	76,970	661,374	-580	950	56,280
Capital/Employee (`000)	5,589	24,416	695	38,482	4	39	273
Sales/Employee (`000)	5,229	22,678	299	364	67	191	608

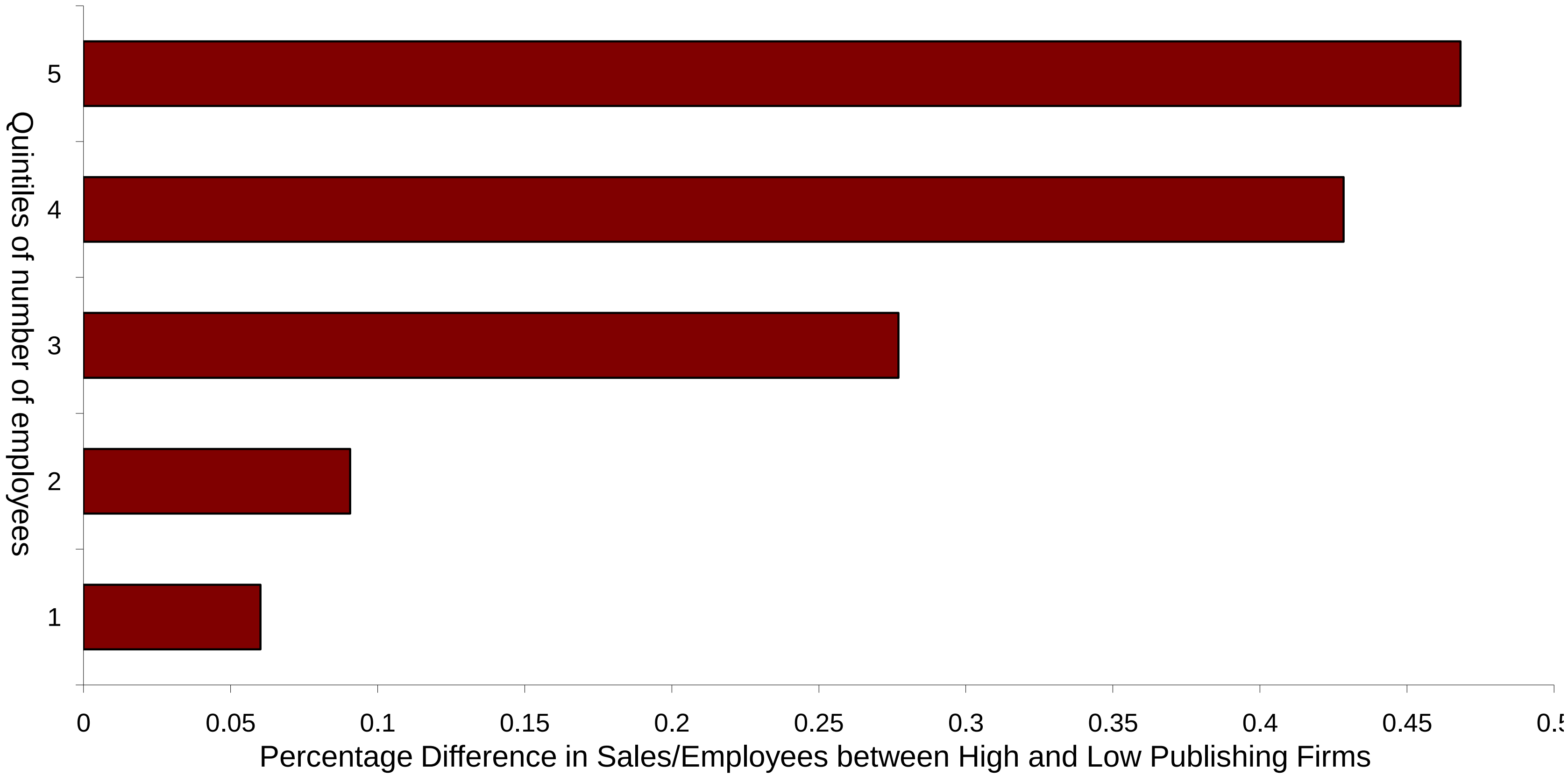
Notes: These tables provide summary statistics for firms in our estimation sample over the period 1995-2004. Panel A includes only firms that have at least one patent in the period 1978-2004 and panel B includes all firms that have at least one academic publication in the same period. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic 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). 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.

