NBER WORKING PAPER SERIES

THE DETERMINANTS OF FACULTY PATENTING BEHAVIOR: DEMOGRAPHICS OR OPPORTUNITIES?

Pierre Azoulay
Waverly Ding
Toby Stuart

Working Paper 11348
http://www.nber.org/papers/w11348

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2005

Send all correspondence to pa2009@columbia.edu. This paper was written for the NBER Conference "Academic Science and Entrepreneurship: Dual Engines of Growth?" Santa Fe, NM, April 1*st* 2005. We thank Scott Stern, Iain Cockburn, and Paula Stephan for many useful comments, as well as seminar audiences at the University of Toronto, Harvard Business School, and NBER. The usual disclaimer applies. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

The Determinants of Faculty Patenting Behavior: Demographics or Opportunities?
Pierre Azoulay, Waverly Ding, and Toby Stuart
NBER Working Paper No. 11348
May 2005
JEL No. O31, O32, O33

ABSTRACT

We examine the individual, contextual, and institutional determinants of faculty patenting behavior in a panel dataset spanning the careers of 3,884 academic life scientists. Using a combination of discrete time hazard rate models and fixed effects logistic models, we find that patenting events are preceded by a flurry of publications, even holding constant time-invariant scientific talent and the latent patentability of a scientist's research. Moreover, the magnitude of the effect of this flurry is influenced by context --- such as the presence of coauthors who patent and the patent stock of the scientist's university. Whereas previous research emphasized that academic patenters are more accomplished on average than their non-patenting counterparts, our findings suggest that patenting behavior is also a function of scientific opportunities. This result has important implications for the public policy debate surrounding academic patenting.

Pierre Azoulay
Columbia University
Graduate School of Business
3022 Broadway, Uris Hall 704
New York, NY 10027
and NBER
pa2009@columbia.edu

Waverly Ding
University of California
Haas School of Business
545 Student Services #1900
Berkeley, CA 94720
wding@haas.berkeley.edu

Toby Stuart
Columbia University
Graduate School of Business
3022 Broadway, Uris Hall 710
New York, NY 10027
toby.stuart@columbia.edu

1 Introduction

In the past few decades, universities and other public-sector research organizations have become more proactive in their efforts to commercialize scientific discoveries (e.g., Jaffe and Lerner, 2001; Jensen and Thursby, 2001; Thursby and Thursby, 2002). This change has spawned a growing academic literature on university technology transfer, one stream of which has assessed trends in university patenting and the spillover of university science into the private sector (Jaffe, 1989; Mansfield, 1995; Henderson et al., 1998). Underlying the well documented upswing in university patenting has been a sharp increase in the number of individual academic scientists who are listed as inventors on patents. In this paper, we examine the individual, contextual, and institutional determinants of academic patenting in a panel dataset of 3,884 academic life scientists.

Past research on this topic has emphasized three related aspects of faculty patenting behavior. First, academic patenters are disproportionately recruited from the ranks of elite scientists and institutions (Zucker et al., 1998; Azoulay et al., 2005b). Second, there are important differences in the propensity to patent across fields, and in the motivations underlying patenting activity, most notably between the life and physical sciences/engineering (Owen-Smith and Powell, 2001). Finally, institutional context exerts a strong influence on the propensity to patent, either in the form of well-funded technology licensing offices, or through the presence of prominent peers who themselves are engaged in this activity (Di Gregorio and Shane, 2003). Most of this evidence, however, stems from analyses of survey data or from qualitative accounts. While consistent with these previous findings, the results in this paper qualify them in some important respects. Our study also generates a novel set of results, underscoring the benefits of fine-grained longitudinal data at the researcher level of analysis.

The paper's findings include the following. First, we estimate pronounced life-cycle effects on the propensity to patent, with mid-career academics being much more likely to patent than younger and older faculty members. Second, we establish a relationship between the latent patentability of a faculty's research and his/her propensity to patent. While latent patentability is often thought to be unobservable, we compute a patentability score for each scientist in our sample by using keywords in the publications of scientists that have already applied for patent rights as a benchmark for patentable research, and then comparing the research of each scientist in our dataset to this benchmark. Although there is noise in this proxy, it nevertheless quite strongly predicts the likelihood of a patenting event.

Third, we document that patenting is often accompanied by a flurry of publication activity in the year preceding the patent application, even after accounting for the lagged stock of publications (in hazard rate models) or controlling for scientist fixed effects. This result highlights the fact that academic patenting, rather than merely reflecting the influence of time-invariant demographic factors, also responds to variation in scientific opportunities. Holding life-time scientific achievement constant, we find that surges of scientific productivity, not steady performance, is most likely to be associated with a patent. Moreover, the magnitude of the effect of this flurry decreases with the presence of a patenting coauthor, or with the intensity of patenting activity in the scientist's university. These findings suggest that institutional and contextual factors may partially substitute for scientific opportunities in determining individual rates of patenting.

Lastly, independent of any specific finding, the general analysis herein is relevant to the broader question of the impact of patenting on the development of academic science. Surveys of university faculty have found rampant concern that patenting is skewing research agendas toward commercial priorities, causing delay in the public dissemination of research findings, and crowding out effort devoted to the pursuit of fundamental knowledge (Blumenthal et al., 1996; Campbell et al., 2002; Krimsky, 2003). Insofar as our results relate to this issue, the finding that patenting follows a flurry of publications suggests to us that the crowd-out hypothesis is unlikely to hold true.[1] Although we cannot adjudicate between opposing claims regarding the *effect* of patenting on individual-level or university-level outcomes in the present study, one can construe our results as providing the "first stage" of an econometric analysis of the effect of academic patenting on the rate and direction of scientific progress, an evaluation we are pursuing in other research (Azoulay et al., 2005a).

[1] However, if scientific trajectories associated with patents exhaust themselves more quickly than those remaining free of associations with the world of commerce, then intertemporal substitution of "basic," fundamental knowledge by "applied," patentable output could still be consistent with the patterns we observe in the data.

The rest of the paper proceeds as follows. In the next section, we situate our contribution in the large and growing literature on academic patenting, and highlight what we regard as outstanding issues that can only be resolved with researcher-level longitudinal data of the kind we analyze. Section 3 describes data sources and the construction of the sample and discusses our econometric approach. Section 4 presents descriptive statistics and reports our results. Section 5 concludes.

2 Who Patents?

In recent times, the region of overlap between the spheres of academic science and commercial markets has experienced significant growth. The expanding interface between these two domains raises myriad questions, ranging from the amount of near-term economic value created by the spillovers of university research, to the emergence of select universities as engines of entrepreneurial activity, to the influence of opportunities to commercialize scientific research on the traditional incentive systems that have governed academic science. Researchers have engaged a variety of these questions, and advancement in our understanding is occurring along many fronts.

Spurred in part by accessible data, many studies have assessed the role of universities as direct sources of commercial innovations, primarily considering the quality and quantity of their innovative outputs. For instance, Henderson et al. (1998) examine the relative importance of university patents, finding that there has been a secular decline in the positive quality gap separating university patents from those assigned to for-profit firms. Mowery et al. (2001) have investigated the consequence of the policy changes brought about by the Bayh-Dole Act. They challenge the conventional wisdom that Bayh Dole has accelerated universities' production of patents, showing that the legislation was not a primary factor in explaining the uptick in patenting at three prominent universities. At the level of the university, Thursby and Thursby (2002) find that university administrators have become more proactive in pursuing patents and licensing opportunities. Di Gregorio and Shane (2003) explore cross-university differences in the formation of start up companies, discovering that intellectual eminence is a central factor distinguishing the universities that spawn start up companies.

The majority of the archival work that has looked at the commercial outputs of not-for-profit organizations has treated the university as the level of analysis. Because the preponderance of the empirical studies have been performed at the university level (notable exceptions include Murray, 2002; Agrawal and Henderson, 2002; and Stephan et al. 2004), less is known about the factors that underlie individual scientists' participation in patenting. In this article, we analyze the probability of patenting in a large, longitudinal sample of university faculty in the biomedical area. Our analysis is guided by an interest in four issues. First, how does the proclivity to patent vary with scientists' experience in the profession? Second, what is the relationship between scientific productivity (measured as papers published) and patenting? Third, are there significant differences across research areas within scientific disciplines in terms of the apparent "patentability" of the work, and is there any evidence to suggest that scientists may be altering their research to move toward patentable research? Fourth, to what extent is the propensity to patent sensitive to the work context of the individual scientist, particularly the level of commercial orientation of a scientist's university and his or her coauthors?

Treating each of these in turn, we first ask, how does the propensity to patent change over the scientific career? Economists and sociologists alike have a long-standing interest in career dynamics in academe, in part because incentives in science vary over the professional life cycle. Two elements of the institutionalized reward system in science are generally thought to be tenure-invariant: the tying of peer recognition to priority in research discovery, and the intrinsic satisfaction garnered from solving vexing problems. However, monetary incentives to produce research output in science do depend on the career stage, and it is well known that the wage-tenure profile in academic science is not steep (Stephan, 1996). Given the shallow slope of post-tenure salary increases, Levin and Stephan (1991) suggest that levels of investment in research should vary over the career life cycle. In particular, senior scientists with tenured appointments may reallocate some of their effort to consulting and other extra-university income generating opportunities. Therefore, if the widely held assumptions about changing incentives over the career do in fact hold, we should observe that the rate of patenting accelerates in the post-tenure interval.

A countervailing possibility is implied by a growing body of ethnographic research that portrays the increasing acceptance of patenting as a legitimate activity in academic science (Etzkowitz, 1998). If the pendulum has swung to the point that patenting is perceived to contribute to scientists' reputation and influence, we would expect to observe that, viewing successive cohorts of scientists, patenting occurs with increasing frequency in the early career stage. Consistent with this perspective, Owen-Smith and Powell (2001) describe interviews with scientists that have come to view patents as reaffirmations of the originality of their work and as contributing to their scientific visibility. Recent interview-based accounts thus raise the possibility of a significant shift in the norms and reward system in science, with implications for life-cycle effects in patenting.

Next, we seek to identify the relationship between scientists' productivity and the likelihood that they patent. Existing evidence suggests that the scientists with the most stellar academic credentials are also the most likely to be involved in commercial endeavors. In particular, Zucker et al. (1998) describe the importance of the geographic location of star scientists in the emergence of the biotechnology industry. They argue that the direct participation of leading academic scientists in early stage biotechnology companies was so important that the locations of star scientists served as geographical constraints on the development of the industry. Stuart and Ding (2005) directly analyze the probability that academic scientists either found or join scientific advisory boards of biotechnology firms. They find that standard measures of human capital strongly associate with the participation of scientists in entrepreneurial initiatives.

The existing literature thus provides reason to expect that patenting is concentrated among the group of eminent scientists. Yet, beyond the general association between research output and the likelihood of engaging in market-related activities, identifying more precisely the relationship between the production of papers and patents may adjudicate among the competing mechanisms that could generate the relationship. In particular, if the magnitude of the *stock* of scientists' research output predicts the onset of patenting, it is likely that faculty members' scientific reputations are important considerations in the decision to patent. If this proves to be the case, a plausible explanation is that the prominence of the inventor on a patented technology may influence the university's ability to capitalize on the intellectual property by affecting the probability that potential licensees become aware of and interested in the technology.

