NBER WORKING PAPER SERIES

ENTRY, EXIT AND PATENTING IN THE SOFTWARE INDUSTRY

Iain M. Cockburn
Megan J. MacGarvie

Working Paper 12563
http://www.nber.org/papers/w12563

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2006

We thank LECG Inc. for support for this research and providing access to data via an unrestricted grant from Microsoft Corporation. We are grateful to Anne Layne-Farrar, Alfonso Gambardella, Daniel Garcia-Swartz, Shane Greenstein, Josh Lerner, Robert Merges, Marc Rysman, Mark Schankerman, Ken Simons, Manuel Trajtenberg, and Sam Thompson for helpful comments. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research..

©2006 by Iain M. Cockburn and Megan J. MacGarvie. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Entry, Exit and Patenting in the Software Industry
Iain M. Cockburn and Megan J. MacGarvie
NBER Working Paper No. 12563
October 2006
JEL No. L1, L6, O34

ABSTRACT

We examine the effects of software patents on entry and exit in 27 narrowly-defined classes of software products, using a dataset with comprehensive coverage of both mature public firms and small privately held firms between 1994 and 2004. Reflecting the complex economics underlying the relationship between patent protection, entry costs and industry structure, we find that patents have a mixture of effects on entry and exit. Controlling for firm and market characteristics, firms are less likely to enter product classes in which there are more software patents. However, all else equal, firms that hold software patents are more likely to enter these markets. The net effect on entry of increasing the number of software patents is difficult to measure precisely: estimates of the effect of an across-the-board 10% increase in patent holdings on the number of entrants into the average market in this sample range from -5% to +3.5%, with quite large standard errors. Evidence on exit and survival is consistent with these findings - holding patents appears to enhance the survival prospects of firms after entering a market.

Iain M. Cockburn
NBER
1050 Massachusetts Ave
Cambridge, MA 02138
cockburn@bu.edu

Megan J. MacGarvie
Boston University
School of Management
595 Commonwealth Avenue, Room 522H
Boston, MA 02215
and NBER
mmacgarv@bu.edu

Introduction

The evolution of the judicial treatment of the patentability of software over the last two decades has led to a dramatic increase in the number of patents on software and in the number of firms seeking patent protection in this area. Critics of the increased patentability of software have argued that these legal changes will stifle innovation and competition by holding up the development of technology that builds on patented prior art and swamping inventors in patent infringement suits. For example, Bessen and Maskin (2002) argue that because innovation in software is sequential and complementary, increased patent protection has led to a reduction in the rate of innovation in software.[1]

Conversely, increased use of patents may lead to greater innovation and competition in software (see, for example, Smith and Mann (2004).) This may happen through obvious mechanisms such as the incentive effect of increased appropriability of returns from R&D, as well as through more subtle mechanisms such as the role of patents as a signal of quality for start-up firms, greater R&D productivity as a result of increased disclosure of useful information, easier access to venture capital or other sources of finance, or more efficient transactions in knowledge in a market with stronger property rights.

Direct evidence on the "stifling" versus "stimulating" impacts of patents on innovation and competition is not easy to find. Some researchers have looked for evidence of the impact of changing patent rights on innovation from quasi-natural experiments associated with legal changes. These changes include the strengthening or instituting of

---

[1] This phenomenon is not unique to the software industry. Heller and Eisenberg (1998) describe the "tragedy of the anti-commons" in biomedical science, in which a proliferation of intellectual property rights on upstream technologies is thought to have the potential to slow the progress of research in that field.

patent systems in countries that previously had weak or non-existent formal IPRs (Sakakibara and Branstetter (2001) for Japan, Lanjouw and Cockburn (1997) for India), changes in patent rights in confined to specific technologies (Scherer and Weisburst (1995) on pharmaceuticals in Italy), or a variety of other changes to patent law or patent office practice that enhance the strength of patent protection (Lerner (2002)).

In software, Bessen and Maskin (2002) and Bessen and Hunt (2003), have argued that more-and-stronger patent rights have induced a decline in R&D spending in software. One of the problems with this type of study is that R&D spending may be a poor proxy for the rate of innovation. Furthermore, R&D spending is jointly determined by competitive interaction, knowledge spillovers, the nature of technological opportunities, and appropriability conditions, as well as constraints imposed by blocking IP. The interplay of these factors has proven prohibitively difficult to model and identify empirically. Cockburn (2005), for example, finds puzzlingly few patents in the economically and technologically dynamic software-intensive field of bioinformatics.