FIGURE 2: PERCENTAGE DIFFERENCE IN LABOR PRODUCTIVITY BETWEEN HIGH AND LOW PATENTING FIRMS ACROSS FIRM SIZE CLASSES



Notes: This figure plots percentage differences in labor productivity (sales per employee) between high and low publishing firms across quintiles of number of employees. A firm is assumed to be high (low) patenting if its number of patents is higher (lower) than the sample median number of patents. The sample includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. 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: PERCENTAGE DIFFERENCE IN LABOR PRODUCTIVITY BETWEEN PUBLISHING AND NON-PUBLISHING FIRMS ACROSS FIRM SIZE CLASSES



Notes: This figure plots percentage differences in labor productivity (sales per employee) between high and low publishing firms and across quintiles of number of employees. A firm is assumed to be high (low) publishing if its number of publications is higher (lower) than the median number of publications. The sample includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database.

TABLE 3-

PRODUCTIVITY-PATENTS RELATIONSHIP AND FIRM SIZE

DEPENDENT VARIABLE: LOG(SALES)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Firms:	*Patenting and Publishing*								*Excluding non-patenting*			
# Employees $_{t-1}$:	*All*	*All*	*All*	*All*	*≤ 25th (25)*	*≤ median (110)*	> median (110)	> 75th (435)	*All*	*All*	*≤ median (135)*	> median (135)
$\log(\text{Employment})_{t-1}$	0.806***	0.816***	0.345***	0.357***	0.857***	0.837***	0.740***	0.676***	0.783***	0.795***	0.802***	0.694***
	(0.009)	(0.009)	(0.017)	(0.017)	(0.021)	(0.014)	(0.018)	(0.028)	(0.011)	(0.011)	(0.016)	(0.018)
$\log(\text{Capital})_{t-1}$	0.186***	0.184***	0.107***	0.106***	0.166***	0.157***	0.230***	0.280***	0.197***	0.495***	0.163***	0.267***
	(0.007)	(0.007)	(0.009)	(0.009)	(0.011)	(0.009)	(0.013)	(0.021)	(0.008)	(0.008)	(0.011)	(0.013)
$\log(\text{Patents stock})_{t-1}$	0.055***	0.147***	0.073***	0.177***	0.126***	0.082***	0.051***	0.038***	0.065***	0.141***	0.096***	0.056***
	(0.009)	(0.024)	(0.014)	(0.048)	(0.044)	(0.025)	(0.009)	(0.010)	(0.009)	(0.025)	(0.023)	(0.009)
$\log(\text{Patents stock})_{t-1} \times \log(\text{Employment})_{t-1}$		-0.012***		-0.016***						-0.011***		
		(0.003)		(0.007)						(0.003)		
Country dummies (12)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Three-digit SIC dummies (183)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed-effects	No	No	Yes	Yes	No	No	No	No	No	No	No	No
R^2	0.881	0.891	0.976	0.976	0.562	0.703	0.856	0.849	0.887	0.899	0.720	0.876
Observations	53,842	53,842	53,842	53,842	13,380	27,009	26,833	13,428	38,000	38,000	19,030	18,970
Number of firms	12,326	12,326	12,326	12,326	4,543	7,624	5,633	2,817	8,284	8,284	5,057	3,920

Notes: This table reports the results of OLS regressions examining the relationship between productivity, patents and firm size. The sample covers the period 1995-2004 and includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic 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). Capital is defined as fixed-assets. Patents and publication stocks are computed using the perpetual inventory method using a depreciation rate of 15 percent. 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 4-