Consider instead the implication of a positive relationship between the *flow*, but not the stock, of scientists' research output and the probability that a patent is issued. If the flow of output is the determining factor, we would suggest that technological "opportunity" looms large in the transition to patenting. A flurry of scientific output — a high flow of publications — occurs when a scientist unearths a productive domain of research. If patenting is a byproduct of a surge in productivity, we think it reasonable to conclude that a patent is often an opportunistic response to the discovery of a promising research area.

The third issue we consider is how the specific areas of expertise of academic scientists affect the likelihood of patenting. Obviously, there exists heterogeneity across scientists in the potential commercial value of the research they produce. If one needs to account for such differences, it is tempting to argue that the analyst can accommodate them by incorporating scientist fixed effects in the analysis. We believe, however, that this represents just a partial solution given the volume and the diversity of research projects that scientists participate in throughout their careers. We therefore attempt to develop a direct measure of the "patentability" of scientific research. The intuition behind the measure is that knowledge of the research foci of academic scientists who have already patented can be used to identify the domains of science in which research is patentable. With such a measure in hand, we ask two questions. First, does the latent patentability of scientists' research in fact affect the probability of patenting? Second, is it the patentability of the stock or the flow of research outputs that most consequentially influences the propensity to patent?

Fourth, we explore two elements of scientists' work contexts. While it is well established that propensities to patent vary substantially across universities, we do not have a clear sense for the influence of organizational characteristics on the patenting rates of otherwise similar scientists within different universities. Numerous studies suggest that the decision to engage in commercial activity of all sorts is strongly influenced by factors ranging from the norms and culture of an institution vis-à-vis commercial activity, to the quality of the university's technology transfer office (Thursby and Thursby 2002; Owen-Smith and Powell, 2001). Two prevalent considerations are thus the (potentially endogenous) role of a smooth functioning technology transfer office in encouraging faculty to disclose possibly patentable research findings, and more generally, a pro-commercialization "entrepreneurial culture" at a university. In our analysis, we ask whether university-level variables influence the patent rate net of controls for many observable individual-level characteristics.

A related question concerns the influence of proximate colleagues on the patent proclivities of individual scientists. There are a set of reasons to expect that scientists who work closely with commercially-inclined peers will be more likely to pursue commercial applications of their scientific research. Stuart and Ding (2005) argue that there are two mechanisms through which colleagues affect the probability that a particular scientist engages in commercial activities. First, peers exert attitudinal influences, in particular shaping the degree to which a given scientist is likely to embrace patenting as both a legitimate undertaking for an academic scientist and as a potential contributor to his or her professional standing. Second, peers convey information that may lower the cost of patenting, such as contacts in the technology transfer office and advice about how to minimize the amount of time consumed in patenting. We thus look for what might be labeled as "peer effects" on the transition to patenting. Specifically, we examine whether scientists who have co-authorship links with patent holders, or with researchers employed in the private sector, are themselves more likely to patent.

A necessary caveat pertains to the thorny issue of causality. Many of our independent variables, such as publications or latent patentability, could be considered outcomes of interest. Moreover, it would be incorrect to interpret our findings as providing evidence, *inter alia*, that publications and patents are complements, or that latent patentability "causes" patent applications. Rather, we have identified correlates of patenting. The conditional correlations we estimate can still be useful insofar as they help narrow the range of plausible theories regarding the *effect* of academic patenting on scientific productivity. In addition, since our most interesting results pertain to what are in fact lagged dependent variables, the study highlights the need to use correct econometric methodologies to recover causal effects. This is pursued in a companion paper (Azoulay et al., 2005a).

3 Data, Sample Characteristics, and Econometric Approach

We examine the determinants of faculty patenting behavior in a panel dataset of academic life scientists employed at universities and non-profit research institutes. This area was chosen because the biomedical fields have accounted for the preponderance of university patenting and licensing activity (Mowery et al., 2001). While we have not selected scientists because they have patented, we have sampled from scientific disciplines that we know to have significantly contributed to a vibrant area of technological development. We began by drawing 12,000 doctoral degree recipients from UMI Proquest Digital Dissertations, which lists Ph.D. recipients from more than one thousand universities. In forming the sample, we randomly selected individuals, but only those with Ph.D.s in scientific disciplines that have informed commercial biotechnology.[2] This assures a random sample of Ph.D.s in areas in which academic research may have significant, short-term commercial value.

Next, we obtained scientists' publication records from the ISI's *Web of Science* database. Because the Web of Science includes authors' affiliations, we were able to identify Ph.D. graduates who pursued careers outside of academe. After removing individuals that (i) had no publications in any post-graduate year, (ii) published exclusively under corporate affiliations, or (iii) exited academe early in their careers,[3] we were left with 3,884 scientists, all of whom we know to have been employed at research institutions. Each scientist is

[2]To identify the scientific disciplines that have been most important to biotechnology, we coded the educational backgrounds of the Ph.D.-holding, university-employed scientific advisory board members of all publicly traded biotechnology firms. The source of information on scientific advisors' degrees was the IPO prospectuses of the 533 U.S.-based biotechnology firms that have been filed with the U.S. Securities and Exchange Committee. We then stratified the random draw from UMI to correspond to the disciplines and Ph.D. grant years of firms' scientific advisors. For example, 22 percent of biotechnology company scientific advisors hold biochemistry Ph.D.s; we drew a corresponding proportion of biochemists into our sample. Table 1 lists the Top 15 disciplines from which scientists in our sample are selected.

[3]Ph.D.s with academic affiliations lasting less than five years were dropped from the dataset to exclude post-doctoral fellows that later moved to jobs in industry.

observed from the year after he or she earned a Ph.D. until 1999, unless the individual exited academia.[4] The final panel contains 59,069 person-year observations between 1967 and 1999.

3.1 Variables

A brief description of the patenting process in academia is useful to interpret the results we will present. The process begins when a faculty member discloses an invention to the university's Technology Transfer Office (TTO).[5] The commercial potential of this invention is then evaluated by the TTO, which may decide to seek patent rights on the invention. Concurrently, the TTO will market the innovation to potential licensing partners in industry. A typical licensing agreement specifies a 40% royalty rate for the individual faculty inventor, to be assessed on the gross licensing revenues the invention accrues.

The patents of the academic scientists in our data were assembled from the NBER patent database (Hall, Jaffe, and Trajtenberg, 2001). To identify academic patenters, we matched the scientists in our dataset to the list of inventors in the NBER patent database. Matches were performed on the basis of last and first names, and we used information on assignee (university) and geographic region to eliminate false matches.[6] For each scientist in our data, we generated two dependent variables: time of first patent application and a dummy variable indicating whether the researcher applied for at least one patent in a given year.

Research Output and Latent Patentability. We create three measures of scientists' research output. From the *Web of Science* we computed annual paper publication counts for each scientist. We count all papers on which a scientist is listed as an author.[7] While this

[4]We assume a researcher has exited academia when he or she fails to publish for five consecutive years, or in fewer instances, when the scientist begins to publish almost exclusively under a corporate affiliation. In either case, we censor observation in the year in which a scientist last publishes under a university affiliation.

[5]Faculty members are contractually obligated to disclose potentially commercializable discoveries developed on university premises to the TTO. They do not have the option to patent university-originated discoveries without going through the official channels. Because university-employed scientists rarely patent their research outside of the TTO, we are able to incorporate assignee (university) information in matching scientists' names to inventors listed on patents.

[6]Because we know the affiliations of the scientists in our data, we do not face the daunting name-matching challenges described in Trajtenberg (2004). We are able to rule out false positives by insisting that both scientists' names and affiliations match the inventor and assignee fields in the patent data.

[7]In other words, we treat sole-authored and co-authored papers as equivalents. By convention in the life sciences, the first author has made the greatest intellectual contribution to the project, and the last author

seems a good proxy for the rate of a given scientist's output, we would also like to measure the content of this output. We do this in two different ways. First, we use the affiliation data available from *Web of Science* to identify all instances in which a scientist wrote a paper that was coauthored with one or more individuals in a corporate research and development lab. We assume that papers coauthored with researchers in industry are more likely to be of an applied nature, and thus we consider publishing with coauthors in industry as an indicator of the degree to which scientists are engaging in commercially-oriented research.

Second, it would be desirable to directly account for differences among scientists in the inherent "patentability" of their research. To construct such a measure, we have used the title words in scientists' publications to identify the areas in which they have conducted research, and then applied weights to theses areas based on an (endogenous-to-the-sample) measure of the extent to which other scientists working in these areas have patented their discoveries. Intuitively, we use the publications of scientists that have already applied for patent rights as the benchmark for patentable research, and then compare the research of each scientist in our dataset to this benchmark to generate a research patentability score for each scientist-year. Specifically, the research patentability score for scientist i in year t is defined as:

$$PATENTABILITY_{it} = \sum_{j=1}^{J} w_{j,t-1}^{i} \frac{n_{ijt}}{\sum_k n_{ikt}}$$

where $j = 1, \ldots, J$ indexes each of the scientific keywords appearing in the titles of the journal articles published by scientist i in year t,[8] n_{ijt} is the number of times each of the keywords j has appeared in scientist i's articles published in year t, and w_{jt}^{i} is a weight for each keyword that measures the frequency with which word j is used in the titles of articles published by scientists who have entered the patenting regime in year t or earlier, relative to those who have not entered the patenting regime as of year t (the calculation of w_{jt}^{i} is detailed in the data appendix). Intuitively, the patentability of a scientist's research can

is the research group leader, typically the senior scientist on the team. Restricting the set of papers to those where the focal scientist appears first or last in the authorship list generates results substantively similar to those we present below.

[8]We relied on title words in journal articles instead of journal- or author-assigned keywords because the Web of Science database did not begin to include keyword descriptors until 1992. However, the titles of biomedical research papers typically indicate the research area and the methodology used in the paper. We find high overlap between title words and keywords in papers for which both are available.

change because of a change in the direction of the research of that scientist, or because other patenters' research increasingly comes to resemble that of the scientist. The former effect is captured by the ratio $\frac{n_{ijt}}{\sum_k n_{ikt}}$, the latter by the weights $w^i_{j,t-1}$. Because the benchmark in year $t-1$ is used to weight title words in year t, year-to-year changes in the research patentability score will only reflect actions of the scientist (through their choices of title keywords), rather than contemporaneous changes in the benchmark.[9]

Finally, to capture the idea that the inherent patentability of past research might still influence the current propensity to patent, we compute a depreciated stock of the research patentability score using a perpetual inventory model. Through the impact of the depreciation rate δ[10], this formulation captures the fact that the recent substantive research orientation of a scientist's research should influence current behavior more strongly than scientific trajectories that unfolded in the more distant past:

$$STOCK\_RP_{it} = (1-\delta)STOCK\_RP_{i,t-1} + FLOW\_RP_{it} = \sum_{\tau=0}^{t}(1-\delta)^{t-\tau} \cdot FLOW\_RP_{i\tau}$$

Following a number of studies of the determinants of scientists' productivity, we were also able to construct many control variables to account for individual and institutional attributes that may influence rates of publication and patenting. To account for life-cycle effects (Stephan, 1996), we include the number of years since a scientist earned his or her Ph.D. Because the time involved in publishing scientific research varies across fields, the regressions include a full set of dummies for researchers' dissertation subject areas. Some of the regressions control for time invariant quality differences among researchers through the inclusion of scientist fixed effects. In specifications without fixed effects, we enter a dichotomous measure of the quality of a scientists' Ph.D.-degree granting institution — a dummy variable indicating whether or not a scientists' doctoral program was ranked in the Top 20. Specifically, we collected Gourmand Report rankings for all institutions in our

[9]Previous researchers have developed other measures of proximity in technological space. For instance, Jaffe (1986) used a cosine-based measure to assess the proximity between the R&D portfolio of any given pair of firms. While this approach works well for measuring technological distance between dyads, it is not well suited to our setting, since we need to measure the distance between the scientific trajectory of any given scientist relative to that of a benchmark *group* of (patenting) scientists.