In this paper we therefore focus on a somewhat different indicator of the impact of patents: exit, entry, and industry dynamics. Entry by new firms is closely associated with new product introductions and technological change in many industries, including software.[2] In industries such software, with a rapid underlying pace of technological change and an active entrepreneurial sector, any "blocking" or "stifling" effect of patents

[2] See Graham and Mowery (2003) and Campbell-Kelly (2003) for surveys of the development of the industry, and Prusa and Schmitz (1991, 1994) on entry and new products.

should benefit incumbents at the expense of entrants, and ought to be visible in decreased rates of entry, exit, and turnover of firms.[3]

On the other hand, in "complex" industries like software, patents may increase rates of entry, through mechanisms such as signaling the quality of entrants' technology in a way that improves their ability to gain access to financing (Hall (2004), Mann and Sager (2005)), or by giving entrants quick access to profits through licensing (or being acquired). Hall and MacGarvie (2006), while finding that the market value of "downstream"[4] software producers is negatively affected relative to "upstream" players by the legal decisions expanding software patentability, show that software patents are significantly more valuable than other types of patents for the firms that hold them. A similar result on the value of software patents is found in Noel and Schankerman (2006), which also quantifies the cost to software firms of the potential hold-up problem associated with having to license from many rival patent holders. Mann (2005) argues that not only do patents not impede innovation in software, rather they actually benefit firms that are able to use them in cross-licensing negotiations. Lerner and Zhu (2005) find that the increased use of patents by software firms following the Lotus v. Borland decision was associated with improvements in firm performance (as measured, for example, by the growth of sales). In this regard, the ability to patent software may actually facilitate the entry and growth of new ventures. Cockburn and Wagner (2006), for example, show that internet companies filing patents were more likely to survive the collapse of the dot-com bubble after 2001. Merges (2006) finds evidence that firms have adjusted to the presence of patents, and that

[3] Indeed, Acs and Audretsch (1990) and Geroski (1989) have shown that innovation is positively associated with entry rates across industries. Gort and Klepper (1982) document high entry rates at the beginning of the product life cycle after a new innovation has been developed and introduced to a market.

[4] Where "downstream" refers to applications and services firms, and "upstream" refers to middleware/systems software and hardware.

effort put into acquiring patents correlates with indicators of market success. Indeed, reviewing the experience of the software industry since the late 1980s, Merges concludes that predictions that patents would "kill" the industry by choking off entry by new firms and entrenching large bureaucratic incumbents have proven to be "mostly, but not completely, wrong."

The evolution of software patentability

The law concerning the patentability of software in the United States has evolved through a series of decisions following the passage of the 1952 Patent Act, to the point where algorithms may be patented if "there is practical application for the algorithm or if it is associated with a tangible medium."[5] In 1972, the Supreme Court's ruling in *Gottschalk v. Benson* stated that because software is essentially a collection of algorithms, it could not be patented. However, in 1981 the court allowed for patenting of software tied to physical or mechanical processes, such as the program implicated in the method for curing rubber at issue in *Diamond v. Diehr*. The Federal Circuit stated in 1994 (in *re Alappat*) that unpatentable software was that which represented " a disembodied mathematical concept…which in essence represents nothing more than a 'law of nature,' 'natural phenomenon,' or 'abstract idea.'" Software that could be patented was "rather a specific machine to produce a useful, concrete, and tangible result."[6] A series of decisions in 1994 and 1995 following *Alappat* culminated in a new set of guidelines, issued by the Commissioner of Patents in May of 1996, which allowed inventors to patent any software embodied in physical media.[7]

---

[5] Sterne and Bugaisky, p. 221
[6] *In re* Alappat, 33 F.3d 1526, 1544 (Fed. Cir. 1994), quoted in Sterne and Bugaisky, p. 222
[7] Sterne and Bugaisky, p. 223

Beyond industry-wide impacts of this gradual evolution of legal and administrative doctrine over the 1980s and 90s on patenting, we hypothesize that these changes may have had an uneven impact within the software industry. Some categories of software may have become more easily patentable, or some categories of software patents more easily enforceable, before others. This appears to be reflected in differences in the volume and growth rates of patenting in different patent classes within software during the 80s and 90s (see Figure 1). Obviously, these trends confound changes in patentability and perceived value of patents with technological change in that category, but we believe that cross-category variation over time in the availability and effectiveness of patents driven by exogenous changes in legal doctrine may be an important source of identification of the economic impact of software patents.