PRODUCTIVITY-INNOVATION RELATIONSHIP AND FIRM SIZE

DEPENDENT VARIABLE: LOG(SALES)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
# Employees $_{t-1}$:	*All*	*All*	*All*	*All*	*≤ median (110)*	*> median (110)*	> 75^{th} (435)	> 90^{th} (1600)
log(Employment)$_{t-1}$	0.806***	0.809***	0.816***	0.813***	0.837***	0.731***	0.667***	0.626***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.014)	(0.016)	(0.027)	(0.045)
log(Capital)$_{t-1}$	0.186***	0.189***	0.184***	0.183***	0.158***	0.237***	0.276***	0.277***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.012)	(0.021)	(0.038)
log(Patents stock)$_{t-1}$	0.052***		0.153***	0.168***	0.085***	0.044***	0.029***	0.029**
	(0.009)		(0.025)	(0.025)	(0.025)	(0.009)	(0.009)	(0.013)
log(Publications stock)$_{t-1}$	0.030**	0.041***	0.041***	-0.037	-0.016	0.048***	0.055***	0.083***
	(0.015)	(0.015)	(0.015)	(0.036)	(0.034)	(0.015)	(0.016)	(0.018)
log(Patents stock)$_{t-1}$ × log(Employment)$_{t-1}$			-0.014***	-0.016***				
			(0.003)	(0.003)				
log(Publications stock)$_{t-1}$ × log(Employment)$_{t-1}$				0.012***				
				(0.004)				
Country dummies (12)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Three-digit SIC dummies (1	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.891	0.891	0.891	0.891	0.703	0.860	0.851	0.846
Observations	53,842	53,842	53,842	53,842	27,009	26,833	13,428	5,491
Number of firms	12,326	12,326	12,326	12,326	7,624	5,633	2,817	1,163

Notes: This table reports the results of OLS regressions examining the relationship between productivity, innovation and firm size. The sample covers the period 1995-2004 and includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic 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). 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. 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-

GROWTH-INNOVATION RELATIONSHIP AND FIRM SIZE

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:	*log(Employment$_t$/Employment$_{t-1}$)*			*log(Sales$_t$/Sales$_{t-1}$)*		
log(Employment)$_{t-1}$	-0.044***	-0.040***	-0.041***	-0.024***	-0.021***	-0.022***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
log(Patents stock)$_{t-1}$	0.032***	0.078***	0.081***	0.022***	0.060***	0.064***
	(0.003)	(0.011)	(0.012)	(0.003)	(0.010)	(0.011)
log(Publications stock)$_{t-1}$	0.021***	0.026***	0.008	0.009*	0.013**	-0.013
	(0.005)	(0.005)	(0.013)	(0.005)	(0.005)	(0.013)
log(Patents stock)$_{t-1}$ × log(Employment)$_{t-1}$		-0.007***	-0.007***		-0.005***	-0.006***
		(0.001)	(0.001)		(0.001)	(0.001)
log(Publications stock)$_{t-1}$ × log(Employment)$_{t-1}$			0.003**			0.004***
			(0.001)			(0.001)
Country dummies (12)	Yes	Yes	Yes	Yes	Yes	Yes
Three-digit SIC dummies (183)	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.043	0.045	0.045	0.037	0.038	0.039
Observations	63,164	63,164	63,164	57,779	57,779	57,779
Number of firms	14,195	14,195	14,195	13,354	13,354	13,354

Notes: This table reports the results of OLS regressions examining the how the effects of innovation on growth vary with firm size. The sample covers the period 1995-2004 and includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database. Patents and publication stocks are computed using the perpetual inventory method using a depreciation rate of 15 percent. 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-