[10]We set δ equal to .15 — the Griliches constant — which has been used by many innovation researchers on whose work this paper builds. We verified that our core results are not sensitive to this arbitrary choice.

dataset. Gourmand ranking are available at the field level and were issued for the first time in 1980. We assigned universities their original rating for all years prior to 1980 (and updated them every other year for the subsequent period).

We also include a number of employer-level variables that may influence scientists' patenting. These covariates are updated each year and when scientists change employers. First, given the existing evidence that prominent universities are more likely to be involved in commercial activities, we include a quality rank dummy variable analogous to the one constructed for Ph.D.-granting institutions. Second, we used the AUTM surveys to create a technology transfer office (TTO) dummy variable, which is set to one in all years in which a scientist's employing university has an active TTO. Third, a university's stock of patents (excluding those of the focal scientist) is entered in the model, among other things to further control for institutional differences in support for patenting. We include a five year patent stock measure for scientists' doctoral training universities.

Finally, to capture the patenting proclivities of our scientists' coauthors, we measure both the number of coauthors and whether the coauthors have applied for patents. We are able to identify patenting behavior only for coauthors that are also members of our sample. Since the set of scientists analyzed here are drawn randomly from the population, this limitation should not introduce bias, although the resulting count is clearly a noisy proxy for the underlying concept. Furthermore, to distinguish the coauthor peer effect from the influence of peers at the same institution, we exclude coauthors that are also co-workers when creating these two variables.

3.2 Econometric Considerations

Estimating the determinants of faculty patenting behavior requires a procedure that accommodates the discrete nature of the event. Since our interest lies in analyzing the dynamics associated with the onset of patenting in scientific careers, we employ discrete-time hazard rate models (Cob 1972, Ayers, Hanky and Mantel 1973, Alison 1982). The use of discrete-time models (as opposed to continuous-time models such as the Cob) is motivated by the fact that our failure time variable displays multiple events within each time period. For a researcher i

during experience interval t, let the discrete time hazard rate be $p_{it} = Pr[T_i = t | T_i \geq t, X_{it}]$, where T_i is the time at which research i experiences an event and X_{it} a vector of covariates. We use a logistic regression function to link the hazard rate with time and the explanatory covariates:

$$Ln[\frac{p_{it}}{1 - p_{it}}] = \alpha_t + \beta' X_{it}$$

where α_t is a set of experience interval dummies. In practice, we estimate a simple logit of the decision to apply for a patent, where the observations corresponding to years subsequent to the first event have been dropped from the estimation sample.

These models essentially rely on between-scientist covariate variation to identify the determinants of the first transition to patenting. A complementary approach is to consider how within-scientist changes in covariates influence the propensity to patent. We do so by estimating so-called "fixed-effects" logit models by conditional maximum likelihood (Chamberlain, 1984). In contrast to our implementation of the standard logits, this approach analyzes the careers of patenting scientists in their entirety, rather than just until the year of first patent application. In other words, we treat patenting as a repeatable event in the fixed-effects logit regressions. There is, however, a countervailing cost in the fixed-effects approach, in that it drops all observations corresponding to scientists who never patent.[11] We believe that, together, the discrete-time hazard models and the fixed effects logit models provide a comprehensive picture of the academic patenting phenomenon.

4 Results

Among the 3,884 researchers in our sample, 758 (20%) hold one or more patents. In Figure 1, we plot the distribution of patents for the patenting researchers in our sample. The histogram illustrates a rapid drop off after one — most patenters are listed on 1 or 2 patents throughout their career, and very few scientists in our data receive more than 10 patents. Figure 2 displays the distribution of scientists' total publication counts, broken out by their patenting status. Consistent with the conventional wisdom that patenting is concentrated among the

[11] Conditional maximum likelihood estimation requires some variation in the dependent variable to condition out the individual scientist effects. Because scientists that have never patented have no variation on the outcome variable, they must be dropped from the analysis.

group of academically productive scientists, the distribution for the patenter subsample is much less skewed than that of the non-patenter subsample.

Descriptive statistics. Table 2 presents the summary descriptive statistics for variables used in our analysis. Table 3 reports, by scientists' patenting status, the mean research and employer characteristics measured at five career stages. This table shows that researchers who have sought and received patent rights for their discoveries are more productive at each career stage: they publish almost twice as many research papers as those who have not yet entered the patenting regime. Scientists who have applied for patent rights are closer to commercial research than their non-patenting counterparts, as indicated by the fact that they have collaborated more often with researchers in the private sector. Likewise, the intrinsic patentability of their research appears higher. Finally, patenters are more likely to work in settings where a technology transfer office exists and patenting activity is intensive, and they are more likely to have coauthors that have themselves patented.

Figure 3 displays the distribution of patenting events over time. Although we observe an uptick in the years following Bayh-Dole, it is also clear that patenting activity was taking place even before the adoption of the Act. This is consistent with the findings of Mowery et al. (2001).

Figure 4 displays, for the first decade of scientists' careers, the unconditional hazard of first patent application against experience (as measured by years since graduation) for three distinct cohorts of scientists: those who received their Ph.D. between 1967 and 1975, those who earned their degree between 1976 and 1985, and those who matriculated between 1986 and 1990. It is clear from Figure 4 that, over successive cohorts, the probability of patenting in an early career stage has increased, and in the latest cohort of life scientists in our data, it is increasing at a greater rate.[12]

One possible explanation for the greater incidence of patenting among early career scientists is that, in recent years, post-doctoral fellows are more likely to be listed as co-inventors

[12]The decline in the unconditional hazard for the third cohort after the fifth year of experience is caused by the gradual censoring of the patent data. Specifically, the NBER patent database contains data on patents issued until 1999. Because our measure of patenting is dated to the time of the filing of an application for a patent that eventually issues, the final years of our data contain fewer patenting events because we do not observe patents that were applied for prior to 1999, but did not issue until after this year.

on patents emanating from the research conducted in their advisors' labs. Examining data from the 1970s and 1980s, Stephan and Ma (2004) report that there has been an increase during this period both in the proportion of scientists who begin their careers as post-doctoral fellows, and in the duration of these fellowships. We cannot explore this possibility directly because our data neither allow us to distinguish post docs from regular faculty, nor do they identify advisor/post-doc pairings. However, we are able to document general trends in the incidence of patent co-inventorship.

Figure 5 presents the proportion of all patents in our data that list (i) a sole inventor (dashed line) or (ii) list three or more inventors (solid line), plotted against the number of years since the patenting academic scientist received his or her Ph.D. degree. The figure demonstrates a clear negative trend in scientists' proclivities to receive sole invented patents over their careers, and a slightly positive trend in the incidence of multiple-inventor patents over the career. These data alone do not permit us to firmly rule out the possibility that early career patenting is somehow associated with changes in the duration and prevalence of post-docs, but it is evidently the case that the life-cycle trend is from sole to multiple-inventor patents, and not vice versa. As a result, we consider it likely that the increase in slope in the early career hazard of patenting observed in Figure 4 reflects the fact that patenting increasingly has come to be recognized as a legitimate form of scientific output in the academic life sciences.

Discrete-time hazard rate models. We now present results from the discrete-time hazard rate regressions. The results can be found in Tables 4a, b and c. Model (1) includes the variables often thought to be associated with academic patenting, but without the paper count and the patentability variables. All models control for (unreported) Ph.D subject areas and calendar year dummies. The results are consistent with the findings of previous studies, and confirm the patterns that were already apparent in the descriptive statistics. We find evidence that controlling for the number of coauthors, scientists with at least one patenting coauthor are more likely to patent. We caution readers against interpreting this correlation as evidence of patenting peer effects, as it could merely reflect assortative matching among scientists along some other dimension correlated with patenting. We also find a strong influence of co-authorship with corporate researchers on the likelihood of first patent application. At the mean of the other covariates, having coauthored with researchers in industry increases the predicted probability of patenting by 76%.

In contrast to the individual-level covariates, the impact of employer-related variables is mixed. We fail to find an effect of the presence of a technology licensing office (although this could be due to the fact that this organizational innovation diffused quite rapidly among Tier-1 universities following the Bayh-Dole Act). However, we do find an effect for the intensity of patenting at the university where the scientist earned his/her doctorate in the five years preceding the award of the degree (but not for the patent stock of the current employer).[13]

Model (2) adds two variables to the specification: a scientist's count of publications in year $t-1$, and a cumulative stock of publications up to year $t-2$. Only the flow variable is significant, suggesting that patenting is accompanied by a flurry of scientific activity. At the mean of the data, each additional research publication increases the researcher's odds of entering the patenting regime during the next year by 7.6%; a one standard deviation increase (2.7) in the flow of research publications is associated with a 18% increase in the likelihood of patenting relative to the baseline rate. In Models (3) and (4), we explore further the timing of this flurry by using more flexible specifications for the distributed lag of publications. In Model (3), we include the flow of publications in year $t-2$ and the stock up to year $t-3$. In Model (4), we include the flow of publications in year $t-3$ and the stock up to year $t-4$. In both cases, only the coefficient for the one-year lagged variable is significant; in other words, Model (2) appears to capture accurately the timing of the publication flurry associated with patenting.

This conditional correlation strikes us as being an important finding, for it can help distinguish between competing interpretations of the association between scientific productivity

[13]In Table 4, the hazard of patenting appears to be monotonically decreasing in experience. However, this trend merely reflects our decision to limit the analysis to the first patenting event. Because we drop scientists from the data once they have patented, we would expect to observe negative duration dependence as only those scientists that have not yet patented prior to an experience interval remain in the risk set during that interval. In other words, the scientists that remain in the risk set to inform the coefficient estimates for the later experience intervals are predominantly non-patenters.

and involvement with the world of commerce. In the first interpretation, commercialization activities correspond to attempts by academics to monetize established reputations and professional status. In the second interpretation, publications and patents are co-occuring outputs that encode the same set of scientific insights; patents, just like publications, reflect genuine shocks to scientific opportunities. We see the correlation between the onset of patenting and the lagged flow, but not the stock, of publications as much more consistent with the latter interpretation.[14] The plausibility of this interpretation is reinforced by a peculiar aspect of US patent law, which grants inventors a one-year grace period from the date of publication for the filing of a patent application (Merges, 1997, p. 226). In other words, an academic inventor wishing to maximize the effective life of a patent would apply to the USPTO exactly 364 days after the date of publication, provided that he/she is willing to forego patent protection in foreign jurisdictions.