Theoretical background

In the literature arising from Gort and Klepper (1982), innovation plays a crucial role in determining the rate of entry and the number of firms in the market. This literature envisages the evolution of a market in the following way. In the initial stage, a ground-breaking innovation emerges and the innovator enjoys temporary monopoly profits that attract entry. Entrants carry out both product R&D, which stochastically results in related innovations or modifications of the new product, and process R&D, which lowers costs. Over time, as the results of the product R&D are imitated, and the process R&D reduces costs and product prices, fewer and fewer entrants are able to realize positive profits and the rate of entry falls while the rate of exit rises.

We speculate that an increase in the strength of software patents may affect entry and exit over the product life cycle in (at least) the following ways. (1) It may lengthen the first stage in which the initial innovator enjoys a monopoly. (2) It may reduce the rate of entry, as imitation becomes more difficult, for example if blocking IP makes sequential improvements more costly. (3) By reducing uncertainty about the entrants' quality and acting as a signal to venture capitalists, patents may result in better screening of potential entrants by venture capitalists thus raising the survival rate of the firms that do enter.

A challenge of identification arises from the fact that Gort and Klepper document a surge in the rate of patenting in a market in last stage of the product cycle, when entry is low.[8] This observation may be less relevant in fast-moving technologies with short product lifecycles, the current economic environment, where patenting strategy is increasingly directed towards "upstream" and basic technology, and firms are under increasing pressure to patent "early and often." Nonetheless, if we observe a negative correlation between patents in a market and the rate of entry, it is therefore possible that we are picking up spurious correlation associated with the stage of the product cycle.

To address this issue, we control of the stage of the product cycle using the modal citation lag to patents in the SOF class as an indicator of the maturity of the technology.[9] Since the number of citations to a patent is a function of the number of potential citations, we estimate the modal lag using a framework that adjusts for this effect. For each SOF class and citing-cited year pair, we compute the citation frequency, or ratio of actual to

---

[8] Gort and Klepper also show that the rate of *innovation* is "at variance with the trends in patenting", and argue that the increased rate of patenting in the final stage of the product cycle reflects strong incentives for innovative effort arising from continued growth in the size of the market, while the actual success rate of innovative effort has declined (p. 648-650).

[9] See Adams, Clemmons and Stephan (2006), who use the modal citation lag to study the rate of diffusion of scientific knowledge.

potential citations (see Jaffe and Trajtenberg (1999)), and then identify the citation lag (citing year – cited year) with the highest citation frequency for a given SOF and citing year. [10] If the modal lag in a product category is short, it implies that the most highly cited patents in that market were granted recently, which suggests that the market is at a relatively early stage of the product cycle.[11] Table 2 lists the mean modal lags by SOF class for 1994-2004.[12]

Another theoretical model that informs our work is that of Gilbert and Newbery (1982), which examines the role played by patents in pre-empting entry and extending monopoly power. Gilbert and Newbery describe a monopolist's incentive to pre-empt entry by filing patents, including "sleeping patents" on inventions that are not commercialized. Gilbert and Newbery state that "preemption would be very hard to identify … because it is difficult to distinguish product development that is the result of superior foresight and technological capabilities from development that is motivated by entry deterrence."[13] One of the novel features of this paper is that it takes advantage of the shifts in software patentability in the mid-nineties to identify preemptive patenting. Because software was not explicitly patentable at the beginning of our sample period, it is likely that a significant share of the observed growth in software patents represents a "filling in" of the intellectual property landscape that goes beyond the increased numbers

---

[10] We compute the citation frequency as the ratio of the number of observed citations to the number of potential citations. That is, if Ckgd is the number of citations made to patents in SOF class k in citing year g to patents granted in SOF class k in cited year d, Pkg is the number of patents granted in class k in year g, and Pkd is the number of patents granted in class k in year d, the citation frequency is Ckgd/(Pkg Pkd)

[11] The usefulness of this variable as an indicator of the stage of the product cycle obviously depends on the assumption that the key inventions are patented, or at least that the patented inventions

[12] Another approach is to try to identify stages of the product lifecycle through turnover rates. We experimented with this by computing SOF-year turnover rates (using the formula [|entrants-exits|/(0.5(entrants+exits))] and estimating models on subsets of the data broken out by bottom quartile, middle-two quartiles, and top quartile of turnover rates. We did not find evidence that the effect of patents on entry changes monotonically as the rate of turnover changes. The patent coefficient was quite stable across the three groups: -1.07 (2.76), -0.51 (0.24) and -1.34 (1.07).