PRODUCTIVITY-INNOVATION RELATIONSHIP AND FIRM SIZE: US COMPUSTAT FIRMS VERSUS LARGE EUROPEAN FIRMS

DEPENDENT VARIABLE: LOG(SALES)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Firms:		*US Compustat firms*			*Large Europeam firms (>1600 Employees)*			*Large Europeam firms (>1000 Employees)*	
log(Employment)$_{t-1}$	0.676***	0.674***	0.671***	0.607***	0.621***	0.622***	0.647***	0.666***	0.667***
	(0.028)	(0.029)	(0.029)	(0.050)	(0.054)	(0.054)	(0.038)	(0.041)	(0.042)
log(Capital)$_{t-1}$	0.314***	0.314***	0.315***	0.289***	0.289***	0.289***	0.282***	0.281***	0.281***
	(0.025)	(0.025)	(0.025)	(0.042)	(0.042)	(0.042)	(0.031)	(0.031)	(0.031)
log(Patents stock)$_{t-1}$	-0.010	-0.022	0.008	0.030**	0.087	0.084	0.025**	0.106**	0.103**
	(0.007)	(0.024)	(0.031)	(0.014)	(0.058)	(0.063)	(0.011)	(0.045)	(0.046)
log(Publications stock)$_{t-1}$	0.022***	0.021**	-0.043	0.081***	0.083***	0.096	0.070***	0.074***	0.085
	(0.008)	(0.009)	(0.049)	(0.019)	(0.019)	(0.100)	(0.017)	(0.018)	(0.079)
log(Patents stock)$_{t-1}$ × log(Employment)$_{t-1}$		0.001	-0.002		-0.006	-0.007		-0.009*	-0.009*
		(0.002)	(0.003)		(0.006)	(0.006)		(0.005)	(0.005)
log(Publications stock)$_{t-1}$ × log(Employment)$_{t-1}$			0.007			-0.001			-0.001
			(0.005)			(0.010)			(0.008)
Country dummies (12)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Three-digit SIC dummies (183)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.951	0.951	0.951	0.854	0.854	0.854	0.847	0.847	0.847
Observations	21,416	21,416	21,416	4,793	4,793	4,793	7,706	7,706	7,706
Number of firms	1,502	1,502	1,502	1,020	1,020	1,020	1,615	1,615	1,615

Notes: This table reports the results of OLS regressions examining the relationship between productivity, innovation and firm size for US Compustat firms and for large European firms. The sample covers the period 1995-2004 for European firms and includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. For the US, the sample covers the period 1980-2001 and includes only firms with at least one patent or publication over the period 1969-2005. The sample included in columns 4-6 matches the average number of employees for the sample of Compustat firms. For Europe, patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic 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). Patents and publication stocks are computed using the perpetual inventory method using a depreciation rate of 15 percent. 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 EFFECTS OF FIRM SIZE ON INNOVATION INTENSITY

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Depedent variable	*Patents stock / Sales*				*Publications stock / Sales*			
Firms:	*All*		*Only patenting*		*All*		*Only publishing*	
Employees$_{t-1}$ ('000)	-0.894***	-2.722***	-1.161***	-4.417***	-0.885*	-2.854***	-1.290*	-4.615***
	(0.244)	(0.654)	(0.358)	(1.074)	(0.483)	(1.045)	(0.678)	(1.586)
Employees$^2_{t-1}$ ('000)		0.008***		0.013***		0.009***		0.014***
		(0.002)		(0.004)		(0.003)		(0.004)
Country dummies (12)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Three-digit SIC dummies (183)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.079	0.079	0.084	0.115	0.046	0.046	0.071	0.072
Observations	10,415	10,415	7,367	7,367	10,415	10,415	4,208	4,208