Using the specification in Model (2) as a benchmark, Table 4b examines the influence of the latent patentability of the scientist's research on his/her propensity to enter the patenting regime. We proceed with the analysis parallel to the approach taken in Table 4a. Model (5) adds the flow of our research patentability score in year $t-1$ (*i.e.,* based on our endogenous-to-the-sample measure, the extent to which the papers a scientist has published in the previous year are substantively similar to the work previously published by patenting scientists) and the corresponding cumulative stock up to year $t-2$. Here again, we find that only the flow influences the likelihood of patenting. At the mean of the data, increasing the patentability score by one standard deviation increases the likelihood of first patent application by 15%. Moreover, as can be seen in Models (6) through (8), the conclusion is not altered when using a more flexible functional form to model the distributed lag of the latent patentability score.[15] Just as in the case of publications, the onset of patenting appears simultaneous with

[14] This interpretation is also consistent with Murray and Stern's (2005) analysis of paper-patent pairs, but it suggests that this phenomenon is not confined to the single journal whose articles they analyze. Of course, since we do not examine the actual content of patents and papers, we can only provide circumstantial evidence in favor of a substantive linkage between these two forms of output. In practice, it seems likely that patentable claims will be spread over a number of papers revolving around a common theme, some published before, some after the filing of the patent application.

[15] Specifically, in Model (8), we replace the research patentability flow at $t-1$ with two dummies: one for observations that lie strictly above 0 but below the to 75^{th} percentile of the research patentability flow variable; and the other for observations above the 75^{th} percentile of the variable.

a change in the content of a scientist's research in a direction that makes it more similar to that of scientists who have already applied for patent rights. But because it is the flow, and not the stock of this measure that seems to matter, the evidence is consistent with the idea that a patent application does not constitute merely a response to changes in the formal and informal incentives faced by academic scientists over their careers, but also reflects the seizing of opportunities along a novel research trajectory.

Using Model (8) as the benchmark, Table 4c examines a variety of interactions between known determinants of patenting behavior and the flurry of publications observed in Table 4a. Model (9) interacts the flurry with five dummies corresponding to different levels of experience. Although the patterns are not very pronounced, it appears that the magnitude of the publication flurry required to shift a scientist into the patenting regime varies over the life cycle and follows an inverted U-shape. In particular, it is during the first five to eight years of the experience clock that the effect of the flurry is most pronounced. For life scientists, this typically corresponds to their first job as established, independent investigators. The decrease observed in subsequent years is consistent with human capital vintage effects that have been frequently mentioned (though not often estimated) in the economics of science literature. Models (10) through (13) interact the flurry with different institutional and contextual measures. We find that the magnitude of the flurry is smaller for scientists working in "patent-intensive" universities (Model 10) and for scientists who have coauthors who themselves have patented (Model 13). In other words, the evidence suggests that the magnitude of the opportunity necessary to shift an individual into the patenting regime is larger in academic environments in which the costs of patenting are higher, either because of bureaucratic hurdles, or a lack of cultural support for involvement in commercial activity.

In summary, we find that individual rates of patenting respond to scientific opportunities, and that patenting coincides with a genuine change in the content of these scientists' research. In addition, our results suggest that social influences operating in graduate school, in the scientist's current university, or through his/her "invisible college" of collaborators shape the intensity of commercial activities among academics.

Fixed-effects logit models. The results presented above suffer from two limitations. First, they only pertain to the decision to apply for the *first* patent. For a sizable proportion of scientists, patenting is a repeated event, and the determinants of patenting could differ in the group of serial patenters. Moreover, one might object that our result regarding the flurry of publications contemporaneous with patenting assumes that the lagged stock of publications adequately captures differences in talent among scientists. It would be desirable to subject this set of results to a more stringent test. For these reasons, Tables 5a, 5b and 5c replicate the analyses presented in tables 4a, 4b and 4c using fixed-effects logit models. In these models, patenting is treated as a repeated event, and there are as many observations in the estimation sample as there are person-years for patenting scientists.[16]

Table 5a shows that the impact of the one-year lagged count of publications remains even after accounting for time-invariant talent differences among scientists through fixed individual effects, and that the inclusion of additional lags does not modify the result. We interpret this finding as suggesting that within-scientist changes in scientific opportunities influence their likelihood of patenting.

Similarly, Table 5b highlights the role of changes in the latent patentability of a scientist's research, which again appear to correlate with patenting events. The statistical significance of these results is weaker than in the corresponding "cross-sectional" hazard rate models. In Model (8), we partition the one-year lag of the patentability measure into three separate dummy variables corresponding to 0, above 0 but below the 75^{th} percentile, and above the 75^{th} percentile. Using this more flexible specification, Model (8) demonstrates a statistically significant influence of changes in latent research patentability on individual rates of patenting. Finally, Table 5c replicates the specifications in Table 4c. While we cannot replicate the results pertaining to the life cycle, the other results are qualitatively similar. Like the results in Table 4c, the fixed effects logit specifications indicate that an environment conducive to patenting and scientific opportunities are substitute inputs in the decision to patent among "serial patenters." In addition to the negative interactions between one-year lagged flow of publications and the university patent count (Model 10) and that between the one-year

[16] We also drop the stock variables from the specifications, since they move too slowly to be separately identified from the individual effects.

lagged flow of publications and the coauthor patenting variable (Model 13), we also observe a negative and marginally significant interaction with the number of scientists who have founded companies or sit on advisory boards (Model 11).

5 Discussion and Conclusion

The policy debate regarding interactions between industry and academia in general, and academic patenting in particular, has often taken for granted the idea that patenting represents a fundamental departure from the norms of the "Republic of Science." According to this view, academic researchers toil in relative obscurity by producing fundamental knowledge up until the time they receive tenure; subsequently, they may monetize their reputation by involving themselves in commercial pursuits. Patents, though not necessarily remunerative in and of themselves, provide academic researchers with visibility and status in the world of commerce, for example by contributing to the likelihood that they are invited to sit on corporate advisory boards (Stuart and Ding, 2005).

The findings in this paper challenge the standard account. First and foremost, our results suggest that patents and publications correspond to two types of output that have more in common than previously believed. Certainly, the positive relationship between patent applications and the flow, but not the stock, of publications suggest that patents and papers encode similar pieces of knowledge, a fact exploited by Murray and Stern (2005) in their investigation of the anti-commons hypothesis. Second, our results suggest that the academic incentive system is evolving in ways that accommodate deviations from traditional scientific norms of openness. Many patenting events in our data take place in the early years of scientists' careers, and the slope of the patent-experience curve has become steeper with more recent cohorts of scientists. This finding dovetails with qualitative accounts that emphasize that patents are becoming *de rigueur* on academic vitas in many institutions, and are even considered legitimate forms of research output in promotion decisions.

If the present paper investigates the antecedents of academic patenting, much work remains to be done on the effects of this now-prevalent practice on the rate of scientific progress. Does applied research (as embodied in patents) crowd out the fundamental pursuit of knowledge (as measured by publications)? Answering this question is difficult, because patenting is a choice variable for scientists, and the outcome of a decision that could easily reflect expectations of future scientific productivity. Our paper provides an important input into this analysis by presenting the results of a selection equation whose estimation is necessary to recover causal effects of patenting on scientific output (Azoulay et al., 2005a). But our results also alert us to the possibility that the substantive content of post-patent publications might be different from these scientists' pre-patent output, leading naturally to the study of the effect of patenting on the *direction* of scientific progress. Our measure of latent patentability, whose construction is an important contribution of this paper, can be used on the left-hand side of a regression equation to investigate this important question.

Finally, our findings suggest that social contagion might be an important mechanism through which the practice of academic patenting diffuses among the population of life scientists. The result that scientists whose coauthors patent are more likely to patent themselves is consistent with genuine "peer effects," but it is also consistent with assortative matching of coauthors along some other dimension correlated with patenting — such as scientific productivity. Distinguishing between these competing hypotheses remains a valuable goal for future research.

References

Agrawal, Ajay K. and Rebecca M. Henderson. 2002. "Putting Patents in Context: Exploring Knowledge Transfer from MIT." *Management Science,* 48:1, pp. 44-60.

Allison, Paul D. 1982. "Discrete-Time Methods for the Analysis of Event Histories," in Sociological Methodology. S. Leinhardt ed. San Francisco: Jossey-Bass, pp. 61-98.

Azoulay, Pierre, Waverly W. Ding, and Toby E. Stuart. 2005a. "The Impact of Academic Patenting on (Public) Research Output." Working Paper, Columbia University.

Azoulay, Pierre, Ryan Michigan, and Bhaven N. Sampat. 2005b. "The Anatomy of Medical School Patenting, 1981-2000." Working Paper, Columbia University.

Blumenthal, David, Eric G. Campbell, Nancyanne Causino, and Karen Seashore Louis. 1996. "Participation of Life-Science Faculty in Research Relationships with Industry." *New England Journal of Medicine*, 335:23, pp. 1734-39.

Campbell, Eric G. , B. R. Clarridge, N. N. Gokhale, L. Birenbaum, S. Hilgartner, N.A. Holtzman, and D Blumenthal. 2002. "Data Withholding in Academic Genetics — Evidence from a National Survey." *JAMA*, 287:4, pp. 473-80.

Chamberlain, G. 1984. "Panel Data," in *Handbook of Econometrics.* Zvi Griliches and Michael D. Intriligator, eds. Volume 2, pp. 1248-1318. Amsterdam: North-Holland.

Cox, D. R. 1972. "Regression Models and Life Tables." *Journal of the Royal Statistical Society Series B,* 34:2, pp. 187-220.

Di Gregorio, Dante and Scott Shane. 2003. "Why Do Some Universities Generate More Start-ups Than Others?" *Research Policy*, 32:2, pp. 209-27.

Ding, Waverly W., Fiona Murray, and Toby E. Stuart. 2005. "Technology: A New Arena for Gender Stratification in Scientific Careers?" Working Paper, Columbia University.

Etzkowitz, Henry. 1998. "The Norms of Entrepreneurial Science: Cognitive Effects of the New University-Industry Linkages." *Research Policy,* 27:8, pp. 823-33.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools." NBER Working Paper #8498.

Henderson, Rebecca M., Adam B. Jaffe, and M. Trajtenberg. 1998. "Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting, 1965-1988." *Review of Economics and Statistics*, 80:1, pp. 119-27.

Jaffe, Adam B. 1986. "Technological Opportunity and Spillovers from R&D: Evidence from Firms' Patents, Profits, and Market Value." *American Economic Review*, 76:5, pp. 984-1001.

Jaffe, Adam B. 1989. "Real Effects of Academic Research." *American Economic Review*, 79, pp. 957-70.

Jaffe, Adam B. and Josh Lerner. 2001. "Reinventing Public R&D: Patent Policy and the Commercialization of National Laboratory Technologies." *Rand Journal of Economics,* 32:1, pp. 167-98.