[13] Gilbert and Newbery (1982), p. 525

of patent filings driven by changes in the pace of innovation or in the technological capability of firms.

We further attempt to disentangle pre-emptive or strategic patenting from patenting that reflects the technological capabilities of incumbents by distinguishing between the total number of patents filed by firms in a particular software market and the quality of those of patents. A market with many patents could be a market in which incumbents are very innovative, or a market in which incumbents work very hard to protect intellectual property—or both. In all three cases, if entry is deterred by incumbents' behavior, then we will observe a negative correlation between patenting and entry rates. However there are very different implications of such a finding for social welfare and the pace of technological change. Purely strategic patenting that shelters incumbents from price competition and does not reflect any underlying technical change presumably lowers welfare. At the other extreme, patents that reflect innovation by incumbents may well be associated with higher welfare if dynamic gains from innovation offset static welfare losses from higher prices.[14]

Though computing welfare gains and losses is beyond the scope of this paper, we may nonetheless be able to identify which effect dominates by controlling for average patent quality. Suppose there was no "real" underlying technological progress, but incumbents nonetheless obtained a large number of patents. We would then observe a large number of patents, lower average patent quality, and lower entry. On the other hand, if there was a high rate of innovation by incumbents, we would again see a higher level of patenting and lower entry rates, but no diminution of patent quality. By controlling for

[14] A potential countervailing effect on welfare in this particular industry could result from excess product differentiation in the presence of network effects.

patent quality we can distinguish between markets with a lot of patents and markets with important patents, and arguably a finding of a negative association between entry rates and patenting after controlling for patent quality is consistent with some degree of strategic patenting.

We measure the average quality of patents in a market using the mean number of citations received by those patents, which is commonly interpreted as an indicator of patent value or importance.[15] There are some obvious problems with this measure. The number of citations received may reflect the legal "size" of the patent in terms of the scope of its claims and the extent to which its disclosure of the invention constitutes prior art against future applications, rather than its quality in the sense of technological significance. More seriously, the ex post number of citations received is a function of the size of the pool of potential citing patents and the patenting strategy of subsequent applicants, which are likely to be endogenous to entry into the market in question. This hampers our ability to draw "clean" inferences about patent quality from this measure, but unfortunately there are no other easily available indicators.

Empirical literature on entry

The empirical literature on entry has focused on the roles of three main factors in influencing entry: a) demand, b) competition, and c) entry costs. Our focus here will be on c), or whether differences across software markets in the extent of patenting are associated with differences in entry costs.

Cross-industry comparisons of entry rates have yielded several interesting findings (see Geroski (1995) for a discussion). Dunne et al. (1998) contains estimates of entry rates

[15] See Lanjouw and Schankerman (2004).

averaging between 41.4% and 51.8% over five-year census periods for a panel of US industries between 1963 and 1982. Within-industry variation in entry rates appears to dominate between-industry variation[16], and entry takes place in waves, with the highest rates occurring at the beginning of a product cycle (Gort and Klepper (1982)). Not surprisingly, high entry rates are associated with high rates of innovation (Acs and Audretsch (1990), Gort and Klepper (1982), Geroski (1989)).

Many of the more recent papers on entry (i.e.: Berry (1992), Bresnahan and Reiss (1991), Mazzeo (2002), Seim (2004), Toivanen and Waterson (2005)), estimate the parameters of structural models and focus on clearly defined, often isolated markets in which firms offer homogeneous products and there is a clear set of potential entrants. Most of these conditions do not apply in the software industry, which is characterized by differentiated products with high development costs and in which the set of potential entrants is not clearly defined. Greenstein and Wade (1998) study entry, exit and the product cycle in the commercial mainframe computer market, and our empirical approach is closely related to this paper as well as to Scott Morton (1999), which analyzes generic entry in pharmaceuticals and Kyle (2006), which studies international entry patterns in pharmaceuticals. Kyle's model has the advantage of treating development costs as sunk with respect to entry, whereas in our model the cost of product development is an important determinant of the entry decision. Specifically on software, Giarratana (2004) provides a detailed case study of entry in encryption software, including the role of patents in facilitating trade in technology.