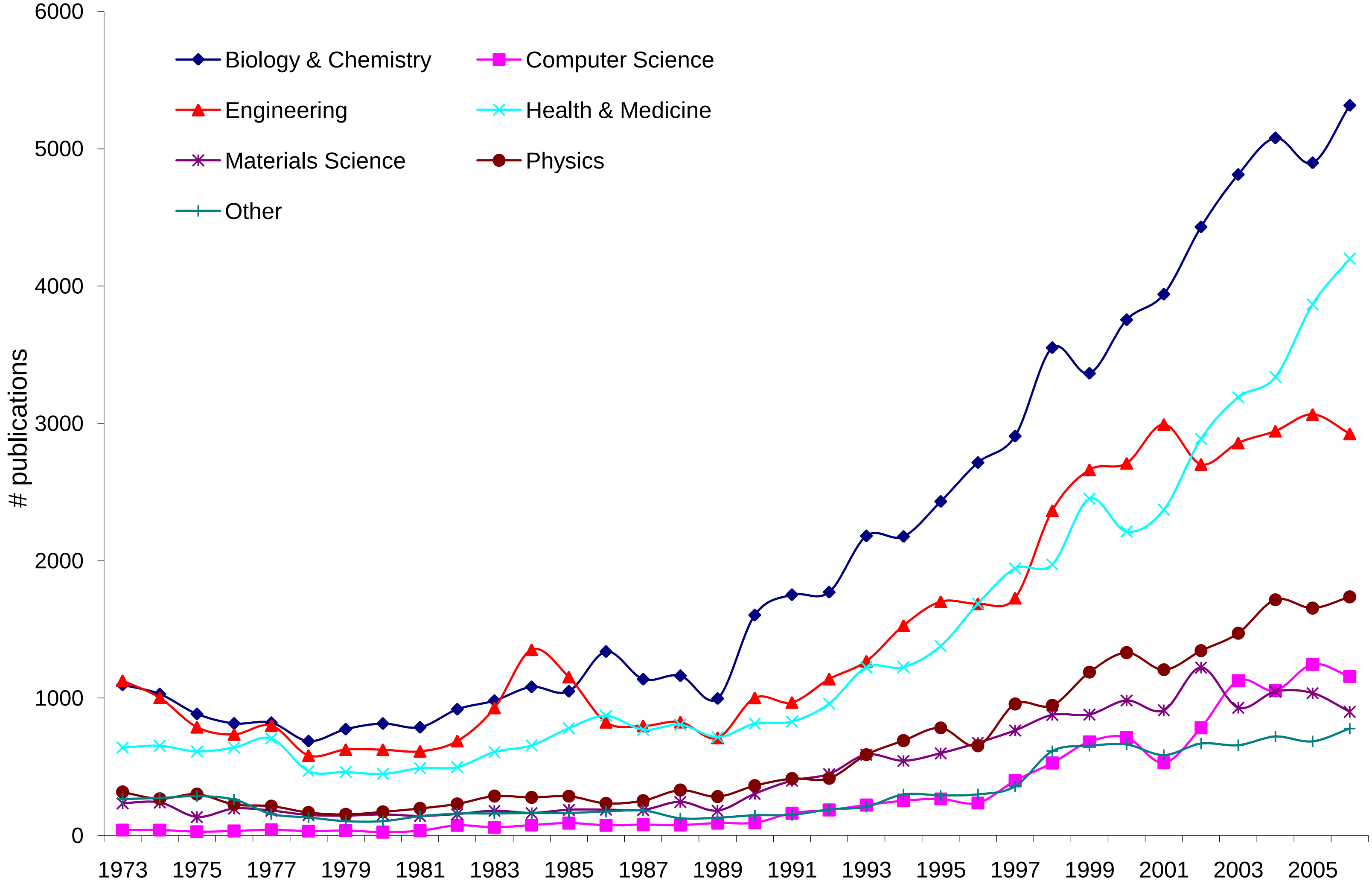
Notes: This table reports the results of OLS regressions that examine the relation between firm size and innovation intensity. The sample is cross-sectional for 2003 (which has most firms) and includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic 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). Patents and publication stocks are computed using the perpetual inventory method using a depreciation rate of 15 percent. 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-

SELECTION INTO INNOVATION						
DEPENDENT VARIABLE: LOG(SALES)						
	(1)	(2)	(3)	(4)	(5)	(6)
# Employees $_{t-1}$:	*All*	*All*	*≤ median (150)*	> median (150)	> 75^{th} (265)	>90^{th} (1500)
$\log(\text{Employment})_{t-1}$	0.745***	0.739***	0.725***	0.729***	0.723***	0.649***
	(0.007)	(0.008)	(0.019)	(0.011)	(0.018)	(0.048)
$\log(\text{Capital})_{t-1}$	0.220***	0.220***	0.196***	0.242***	0.236***	0.246***
	(0.005)	(0.005)	(0.001)	(0.007)	(0.011)	(0.036)
Dummy for innovating	0.060***	-0.004	0.049***	0.056***	0.173***	0.189***
	(0.012)	(0.044)	(0.018)	(0.016)	(0.031)	(0.066)
Dummy for innovating × $\log(\text{Employment})_{t-1}$		0.012				
		(0.007)				
Country dummies (12)	Yes	Yes	Yes	Yes	Yes	Yes
Three-digit SIC dummies (183)	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.837	0.837	0.492	0.835	0.801	0.807
Observations	90,449	90,449	46,202	44,247	16,347	3,400
Number of firms	19,822	19,822	12,435	9,867	3,835	813

Notes: This table reports the results of OLS regressions examining how the effect of innovating on productivity varies with firm size. The sample covers the period 1995-2004 and includes two samples of firms: all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005 and a random sample of 10 percent of all non-innovating firms (firms that were not matched to the patents and publications databases). Dummy for innovating receives the value of one for observations where the firm ever innovates, and zero for observations where the firm never innovates. 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: 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.