Jensen, Richard and Marie C. Thursby. 2001. "Proofs and Prototypes for Sale: The Licensing of University Inventions." *American Economic Review*, 91:1, pp. 240-59.

Krimsky, Sheldon. 2003. *Science in the Private Interest: Has the Lure of Profits Corrupted Biomedical Research*. Lanham, MD: Rowman & Littlefield.

Levin, S. G. and Paula E. Stephan. 1991. "Research Productivity over the Life-Cycle — Evidence for Academic Scientists." *American Economic Review*, 81:1, pp. 114-32.

Merges, Robert. 1997. *Patent Law and Policy: Cases and Materials.* Charlottesville, VA: Michie Co.

Mansfield, Edwin. 1995. *Innovation, Technology, and the Economy: The Selected Essays of Edwin Mansfield.* Aldershot, UK: Brookfield.

Mowery, David C., Richard R. Nelson, Bhaven N. Sampat, and Arvids A. Ziedonis. 2001. "The Growth of Patenting and Licensing by US Universities: an Assessment of the Effects of the Bayh-Dole Act of 1980." *Research Policy*, 30:1, pp. 99-119.

Murray, Fiona. 2002. "Innovation as Co-evolution of Scientific and Technological Networks: Exploring Tissue Engineering." *Research Policy,* 31:8-9, pp. 1389-1403.

Murray, Fiona and Scott Stern. 2005. "Do Formal Intellectual Property Rights Hinder the Free Flow of Scientific Knowledge? Evidence from Patent-Paper Pairs." Working Paper, Northwestern University.

Myers, M.H., B.F. Hankey, and N. Mantel. 1973. "A Logistic Exponential Model for Use with Response-Time Data Involving Regressor Variables." *Biometrics,* 29, pp. 257-69.

Owen-Smith, Jason and Walter W. Powell. 2001. "Careers and Contradictions: Faculty Responses to the Transformation of Knowledge and its Uses in the Life Sciences." *Research in the Sociology of Work,* 10, pp. 109-40.

Stephan, Paula E. 1996. "The Economics of Science." *Journal of Economic Literature*, 34:3, pp. 1199-235.

Stephan, Paula E., Shiferaw Gurmu, A.J. Sumell, and Grant Black. 2004. "Who's Patenting in the University? Evidence from the Survey of Doctorate Recipients." Working Paper, Georgia State University.

Stephan, Paula E. and Jennifer Ma. 2004. "The Increased Frequency and Duration of the Postdoctorate Career Stage." Working Paper, Georgia State University.

Stuart, Toby E. and Waverly W. Ding. Forthcoming. "The Social Structural Determinants of Academic Entrepreneurship: An Analysis of University Scientists' Participation in Commercial Ventures." *American Journal of Sociology.*

Thursby, Jerry G. and Marie C. Thursby. 2002. "Who is selling the Ivory Tower? Sources of Growth in University Licensing." *Management Science*, 48:1, pp. 90-104.

Trajtenberg, Manuel. 2004. "The Names Game: Using Inventors Patent Data in Economic Research." Presentation to the NBER Summer Institute, July 2004.

Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer. 1998. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review*, 88:1, pp. 290-306.

Table 1
Top 15 Scientific Disciplines in the Sample

UMI Subject Code	UMI Subject Description	Frequency	
487; 303	Biochemistry	861	(22.2%)
306	Biology, General	568	(14.6%)
410	Biology, Microbiology	469	(12.1%)
419	Health Sciences, Pharmacology	240	(6.2%)
490	Chemistry, Organic	213	(5.5%)
786	Biophysics, General	211	(5.4%)
369	Biology, Genetics	191	(4.9%)
433	Biology, Animal Physiology	171	(4.4%)
982	Health Sciences, Immunology	167	(4.3%)
307	Biology, Molecular	102	(2.6%)
301	Bacteriology	63	(1.6%)
287	Biology, Anatomy	54	(1.4%)
571	Health Sciences, Pathology	52	(1.3%)
349	Psychology, psychobiology	37	(1.0%)
572	Health Sciences, Pharmacy	34	(0.9%)

Legend: Table 1 reports the Top 15 disciplines from which our sample was drawn. These disciplines have spawned the greatest number of biotechnology firm founders, scientific advisors and executives. The table also reports the frequency and the proportion of scientists in our sample for each of these 15 scientific disciplines.

Table 2
Descriptive Statistics

	Mean	Std. Dev.	Min.	Max.
Time-varying (59,069 person-year observations)				
Experience	10.22	7.127	1	32
Patent flow dummy	0.039	0.194	0	1
Patent regime dummy	0.125	0.330	0	1
Publications flow	1.677	2.667	0	100
Publications stock	16.48	27.68	0	645
Research patentability flow	0.084	0.108	0	5.185
Research patentability stock	0.460	0.417	0	5.659
Collaboration tie with company scientists	0.263	0.441	0	1
Average number of identified coauthors per paper	0.131	0.248	0	10
Identified coauthors have patents	0.200	0.400	0	1
Employer graduate school in top 20	0.232	0.422	0	1
Employer has TTO	0.489	0.500	0	1
Employer patent stock (in hundred)	0.717	1.450	0	22
Employer entrepreneurial faculty count	8.634	22.89	0	199
Calendar year	1986	7.741	1968	1999
Time-invariant (3,884 observations)				
Ph.D. univ. grad. school in top 20	0.308	0.462	0	1
Ph.D. univ. 5-yr patent stock (in hundred)	19.02	40.89	0	566
Ph.D. univ. entrepreneurial faculty count	2.294	8.304	0	182

.

Table 3
Scientists' Mean Research and Employer Characteristics at Five Career Stages by Patent Application Status

	Experience = 5		Experience = 10		Experience = 15		Experience = 20		Experience = 25	
Scientist Has at least one patent application	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*	*Yes*	*No*
Research publications stock	9.610	4.456	23.082	12.485	39.579	22.003	55.917	32.843	77.679	41.747
Research patentability stock	0.342	0.244	0.613	0.491	0.856	0.662	1.021	0.788	1.103	0.833
Count of collaboration ties with company scientists	0.968	0.206	2.282	0.697	3.562	1.280	4.838	2.168	7.540	2.562
Identified coauthors have patents	0.197	0.081	0.367	0.176	0.525	0.265	0.611	0.337	0.636	0.396
Employer grad. school ranks in top20	0.261	0.266	0.261	0.220	0.249	0.196	0.195	0.182	0.187	0.170
Employer has TTO	0.463	0.383	0.576	0.482	0.698	0.586	0.735	0.682	0.829	0.727
Employer Patent stock (in 100)	0.738	0.537	1.072	0.650	1.223	0.741	1.279	1.089	1.571	1.212
Employer entrepreneurial faculty count	7.110	6.260	11.815	8.683	14.983	10.736	14.723	12.905	16.652	12.251
Observations	218	3612	330	2286	354	1503	339	978	187	454

Legend: Table 3 reports the mean research and employer characteristics measured at five different stages in scientists' career: the 5th, 10th, 15th, 20th and 25th year after the scientist was granted a Ph.D. Within each career stage, the table is further broken out by whether a scientist has ever received a patent.

Table 4a
First Patenting Event — Discrete-Time Hazard Models

	(1)	(2)	(3)	(4)
Experience [1, 4]	0.700 (0.224)**	0.691 (0.240)**	0.648 (0.238)**	0.625 (0.237)**
Experience [5, 8]	0.627 (0.211)**	0.593 (0.228)**	0.537 (0.227)*	0.490 (0.226)*
Experience [9, 15]	0.609 (0.197)**	0.573 (0.212)**	0.526 (0.212)*	0.487 (0.211)*
Experience [16, 22]	0.547 (0.194)**	0.526 (0.201)**	0.502 (0.201)*	0.482 (0.201)*
Collaboration tie with company scientists$_{t-1}$	0.576 (0.093)**	0.478 (0.098)**	0.474 (0.099)*	0.471 (0.100)**
Average number of identified coauthors per paper $_{t-1}$	0.319 (0.124)*	0.360 (0.122)**	0.364 (0.123)*	0.370 (0.123)**
Identified Coauthors have patent $_{t-1}$	0.522 (0.096)**	0.419 (0.104)**	0.416 (0.105)*	0.412 (0.106)**
Ph.D. University Grad School in Top 20	-0.058 (0.086)	-0.065 (0.087)	-0.065 (0.087)	-0.066 (0.087)
Ph.D. University 5-year Patent Stock	0.002 (0.001)†	0.002 (0.001)†	0.002 (0.001)†	0.002 (0.001)†
Ph.D. University Entrepreneurial Faculty Count	0.004 (0.006)	0.004 (0.006)	0.004 (0.006)	0.004 (0.006)
Employer Grad School in Top 20	-0.034 (0.104)	-0.052 (0.105)	-0.054 (0.105)	-0.054 (0.106)
Employer has a TTO$_{t-1}$	0.120 (0.092)	0.106 (0.093)	0.105 (0.093)	0.103 (0.094)
Employer Patent Stock$_{t-1}$	0.046 (0.030)	0.044 (0.031)	0.044 (0.032)	0.044 (0.032)
Employer Entrepreneurial Faculty Count$_{t-1}$	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Research Publication Stock$_{t-2}$		-0.002 (0.003)		
Research Publication Stock$_{t-3}$			-0.004 (0.003)	
Research Publication Stock$_{t-4}$				-0.006 (0.003)†
Research Publication Flow$_{t-1}$		0.079 (0.021)**	0.069 (0.020)*	0.065 (0.020)**
Research Publication Flow$_{t-2}$			0.025 (0.020)	0.016 (0.022)
Research Publication Flow$_{t-3}$				0.025 (0.020)
Constant	-9.010 (1.022)**	-9.043 (1.026)**	-9.000 (1.025)*	-8.973 (1.025)**
Number of observations	52,466	52,466	52,466	52,466
Number of researchers	3,884	3,884	3,884	3,884
Number of first patenting events	758	758	758	758
Log-likelihood	-3788.78	-3766.02	-3765.05	-3764.17
Wald Chi^2	313.42	321.71	323.42	327.00
Model d.f.	47	49	50	51
Pseudo-R^2	0.04	0.05	0.05	0.05