Note that unlike much of the prior literature on entry, which focuses quite generally on the determinants of market structure and the welfare effects of entry, we are more

[16] Geroski (1995), p.423.

narrowly focused here in testing for any association between intellectual property rights and the pace of innovation, where cross-market entry rates are used as a proxy for the pace of innovation. This reduced form approach frees us from making strong assumptions about a very complex institutional setting but limits the conclusions that can be drawn from our estimates. Our models do not, for example, deal explicitly with strategic interaction between firms, and may therefore be a poor basis for estimating a behavioral response to policy shocks.

Data

Studies of software patenting activity and changes in legal doctrine based solely on the COMPUSTAT universe have a serious selectivity problem: larger incumbent firms' reactions to the evolution of software patent law are likely to differ significantly from those of new ventures or smaller start-ups. Furthermore, they fail to capture a very important source of innovation and competition: new entrants and new products. One of the advantages of this paper is that we use a dataset with comprehensive coverage of both large, mature, public firms and small privately held firms, over a relatively long time span, which offers a uniquely comprehensive window on competition and industry evolution in software. Our data is based on the CorpTech directory of technology companies, which covers 19,717 software companies over the period 1990-2004.[17] We have detailed information on the product classes in which these firms are active, which will form the basis of our analysis of entry into new product areas. We also know the founding date of the firm, revenues and employment for most (but not all) of the firms in the dataset, the

---

[17] We define software companies as the firms listed in CorpTech as having at least one product classification beginning with "SOF", which is CorpTech's code for software.

patents held by the firm, information on corporate parents, funding sources, and a number of other variables.

One of the key CorpTech variables for the purposes of this study is the "SOF category" (or categories) assigned for each firm, which indicate the software product markets in which a firm is active at a very detailed level. This is a self-reported variable, and can include products under development as well as products already launched. CorpTech reports more than 290 SOF categories, however many of these are quite vaguely defined, or appear to be defined in terms of customer segments rather than in terms of a technology—e.g. "secondary school software, dental practice management software, etc." Furthermore mapping patents to markets is a challenging and resource-intensive task. We therefore focus our analysis on 27 of these SOF-defined markets.

These 27 markets were chosen primarily on the basis of our assessment as to whether the technology/product is reasonably distinctive, and we could define a set of keywords that could be fruitfully searched in the abstract of patent documents. Clearly there is some potential for selection bias to influence our results, however we believe that the criteria used to choose these markets are independent of entry and exit dynamics and the sample of 27 SOFs does not appear to be markedly different from the other 262 in terms of firm characteristics and entry and exit rates (see the Appendix). One area in which our sample differs, however, is in terms of the average number of patents held by firms active in the market. The average firm active in one of the sample markets has 29 patents, while the average firm in a market omitted by the sample has only 18 patents, and this difference is statistically significant. Note though that this difference arises by construction: it is difficult, if not impossible to identify patents related to many of the more vaguely defined markets. In our judgment, therefore, this subset of markets is reasonably representative of software products in general.

The markets that we consider are listed in Table 2, along with counts of numbers of incumbents and entrants. The Appendix describes the process by which we selected these categories.

Identifying entry and exit

We identify entry and exit in these markets from information in a dataset based on surveys conducted by CorpTech.[18] This dataset lists firms active in each SOF-defined market every second year from 1990 to 2004, and presence or absence of a firm in these lists allows us to We classify a firm as an entrant if the firm has products in a SOF category after two consecutive sample years (4 years elapsed time) of not having products in that class, or is born no more than two years before its first appearance in the dataset. We define an exit as occurring when a firm that operated in a SOF class for two consecutive sample years exits the class in the third sample year or is dropped permanently from the database. Some summary statistics on entry and exit, by product class are found in Table 2 and Figure 2.[19]

Patents and Entry Costs

Patents that block a would-be entrant from producing or selling its product may be a significant barrier to entry. The entrant must either bear additional costs of "inventing

[18] We thank LECG, Inc. for providing access to these data.