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 publication in "hard" science journals (where 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.

Table A2

SUMMARY STATISTICS FOR FIRM PUBLICATIONS ACROSS MAIN FIELDS: LARGE FIRMS VERSUS SMALL FIRMS

PANEL A: PUBLICATIONS BY LARGE FIRMS (# EMPLOYEES>110, MEDIAN)

Main Field	Publications	Mean of Citations Received	Mean of Citations Made	Mean of Impact Rate of Journal	Average number of Authors
All fields	67,395	7.89	17.22	2.05	3.77
Biology & Chemistry	24,027	12.06	22.68	2.93	4.30
Engineering	16,962	3.02	10.59	0.79	2.97
Health & Medicine	12,661	9.06	16.88	2.84	4.37
Physics	3,748	7.25	17.80	1.66	3.87
Materials Science	5,795	5.22	13.80	1.00	3.21
Other	1,909	7.34	20.73	1.50	3.08
Computer Science	2,293	2.12	15.66	0.74	2.70

PANEL B: PUBLICATIONS BY SMALL FIRMS (# EMPLOYEES ≤ 110, MEDIAN)

Main Field	Publications	Mean of Citations Received	Mean of Citations Made	Mean of Impact Rate of Journal	Average number of Authors
All fields	68,187	6.28	17.07	2.27	4.13
Biology & Chemistry	20,495	8.99	22.03	3.37	4.79
Engineering	14,802	3.16	10.93	0.83	2.93
Health & Medicine	13,457	6.46	14.03	3.24	5.18
Physics	7,810	7.90	20.99	2.21	4.20
Materials Science	4,443	4.72	14.06	1.02	3.78
Other	3,053	5.59	20.00	1.38	3.01
Computer Science	4,127	2.50	18.10	0.67	2.81

Notes: This table reports summary statistics for firm publications, disaggregated according to whether the publishing firm is large or small. A firm is said to be large (small) if its number of employees is greater (lower) than 110, the sample median number of employees. The sample covers the period 1978-2005 and includes all Amadeus firms with at least publication in "hard" science journals. Only firms that report financials are included. 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 LARGE FIRMS (# EMPLOYEES>110, MEDIAN)

Main Field	Publications	Mean of Citations Received	Mean of Citations Made
All fields	23,272	11.53	17.64
Biology & Chemistry	10,429	15.99	24.01
Engineering	360	2.71	11.93
Health & Medicine	3,551	2.18	6.64
Physics	6,192	11.23	14.58
Materials Science	1,264	8.39	15.59
Other	1,067	8.38	15.45
Computer Science	409	9.29	14.28

PANEL B: PUBLICATIONS BY SMALL FIRMS (# EMPLOYEES ≤ 110, MEDIAN)

Main Field	Publications	Mean of Citations Received	Mean of Citations Made
All fields	25,560	8.75	16.29
Biology & Chemistry	10,061	11.19	20.37
Engineering	405	3.42	17.16
Health & Medicine	3,170	3.37	9.31
Physics	7,560	7.14	10.66
Materials Science	681	7.56	17.69
Other	3,084	11.72	23.48
Computer Science	599	5.99	16.59

Notes: This table reports summary statistics for firm publications in leading academic journals. We include only journals in the top quartile of the JCR impact factor. The sample covers the period 1978-2005 and includes all Amadeus firms with at least publication in "hard" science journals. Only firms that report financials are included. Firm publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database.