Table 4b
First Patenting Event — Discrete-Time Hazard Models

	(5)	(6)	(7)	(8)
Experience [1, 4]	0.791	0.846	0.886	0.761
	(0.241)**	(0.243)**	(0.250)**	(0.241)**
Experience [5, 8]	0.598	0.576	0.564	0.593
	(0.227)**	(0.229)*	(0.232)*	(0.227)**
Experience [9, 15]	0.553	0.530	0.518	0.544
	(0.211)**	(0.211)*	(0.212)*	(0.211)**
Experience [16, 22]	0.511	0.498	0.491	0.499
	(0.199)*	(0.199)*	(0.199)*	(0.200)*
Collaboration tie with company scientists$_{t-1}$	0.419	0.401	0.396	0.429
	(0.095)**	(0.094)**	(0.094)**	(0.095)**
Average number of identified coauthors per paper $_{t-1}$	0.302	0.312	0.322	0.300
	(0.127)*	(0.128)*	(0.128)*	(0.125)*
Identified Coauthors have patent $_{t-1}$	0.363	0.353	0.347	0.355
	(0.102)**	(0.101)**	(0.101)**	(0.101)**
Ph.D. University Grad School in Top 20	-0.071	-0.075	-0.076	-0.078
	(0.087)	(0.087)	(0.087)	(0.087)
Ph.D. University 5-year Patent Stock	0.002	0.002	0.002	0.002
	(0.001)†	(0.001)†	(0.001)†	(0.001)†
Ph.D. University Entrepreneurial Faculty Count	0.005	0.005	0.005	0.005
	(0.006)	(0.006)	(0.006)	(0.005)
Employer Grad School in Top 20	-0.052	-0.051	-0.050	-0.056
	(0.104)	(0.104)	(0.104)	(0.104)
Employer has a TTO$_{t-1}$	0.103	0.101	0.100	0.099
	(0.094)	(0.094)	(0.094)	(0.093)
Employer Patent Stock$_{t-1}$	0.051	0.050	0.050	0.049
	(0.032)	(0.032)	(0.032)	(0.032)
Employer Entrepreneurial Faculty Count$_{t-1}$	-0.001	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)
Research Publication Stock$_{t-2}$	-0.002	-0.002	-0.002	-0.003
	(0.002)	(0.002)	(0.003)	(0.003)
Research Publication Flow$_{t-1}$	0.060	0.059	0.060	0.072
	(0.016)**	(0.016)**	(0.016)**	(0.017)**
Research Patentability Stock$_{t-2}$	0.014			0.107
	(0.133)			(0.125)
Research Patentability Stock$_{t-3}$		0.007		
		(0.144)		
Research Patentability Stock$_{t-4}$			-0.003	
			(0.154)	
Research Patentability Flow$_{t-1}$	1.787	1.713	1.693	
	(0.459)**	(0.465)**	(0.469)**	
Research Patentability Flow$_{t-2}$		-0.583	-0.552	
		(0.495)	(0.491)	
Research Patentability Flow$_{t-3}$			-0.216	
			(0.498)	
Intermediate Research Patentability Flow$_{t-1}$				0.465
				(0.097)**
High Research Patentability Flow$_{t-1}$				0.552
				(0.110)**
Constant	-8.885	-8.714	-8.648	-9.206
	(1.026)**	(1.028)**	(1.030)**	(1.027)**
Number of observations	52,064	52,064	52,064	52,466
Number of researchers	3,884	3,884	3,884	3,884
Number of first patenting events	758	758	758	758
Log-likelihood	-3,718.31	-3,714.07	-3,713.31	-3,747.64
Wald Chi^2	369.26	377.28	378.67	372.57
Model d.f.	52	54	56	52
Pseudo-R^2	0.05	0.05	0.05	0.05

Table 4c
First Patenting Event — Discrete-Time Hazard Models

	(9)	(10)	(11)	(12)	(13)
Experience [1, 4]	0.700 (0.256)**	0.757 (0.242)**	0.775 (0.241)**	0.807 (0.243)**	0.862 (0.246)**
Experience [5, 8]	0.474 (0.230)*	0.586 (0.229)*	0.604 (0.228)**	0.635 (0.230)**	0.676 (0.232)**
Experience [9, 15]	0.491 (0.218)*	0.533 (0.211)*	0.548 (0.211)**	0.578 (0.213)**	0.610 (0.214)**
Experience [16, 22]	0.536 (0.210)*	0.480 (0.198)*	0.501 (0.199)*	0.517 (0.200)**	0.528 (0.199)**
Collaboration tie with company scientists$_{t-1}$	0.388 (0.096)**	0.408 (0.095)**	0.419 (0.095)**	0.424 (0.095)**	0.396 (0.095)**
Average number of identified coauthors per paper$_{t-1}$	0.331 (0.125)**	0.314 (0.125)*	0.306 (0.125)*	0.305 (0.125)*	0.289 (0.127)*
Identified Coauthors Have Patents$_{t-1}$	0.328 (0.102)**	0.335 (0.101)**	0.350 (0.101)**	0.346 (0.101)**	0.558 (0.123)**
Ph.D. University Grad School in Top 20	-0.088 (0.087)	-0.089 (0.087)	-0.082 (0.087)	-0.081 (0.087)	-0.089 (0.087)
Ph.D. University 5-year Patent Stock	0.002 (0.001)	0.002 (0.001)	0.002 (0.001)†	0.002 (0.001)†	0.002 (0.001)†
Ph.D. University Entrepreneurial Faculty Count	0.006 (0.005)	0.005 (0.005)	0.005 (0.006)	0.005 (0.005)	0.005 (0.006)
Employer Grad School in Top 20	-0.042 (0.104)	-0.034 (0.103)	-0.050 (0.104)	-0.048 (0.104)	-0.042 (0.103)
Employer has a TTO$_{t-1}$	0.086 (0.094)	0.093 (0.093)	0.090 (0.094)	0.183 (0.112)	0.090 (0.094)
Employer Patent Stock$_{t-1}$	0.052 (0.032)†	0.072 (0.030)*	0.050 (0.031)	0.049 (0.031)	0.050 (0.031)
Employer Entrepreneurial Faculty Count$_{t-1}$	-0.001 (0.002)	-0.001 (0.002)	0.001 (0.002)	-0.001 (0.002)	-0.001 (0.002)
Research Publication Stock$_{t-2}$	0.001 (0.003)	-0.002 (0.003)	-0.002 (0.003)	-0.002 (0.003)	0.001 (0.003)
Research Publication Flow$_{t-1}$		0.096 (0.022)**	0.086 (0.020)**	0.098 (0.030)**	0.116 (0.028)**
Research Patentability Stock$_{t-2}$	0.067 (0.127)	0.074 (0.126)	0.090 (0.125)	0.091 (0.125)	0.028 (0.131)
Intermediate Research Patentability Flow$_{t-1}$	0.432 (0.101)**	0.443 (0.098)**	0.454 (0.098)**	0.453 (0.099)**	0.418 (0.101)**
High Research Patentability Flow$_{t-1}$	0.527 (0.113)**	0.537 (0.111)**	0.546 (0.110)**	0.542 (0.111)**	0.516 (0.113)**
Publication Flow$_{t-1}$ × Experience [1,4]	0.110 (0.073)				
Publication Flow$_{t-1}$ × Experience [5,8]	0.124 (0.024)**				
Publication Flow$_{t-1}$ × Experience [9,15]	0.076 (0.023)**				
Publication Flow$_{t-1}$ × Experience [16,22]	0.027 (0.026)				
Publication Flow$_{t-1}$ × Experience [23,29]	0.025 (0.023)				

	(9)	(10)	(11)	(12)	(13)
Publication $Flow_{t-1}$ × Employer Patent $Stock_{t-1}$		-0.014			
		(0.006)*			
Publication $Flow_{t-1}$ × Employer Entrepreneurial $Faculty_{t-1}$			-0.001		
			(0.001)		
Publication $Flow_{t-1}$ × Employer has a TTO_{t-1}				-0.042	
				(0.031)	
Publication $Flow_{t-1}$ × Identified Coauthors have patent $_{t-1}$					-0.084
					(0.030)**
Constant	-9.163	-9.216	-9.226	-9.271	-9.320
	(1.029)**	(1.027)**	(1.027)**	(1.028)**	(1.028)**
Number of observations	52,466	52,466	52,466	52,466	52,466
Number of researchers	3,884	3,884	3,884	3,884	3,884
Number of first patenting events	758	758	758	758	758
Log-likelihood	-3,741.34	-3,743.40	-3,746.02	-3,745.79	-3,740.71
Wald Chi^2	391.68	388.76	380.13	377.52	381.33
Model d.f.	56	53	53	53	53
Pseudo-R^2	0.06	0.06	0.06	0.06	0.06

Notes:

(1) For all researchers in the sample, only observations on or before the year of the first patenting event or censoring have been used, i.e., for all researchers that have patented, the observations after the year of their first granted patent application were not used in the analysis.

(2) Models (5)-(7) use a restricted sample, in which 402 person-year observations in the unrestricted sample were excluded from the analysis. These 402 observations account for the top 1% of the research patentability flow measure.

(3) All models control for Ph.D. subject areas and calendar year dummies.

(4) Experience [23, 29] is the base category.

(5) A dummy variable indicating whether the researcher has zero publication in year t-1 is included in models (5)-(7), though not reported in the table; a dummy variable indicating whether the researcher has zero publication in year t-2 is included in models (6) and (7), though not reported in the table; a dummy variable indicating whether the researcher has zero publication in year t-3 is included in model (7), though not reported in the table.

(6) Robust standard errors in parentheses, clustered by scientist.

(7) †significant at 10%; *significant at 5%; **significant at 1%.

Table 5a
Logit Models of Patenting with Scientist Fixed Effects

	(1)	(2)	(3)	(4)
Experience [1, 4]	0.293	0.296	0.301	0.303
	(0.318)	(0.318)	(0.319)	(0.319)
Experience [5, 8]	0.652	0.631	0.632	0.632
	(0.268)*	(0.268)*	(0.268)*	(0.268)*
Experience [9, 15]	0.692	0.666	0.666	0.667
	(0.202)**	(0.202)**	(0.202)**	(0.202)**
Experience [16, 22]	0.534	0.512	0.511	0.511
	(0.132)**	(0.133)**	(0.133)**	(0.133)**
Collaboration tie with company scientists$_{t-1}$	0.388	0.341	0.338	0.338
	(0.096)**	(0.097)**	(0.097)**	(0.098)**
Average number of identified coauthors per paper $_{t-1}$	0.367	0.385	0.388	0.389
	(0.148)*	(0.148)**	(0.148)**	(0.149)**
Identified Coauthors Have Patents$_{t-1}$	-0.025	-0.059	-0.062	-0.062
	(0.100)	(0.101)	(0.101)	(0.101)
Employer Grad School in Top 20	0.066	0.085	0.087	0.087
	(0.144)	(0.144)	(0.144)	(0.144)
Employer has a TTO$_{t-1}$	0.098	0.101	0.101	0.101
	(0.094)	(0.094)	(0.094)	(0.094)
Employer Patent Stock$_{t-1}$	-0.010	-0.010	-0.010	-0.010
	(0.031)	(0.031)	(0.031)	(0.031)
Employer Entrepreneurial Faculty Count$_{t-1}$	0.005	0.006	0.006	0.006
	(0.002)**	(0.002)**	(0.002)**	(0.002)**
Research Publication Flow$_{t-1}$		0.036	0.034	0.034
		(0.011)**	(0.011)**	(0.011)**
Research Publication Flow$_{t-2}$			0.004	0.004
			(0.011)	(0.012)
Research Publication Flow$_{t-3}$				0.001
				(0.012)
Number of observations	14,507	14,507	14,507	14,507
Number of researchers	758	758	758	758
Log-likelihood	-3,932.78	-3,927.03	-3,926.97	-3,926.97
Wald Chi^2	805.64	817.13	817.25	817.25
Model d.f.	20	21	22	23
Pseudo-R^2	0.09	0.09	0.09	0.09