[19] One potential pitfall arises because some firms enter CorpTech several years after their founding dates, and we thus do not observe their entry. However, only a relatively small number of these firms actually enter during the period under consideration (1994-2002). We omit SOF codes in which the number of missed entries during the period is more than one standard deviation above the mean. The average number of missed entries across the categories (calculated as the share of firms that are founded after 1990 but do not appear in the sample until more than 2 years after their founding date) amounts to 12.5% of entries, and the standard deviation is 10.08.

around" such patents, pay licensing fees to the patent holder, or accept potentially severe ex post penalties.[20] However, constructing an indicator of the significance of patents for costs of entry to a software market, or even a simple count of the number of patents covering the technologies relevant to that market, is no simple task. The Appendix to this paper describes the resource-intensive process used to match USPTO patent classifications to the CorpTech SOF categories. In short, we used a combination of text searching and reading the manual of patent classification to identify the set of key patents associated with each market. We then estimate the relationship between the rate of entry and the number of patents in the market.

Note that it may not be just the absolute *number* of patents in an area that can deter entry, but also the extent to which those patents form a "thicket". According to Shapiro (2001), "a patent thicket is a dense web of overlapping intellectual property rights that a company must hack its way through in order to actually commercialize new technology. With cumulative innovation and multiple blocking patents, stronger patent rights can thus have the perverse effect of stifling, not encouraging, innovation."[21]

We experiment with two alternative measures of the effects of patent thickets. One of these is intended to capture the number of holders of potential blocking patents, i.e. the number of potential licensors with which an entrant would have to negotiate. The other seeks to measure the concentration of patent rights in a field. Both measures are computed using patent citations.

---

[20] Infringing valid patents can present the entrant with very substantial ex post penalties, such as damages judgments (tripled in the case of "willful infringement") or the loss in value of assets stranded in the wake of an injunction obtained by the patent holder.
[21] P. 2

Patent citations are references to “prior art”, or existing patented technologies, listed in the patent document.[22] Since citations delimit the property rights represented by a patent by describing related claims contained in other patents, citations made by a patent give an indication of the extent to which a technological area is already covered by intellectual property rights and thus foreclosed to entrants who do not obtain a license. The factors determining the total cost of obtaining licenses to allow entry are complex. All else equal, we expect that the more patents that must be licensed, the higher the total cost of entry. However, particularly in complex technologies, patents are frequently bundled or pooled or jointly licensed, thus total costs of entry may not have a simple linear relationship to the number of patents blocking the would-be entrant. Another salient feature of “thickets” is the higher costs associated with negotiating with many parties. To the extent that there are fixed costs of conducting a negotiation, having to deal with more parties will drive up costs. In addition, the outcome of a complex bargaining process conducted with many licensors, each of whom has some holdup power, may result in higher total costs—i.e. the height of the “royalty stack” may rise non-linearly in the number of its components. To capture these effects we therefore use the number of patent-holders cited by patents in a SOF class (the number of cited assignees) to proxy for the number of potential licensors. We hypothesize that as the potential number of licensors with which an entrant would have to negotiate increases, the costs of entry will increase. Note however that the number of licensors may not be the only determinant of entry costs: one offsetting factor may be the concentration of ownership of intellectual property. For a given number of licensors, we hypothesize that entry costs will be lower when the

[22] “Prior art” is not confined to patents, indeed most forms of printed publication describing the claimed invention can constitute prior art, as can public knowledge, use, or sale of the technology.

ownership of patent rights is more concentrated, because there are fewer parties with whom to negotiate.[23]

We follow Noel and Schankerman (2006) in our construction of measures of the concentration of ownership of intellectual property in a market. Assuming that the share of citations received by an assignee proxies the importance of negotiating with that assignee, we postulate that in a market which has many cited assignees but where citations go disproportionately to a small number of firms, entry costs may actually be lower than in a market with fewer assignees each of which receives a similar share of total citations. To capture this effect, we calculate the four-assignee concentration index for citations for each SOF in each year.[24] Because our citation counts are truncated, we apply a truncation correction by weighting citations based on the estimates for computer-related patents in Hall, Jaffe and Trajtenberg (2005). The citation weights can be found in the Appendix.

Descriptive Statistics

Tables 1, 2 and 3 provide descriptive statistics for the data set. The average SOF-defined market in this sample had 148.2 active firms. The average number of incumbents per market grew steadily from 74.4 in 1994 to 201.9 in 2004, while the average number of entrants varied from 7.4 to 22.0 during the sample period. Our measure of the maturity of the technology, the modal citation lag to patents in the market, varied across markets from an average of 4.33 years to 7.17 years. Figure 2 plots the "turnover" in these markets over time, showing considerable turbulence, though the "classic" patterns of industry dynamics

---

[23] An alternative hypothesis is that when citations are more concentrated, the firms holding cited patents have greater bargaining power, which would increase entry costs.