TABLE A4-

PRODUCTIVITY-INNOVATION RELATIONSHIP AND FIRM SIZE. MORE THAN 50 EMPLOYEEES

DEPENDENT VARIABLE: LOG(SALES)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# Employees $_{t-1}$:	*All*	*All*	*All*	*≤ median (280)*	*> median (280)*	> 75^{th} (435)	> 90^{th} (1600)
log(Employment)$_{t-1}$	0.763***	0.776***	0.772***	0.802***	0.699***	0.647***	0.619***
	(0.013)	(0.014)	(0.014)	(0.023)	(0.023)	(0.037)	(0.059)
log(Capital)$_{t-1}$	0.209***	0.208***	0.208***	0.158***	0.261***	0.286***	0.283***
	(0.010)	(0.010)	(0.009)	(0.013)	(0.017)	(0.030)	(0.051)
log(Patents stock)$_{t-1}$	0.049***	0.133***	0.156***	0.100***	0.029***	0.024**	0.023
	(0.009)	(0.026)	(0.027)	(0.021)	(0.009)	(0.011)	(0.015)
log(Publications stock)$_{t-1}$	0.042***	0.050***	-0.067	-0.011	0.058***	0.067***	0.086***
	(0.015)	(0.015)	(0.049)	(0.035)	(0.016)	(0.017)	(0.021)
log(Patents stock)$_{t-1}$ × log(Employment)$_{t-1}$		-0.011***	-0.014***				
		(0.003)	(0.003)				
log(Publications stock)$_{t-1}$ × log(Employment)$_{t-1}$			0.015***				
			(0.005)				
Country dummies (12)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Three-digit SIC dummies (183)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.863	0.864	0.864	0.483	0.852	0.824	0.859
Observations	34,865	34,865	34,865	17,460	17,405	8,552	3,557
Number of firms	7,384	7,384	7,384	4,450	3,648	1,792	739

Notes: This table reports the results of an OLS regression examining the relationship between productivity, innovation and firm size for firms that have more than 50 employees. The sample covers the period 1995-2004 and includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic 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 25 million publications). 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. Standard errors (in brackets) are robust to arbitrary heteroskedasticity and allow for serial correlation through clustering by ultimate owner. * significant at 10%; ** significant at 5%; *** significant at 1%.

TABLE A5-

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 2005. Firm publications are constructed by matching the name of the firm to the address field in the complete ISI Web of Science database.

TABLE A6-

FIRM PUBLICATIONS IN TOP 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 A7-

PRODUCTIVITY-INNOVATION RELATIONSHIP AND FIRM SIZE: CITATIONS WEIGHTED PUBLICATIONS STOCK

DEPENDENT VARIABLE: LOG(SALES)

	(1)	(3)	(4)	(5)	(6)	(7)
# Employees $_{t-1}$:	*All*	*All*	*≤ median (110)*	*> median (94)*	> 75^{th} (350)	> 90^{th} (1200)
log(Employment)$_{t-1}$	0.806***	0.814***	0.084**	0.739***	0.684***	0.635***
	(0.009)	(0.009)	(0.014)	(0.016)	(0.024)	(0.042)
log(Capital)$_{t-1}$	0.186***	0.183***	0.158***	0.233***	0.269***	0.282***
	(0.007)	(0.007)	(0.009)	(0.012)	(0.018)	(0.034)
log(Patents stock)$_{t-1}$	0.054***	0.169***	0.089***	0.044***	0.033***	0.026***
	(0.009)	(0.025)	(0.026)	(0.009)	(0.009)	(0.012)
log(Publications stock)$_{t-1}$	0.008	-0.086***	-0.045	0.032**	0.046***	0.062***
	(0.013)	(0.033)	(0.029)	(0.014)	(0.015)	(0.017)
log(Patents stock)$_{t-1}$ × log(Employment)$_{t-1}$		-0.016***				
		(0.003)				
log(Publications stock)$_{t-1}$ × log(Employment)$_{t-1}$		0.016***				
		(0.004)				
Country dummies (12)	Yes	Yes	Yes	Yes	Yes	Yes
Three-digit SIC dummies (183)	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies (10)	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.881	0.891	0.694	0.859	0.851	0.842
Observations	51,344	51,344	27,009	24,335	12,401	4,919
Number of firms	11,420	11,420	7,624	5,352	2,718	1,119

Notes: This table reports the results of an OLS regression examining the relationship between productivity, innovation and firm size. Publications are weighted according to the number of forward citations they receive. The sample covers the period 1995-2004 and includes all Amadeus firms with at least one patent or one academic publication in "hard" science journals between 1978 and 2005. Patents are constructed by matching the name of all Amadeus firms to all EPO and USPTO patent records. Academic 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). Capital is defined as fixed-assets. Patents and publication stocks are computed using the perpetual inventory method using a depreciation rate of 15 percent. 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%.