Table 5b
Logit Models of Patenting with Scientist Fixed Effects

	(5)	(6)	(7)	(8)
Experience [1, 4]	0.249 (0.319)	0.273 (0.319)	0.284 (0.323)	0.283 (0.318)
Experience [5, 8]	0.540 (0.269)*	0.520 (0.270)†	0.511 (0.270)†	0.596 (0.268)*
Experience [9, 15]	0.593 (0.203)**	0.576 (0.203)**	0.570 (0.204)**	0.633 (0.202)**
Experience [16, 22]	0.482 (0.133)**	0.475 (0.133)**	0.471 (0.133)**	0.494 (0.133)**
Collaboration tie with company scientists$_{t-1}$	0.319 (0.098)**	0.308 (0.098)**	0.307 (0.099)**	0.336 (0.097)**
Average number of identified coauthors per paper $_{t-1}$	0.332 (0.151)*	0.332 (0.152)*	0.336 (0.152)*	0.336 (0.150)*
Identified Coauthors Have Patents$_{t-1}$	-0.063 (0.101)	-0.067 (0.102)	-0.066 (0.102)	-0.072 (0.101)
Employer Grad School in Top 20	0.082 (0.146)	0.084 (0.146)	0.084 (0.146)	0.092 (0.144)
Employer has a TTO$_{t-1}$	0.089 (0.095)	0.087 (0.095)	0.086 (0.095)	0.097 (0.094)
Employer Patent Stock$_{t-1}$	-0.011 (0.031)	-0.012 (0.031)	-0.012 (0.031)	-0.011 (0.031)
Employer Entrepreneurial Faculty Count$_{t-1}$	0.005 (0.002)**	0.006 (0.002)**	0.006 (0.002)**	0.005 (0.002)**
Research Publication Flow$_{t-1}$	0.026 (0.011)*	0.025 (0.011)*	0.025 (0.011)*	0.033 (0.011)**
Research Patentability Flow$_{t-1}$	0.735 (0.383)†	0.719 (0.383)†	0.712 (0.383)†	
Research Patentability Flow$_{t-2}$		-0.394 (0.371)	-0.390 (0.370)	
Research Patentability Flow$_{t-3}$			-0.359 (0.373)	
Intermediate Research Patentability Flow$_{t-1}$				0.173 (0.080)*
High Research Patentability Flow$_{t-1}$				0.252 (0.083)**
Number of observations	14,332	14,332	14,332	14,507
Number of researchers	755	755	755	758
Log-likelihood	-3,881.96	-3,881.39	-3,881.39	-3,922.42
Wald Chi^2	816.11	817.24	817.24	826.36
Model df	23	24	25	23
Pseudo-R^2	0.10	0.10	0.10	0.10

Table 5c
Logit Models of Patenting with Scientist Fixed Effects

	(9)	(10)	(11)	(12)	(13)
Experience [1, 4]	0.208 (0.336)	0.279 (0.318)	0.289 (0.318)	0.281 (0.318)	0.283 (0.318)
Experience [5, 8]	0.645 (0.289)*	0.592 (0.268)*	0.594 (0.268)*	0.592 (0.268)*	0.574 (0.268)*
Experience [9, 15]	0.664 (0.225)**	0.628 (0.202)**	0.627 (0.203)**	0.628 (0.202)**	0.612 (0.202)**
Experience [16, 22]	0.671 (0.161)**	0.495 (0.133)**	0.501 (0.133)**	0.491 (0.133)**	0.486 (0.133)**
Collaboration tie with company scientists$_{t-1}$	0.332 (0.098)**	0.333 (0.097)**	0.328 (0.097)**	0.334 (0.097)**	0.311 (0.097)**
Average number of identified coauthors per paper $_{t-1}$	0.302 (0.154)*	0.331 (0.150)*	0.337 (0.150)*	0.334 (0.150)*	0.303 (0.152)*
Identified Coauthors Have Patents$_{t-1}$	-0.042 (0.102)	-0.067 (0.101)	-0.065 (0.101)	-0.067 (0.101)	0.148 (0.123)
Employer Grad School in Top 20	0.090 (0.145)	0.082 (0.145)	0.081 (0.145)	0.089 (0.145)	0.078 (0.145)
Employer has a TTO$_{t-1}$	0.099 (0.094)	0.090 (0.094)	0.092 (0.094)	0.150 (0.110)	0.093 (0.094)
Employer Patent Stock$_{t-1}$	-0.007 (0.031)	0.030 (0.035)	-0.007 (0.031)	-0.010 (0.031)	-0.010 (0.031)
Employer Entrepreneurial Faculty Count$_{t-1}$	0.006 (0.002)**	0.006 (0.002)**	0.008 (0.002)**	0.005 (0.002)**	0.005 (0.002)**
Research Publication Flow$_{t-1}$		0.046 (0.012)**	0.047 (0.012)**	0.046 (0.017)**	0.071 (0.016)**
Intermediate Research Patentability Flow$_{t-1}$	0.157 (0.080)†	0.172 (0.080)*	0.172 (0.080)*	0.171 (0.080)*	0.160 (0.080)*
High Research Patentability Flow$_{t-1}$	0.240 (0.083)**	0.252 (0.083)**	0.254 (0.083)**	0.252 (0.083)**	0.241 (0.083)**
Publication Flow$_{t-1}$ × Experience [1,4]	0.111 (0.030)**				
Publication Flow$_{t-1}$ × Experience [5,8]	0.034 (0.023)				
Publication Flow$_{t-1}$ × Experience [9,15]	0.042 (0.016)**				
Publication Flow$_{t-1}$ × Experience [16,22]	0.005 (0.014)				
Publication Flow$_{t-1}$ × Experience [23,29]	0.050 (0.018)**				
Publication Flow$_{t-1}$ × Employer Patent Stock$_{t-1}$		-0.011 (0.005)*			
Publication Flow$_{t-1}$ × Employer Entrepreneurial Faculty$_{t-1}$			-0.001 (0.000)*		
Publication Flow$_{t-1}$ × Employer has a TTO$_{t-1}$				-0.016 (0.018)	
Publication Flow$_{t-1}$ × Identified Coauthors have patent $_{t-1}$					-0.055 (0.018)**
Number of observations	14,507	14,507	14,507	14,507	14,507
Number of researchers	758	758	758	758	758
Log-likelihood	-3,915.25	-3,919.65	-3,919.74	-3,922.01	-3,917.53
Wald Chi^2	840.71	831.90	831.71	827.19	836.14
Model d.f.	27	24	24	24	24
Pseudo-R^2	0.10	0.10	0.10	0.10	0.10

Notes:

(1) Models (5)-(7) use a restricted sample, in which 175 person-year observations in the unrestricted sample were excluded from the analysis. These 175 observations account for the top 1% of the research patentability flow measure.

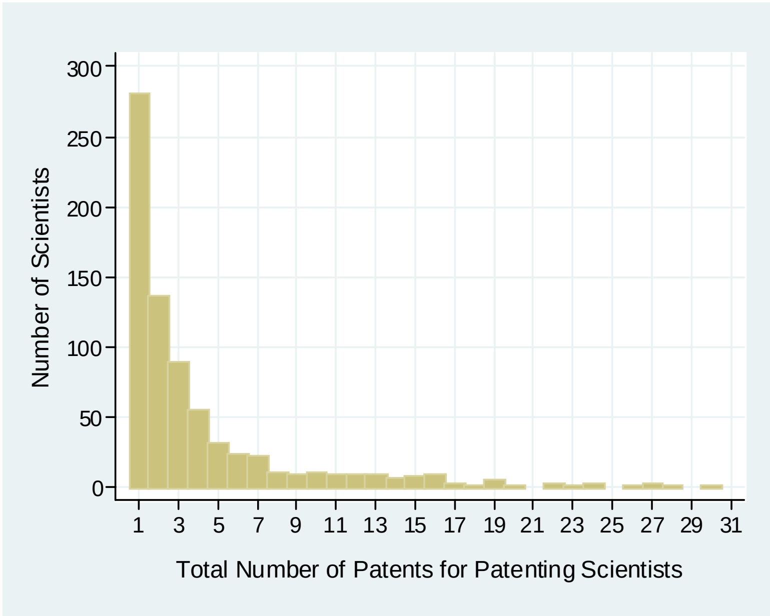
(2) All models control for period dummies 1975-76, 1977-79, 1980-82, 1983-85, 1986-88, 1989-91, 1992-94, 1995-97, 1998-99; base category is 1967-74.

(3) Experience [23, 29] is the base category.

(4) A dummy variable indicating whether the researcher has zero publication in year t-1 is included in models (5)-(7), though not reported in the table; a dummy variable indicating whether the researcher has zero publication in year t-2 is included in models (6) and (7), though not reported in the table; a dummy variable indicating whether the researcher has zero publication in year t-3 is included in model (7), though not reported in the table.

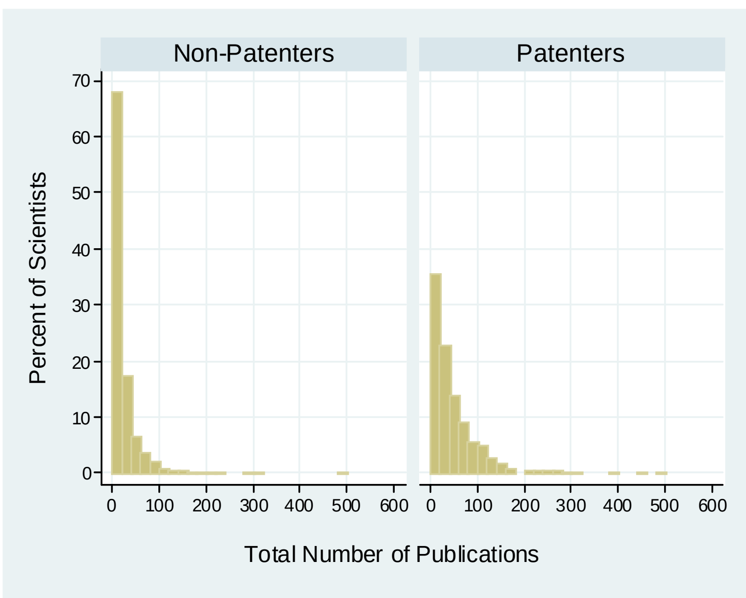
(5) †significant at 10%; *significant at 5%; **significant at 1%.

Figure 1: Distribution of Patent Count for Patenting Scientists



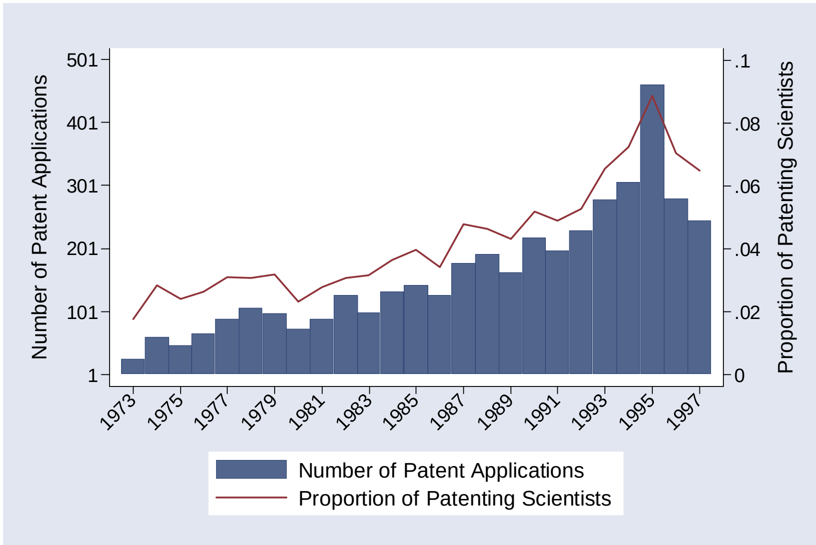
Legend: Figure 1 plots the histogram of the distribution of patents received by our 3,884 scientists over the complete sample period.