[24] We also experimented with using the Herfindahl index of citations across assignees to measure concentration of patent rights, but obtained very similar results to those based on the four-assignee concentration index. These results are available from the authors upon request.

found by Klepper and others are difficult to discern. The number of patents related to each market ranged from an average of 9.31 over the sample period to almost 4000, with an average over the whole sample of 1501. However, patent holding was highly concentrated with only 0.7% of firms having any patents relevant to a given market in a given year, and the average firm in the sample holding a stock of 0.08 patents in a given market. Figures 3 and 4 display the share of firms holding patents in a market and the average stock of patents held by type of firm (incumbents, entrants, and non-entrants). Interestingly, the fraction of incumbents that hold patents in a market is quite similar to the fraction of entrants that hold patents, while firms that never enter a market almost never hold patents in that market. Our measures of bargaining costs are quite interesting. The number of cited assignees per market averages 696 with a high of 2307, and a low of six. Clearly, the average potential entrant is very unlikely to have to obtain licenses to 1501 patents from 696 different entities—only a small fraction of the total number of patents that we have identified as being relevant to a market will be applicable to a specific product. But these figures are consistent with anecdotal evidence that in complex technologies, clearing a product for launch can entail reviewing thousands of patents.[25] The CR4 ratio over assignees of citations to patents in the average market was 0.46, ranging from 0.21 to as high as 0.94, confirming that patent holdings are highly concentrated.

Empirical Approach

In keeping with the theoretical and empirical literature on entry, we expect the probability of entering a market to be decreasing in the number of competitors and increasing in demand. Our hypotheses about cross-market variations in the *costs* of entry

[25] Based on conversations with various corporate patent counsel.

can be stated as follows: (1) entry is decreasing in the number of patents in a market, (2) entry is increasing in the number of patents held by entrants, and (3) entry is decreasing in the cost of bargaining with rival patent holders.

Market level entry model

We do not have information on which firms are potential entrants into which markets. As a result, in the firm-level specification that is described below, we treat all firms in our sample that have not previously entered a market as potential entrants to that market. To begin, however, we present results based on data aggregated to the market level in which we examine the association between the total number entrants to a market and the extent of patenting in that market. The advantage of this specification is that it does not rely on the assumption that all firms are potential entrants. The disadvantage is that it does not allow us to compare the differential effects of software patents on firms with different characteristics, and does not allow us to control for firm characteristics that are correlated with entry such as firm size and age. We also estimate firm-level models and describe them below. Perhaps the paper closest to ours in terms of empirical methodology is Greenstein and Wade (1998). We follow their approach of estimating market-level entry models using econometric techniques for count data and firm-level exit regressions using hazard models.

Table 4 contains the results of a conditional fixed-effects Negative Binomial regression in which the dependent variable is the number of entrants in market j in year t, and the explanatory variables include the number of incumbent firms in the market and the number of incumbents squared, plus the growth in revenues in the market over the previous two years.[26] To control for the stage of the product cycle, we include a set of dummies for each decile of the modal citation lag of patents granted in the SOF class.[27] Our key right-hand-side variables are the log of the number of patents in the market[28], the log of the mean number of forward citations received by patents in the market, the log of the number of patents held by entrants, and the log of the mean number of forward citations received by patents held by entrants. Incumbent and entrant patents are measured as stocks, computed from annual flows of issued patents by application date using a declining balance formula with a 15% depreciation rate. The standard errors are clustered by market, and market fixed effects are included in all regressions.

Table 4 shows that after controlling for demand with the growth of revenues variable, the number of incumbents enters our model with a positive sign, and the number of incumbents squared has a negative coefficient. Both are significant at the 5% level in specifications (1) and (2). Thus, when the number of firms in the market is small, increases in the size of the market are associated with increases in entry—presumably reflecting a reduction in the market power of incumbents or a reduction of barriers to entry created by network effects. For markets with larger numbers of incumbents, however, increases in the number of incumbents reduce the probability of entry, which presumably reflects the fact that large numbers of incumbents indicate more mature, more crowded, and less attractive markets. Our measure of the growth rate of revenues in a SOF is

---

[26] The revenue variable with the fewest missing values in the CorpTech data is a categorical variable that indicates the range in which the firm's revenues fall. We added up these categorical variables for all the firms active in a market.