Figure 2: Distribution of Publication Count for Patenting and Non-Patenting Scientists



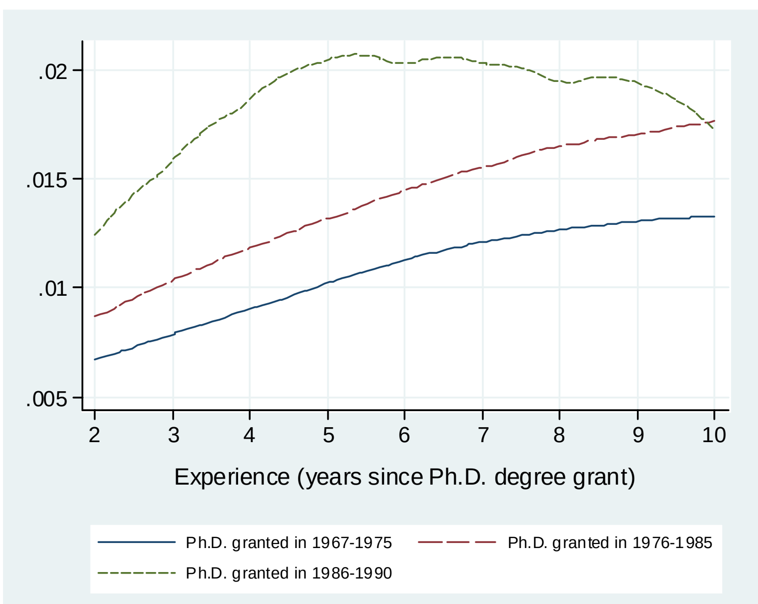
Legend: Figure 2 plots the histogram for the distribution of publication counts for our 3,884 scientists over the complete sample period, separately for patenting and non-patenting scientists.

Figure 3: Distribution of Patenting Activity over Time



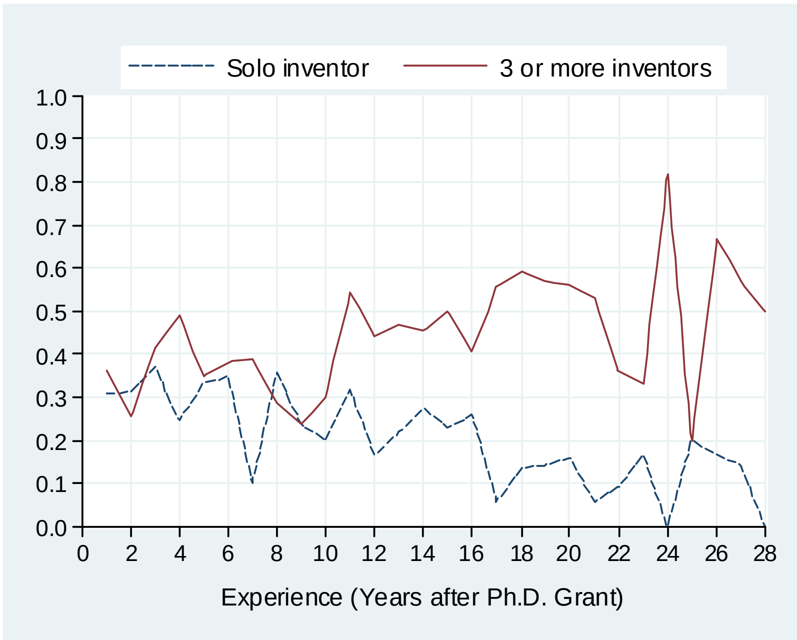
Legend: Figure 3 plots the number of patent applications filed by the scientists in each year and the proportion of scientists that have filed one or more patent applications in each year.

Figure 4: Unconditional Hazard of First Patent Application by Ph.D. Cohort



Legend: Figure 4 plots the kernel-smoothed estimate of the unconditional hazard of first patent application for three cohorts of scientists.

Figure 5: Patent Coinventorship Patterns over Professional Experience



Legend: A "solo" inventor patent is one that lists a single inventor. A "3 or more inventor" patent is one that lists at least three inventors. Proportions represent the percent of all first-time patents applied for by academic scientists in a given post-Ph.D. experience year that belong to either of these two categories.

Data Appendix: Keyword Weights

w^i_{jt}, the patentability weight for each keyword j in year t is defined as:

$$w^i_{jt} = \frac{\sum_{s \in I^p_t - \{i\}} \frac{m_{sjt}}{\sum_k m_{skt}}}{\sum_{s \in I^{np}_t - \{i\}} m_{sjt}}$$

where m_{sjt} denotes the number of times keyword j has appeared in articles published up to year t by scientist s, I^p_t is the subset of scientists in our sample that have already applied for one or more patents as of year t, and I^{np}_t is the subset of scientists in our sample that have not yet applied for any patent as of year t. The weight is also indexed by scientist i, because i's publications are taken out of the set of articles used to compute the formula above.

To create the numerator of w^i_{jt}, we first create a row-normalized matrix with each scientist in the patenting regime listed in a row and each of the keywords used to describe their papers up to year t listed in a column. The sj^{th} cell in the matrix, $[m_{sjt}/\sum_k m_{skt}]$, corresponds to the proportion of title keywords for scientist s that corresponds to keyword j. We then take the column sums from this matrix, i.e., we sum the contributions of individual patenting scientists for keyword j. Turning next to the denominator, we proceed in a similar manner, except that the articles considered only belong to the set of scientists who have not applied for patents as of year t. The numerator is then deflated by the frequency of use for j by non-patenters (in the rare case of keywords exclusively used by patenters, we substitute the number 1 for the frequency).

The weights w^i_{jt} are large for keywords that have appeared with disproportionate frequency as descriptors of papers written by scientists already in the patenting regime, relative to scientists not yet in the patenting regime.

Two things should be noted about the construction of these weights. First, $w^i_{jt} = 0$ for all keywords that have never appeared in the titles of papers written by scientists that have patented before t. Second, the articles written by scientist i him/herself do not contribute at all to the weights w^i_{jt}. Therefore, no scientist can directly influence year-to-year changes in these weights.

The final step for each scientist i in the dataset is to produce a list of the keywords in the individual's papers published in year t, calculate the proportion of the total represented by each keyword j, apply the appropriate keyword weight $w^i_{j,t-1}$, and sum over keywords to produce a composite score. The resulting variable increases in the degree to which keywords in the titles of a focal scientist's papers have appeared relatively more frequently in the titles of other academics who have applied for patents. This score is entered in the regressions to control for the research patentability of scientists' areas of specialization.

To illustrate the construction of the research patentability measure, Table A1 lists some representative keywords, along with their patentability weights in the year 2000. Consider the keyword "Endoplasmic-Reticulum" (italicized in the table) in group 1. In 2000, it had previously appeared 22 times as a keyword in one or more articles of scientists who had patented prior to 2000. Among them is Carlos B. Hirschberg, professor and chair of molecular and cell biology at Boston University, who is listed as an inventor on a patent filed in 1996. To compute the numerator of the patentability weight for this keyword, we begin with the fraction of Hirschberg's research using "Endoplasmic-Reticulum" in the title. In his 29 ISI-listed research papers published between 1970, when he was granted a Ph.D. and 2000, 234 unique keywords have been used a total of 445 times. The word "Endoplasmic-Reticulum" was used 5 times, hence the fraction of Hirschberg's research stock devoted to "Endoplasmic-Reticulum" is 0.011. This procedure is repeated for the other 21 patenting scientists who have used the word. The sum of these fractions taken over all 21 patenting scientists is reported in the second column of the table. Next, to compute the denominator in the above equation, we examine the keywords of all scientists who had not yet received a patent by 2000 for the appearance of Endoplasmic-Reticulum. In the research publications of 3,841 such scientists, this keyword has appeared on 46 occasions. The patentability weight for Endoplasmic-Reticulum is obtained by dividing the sum of proportions of keyword use among patenting scientists (column 2) by the frequency of use for this same keyword among non-patenting scientists (column 3).

Table A1: Sample Title Keywords in the year 2000

	(1)	(2)	(3)	(4)
	Number of times keyword used by patenting scientists	**Sum over all patenting scientists of keyword's proportion of total keywords used**	**Number of times keyword used by non-patenting scientists**	**Keyword weight: Column (2) / Column (3)**
	$\sum_{s \in I_t^p - \{i\}} m_{sjt}$	$\sum_{s \in I_t^p - \{i\}} \frac{m_{sjt}}{\sum_k m_{skt}}$	$\sum_{s \in I_t^{np} - \{i\}} m_{sjt}$	$\frac{\sum_{s \in I_t^p - \{i\}} \frac{m_{sjt}}{\sum_k m_{skt}}}{\sum_{s \in I_t^{np} - \{i\}} m_{sjt}}$
Group 1				
Polyhedrosis-Virus	31	0.144	14	1.030
Trypanosoma-Cruzi	37	0.123	14	0.877
Copolymerization	29	0.067	11	0.612
Schistosoma-Mansoni	44	0.083	17	0.491
Phosphoribosyltransferase	36	0.092	27	0.342
Autocrine	33	0.073	22	0.330
Follicle-Stimulating-Hormone	45	0.061	19	0.323
Etoposide	34	0.052	17	0.306
Atherosclerosis	43	0.101	35	0.289
Methotrexate	73	0.170	63	0.271
Endoplasmic-Reticulum	54	0.114	46	0.247
Antitumor	106	0.220	92	0.240
Integrin	113	0.229	98	0.234
Leukotriene	43	0.078	34	0.231
Monooxygenase	38	0.051	33	0.156
Group 2				
Enzyme	498	1.069	1148	0.097
Escherichia-Coli	552	3.251	1500	0.093
RNA	393	1.383	1097	0.120
Transcription	385	1.795	974	0.120
Receptor	1,328	0.941	3513	0.097
Group 3				
Tropomyosin	7	0.012	62	0.015
Peroxisomal	6	0.023	56	0.023
Aplysia	4	0.015	102	0.026
Photosystem-Ii	4	0.006	80	0.007
Dynein	3	0.038	89	0.062

Legend: To illustrate the construction of keyword weights, we have chosen representative words in three categories. Group 1 keywords are typical of those that appear frequently in the work of patenting scientists, and infrequently in the work of non-patenting scientists. These words receive high patentability weights. Group 2 comprises keywords that occur frequently in the journal articles of both patenting and non-patenting scientists. Words in this group garner intermediate weights. Group 3 contains keywords that are very common in the research of non-patenting scientists but uncommon in the work of patenters. In consequence, these keywords receive low weight.