[27] Dummies for the third. fifth and seventh deciles are omitted because they are collinear with the dummies for the second and fourth deciles (respectively). The value of the modal lag at the second and third decile is 4 years, the value at the fourth and fifth deciles is 5 years, and the value at the sixth and seventh deciles is 6 years.

[28] This is the total number of patents relevant to the market, as defined by the concordance of patent classes to SOF classes found in the appendix. While these patents are held by a set of firms that certainly includes the incumbent firms in the market, they may also be held by firms that are not active in the market.

positively and significantly associated with entry. The modal citation lag coefficients display a pattern in which there is an initial increase in the rate of entry as the modal lag increases, followed by a decrease and then an increase in the coefficient when the modal lag becomes very long. This pattern is particularly evident once we have controlled for the number of cited assignees (which would be expected to confound the estimated effect of the modal citation lag because older markets have more cited assignees on average), in column 4. These coefficients appear to more or less trace out the expected relationship between the rate of entry and the product cycle—an initial increase in the rate of entry in early stages of the product cycle, followed by a decrease as the market matures. The increase in the coefficients at very late stages of the product cycle presumably reflects markets in which there are very few patents relevant to the current state of technology (or the key blocking patents have expired) and as a result entry is easier.

The log of the number of patents in a market is negatively and significantly associated with the rate of entry after controlling for the average quality of the patents in the market (in the form of the mean number of citations in the market). These effects are economically as well as statistically significant. Ceteris paribus, a ten percent increase in the number of patents in a market is associated with approximately a 3-5% decrease in the rate of entry (see columns 2-5 of Table 4). Consistent with the "bargaining chip" and "quality signal" hypotheses, the log of the number of patents and the average quality of the patents held by entrants are positively associated with the rate of entry. A ten percent increase in the number of patents held by entrants is associated with approximately a one percent increase in the rate of entry. After controlling for patent quality, the *number* of patents held by entrants is significant only at the ten percent level (however, the Wald test of the joint hypothesis that all the patent and patent quality coefficients are zero is rejects the null at the 1% level).[29] The average quality of incumbent's patents has a very large and strongly significant negative effect on entry, with estimates of the point elasticity ranging from 0.6 to 0.8. Higher numbers of citations suggest "larger" patents, which are more difficult to invent around, and more significant innovation by incumbents, both of which will tend to deter entry. Conversely, the average quality of entrants' patents has a positive effect on entry: entrants with higher quality patents may find it easier to bargain their way into the market.

In column 4 we include a quadratic in the number of cited assignees in a SOF[30], a proxy for the costs of bargaining with rival patent holders. The quadratic is used because we believe that the relationship between entry and the number of cited assignees is unlikely to be linear. Holding constant other market characteristics, markets with very few cited assignees are likely to be markets in which the existing IP is especially good at blocking entry, and thus increases in the number of assignees (corresponding to more fragmented and therefore weaker IP) will be associated with increases in the rate of entry. Beyond some point, however, any benefits to a prospective entrant of fragmented rights will become fully offset by increases in bargaining costs, and larger number of cited assignees will be associated with reduced entry. The estimated parameters are jointly significant and imply a concave relationship between entry and the number of cited assignees, increasing up to and past the sample mean of cited assignees per patent (696) and reaching a maximum at approximately 918. For markets with relatively low numbers of cited

[29] We also estimated a specification including the interaction of the market's patents with entrants' patents, but the coefficient on this variable was statistically insignificant and the specification was rejected by a likelihood ratio test in favor of a specification without the interaction term.

[30] In the reported regression results, the number of assignees is divided by 100 to reduce the number of decimal places in the reported coefficients.

assignees, greater fragmentation has a substantial positive effect on entry: for example the estimated coefficients imply that a market with only 200 cited assignees would see the number of entrants rise by 1.2% if the number of cited assignees increased 10%. For markets with large numbers of cited assignees, increases in this variable have a substantial negative effect on entry. At 1000 cited assignees, an increase of 10% would reduce entry by 1.1 percent, ceteris paribus. An increase in cited assignees from the sample average to the sample maximum would cause entry to drop by over 54%.

Puzzlingly, the coefficient on the four-firm concentration ratio of citations in that market, our other measure of bargaining costs, is not statistically significant.[31] This may reflect the countervailing effects of concentration discussed above, that is, a reduction in bargaining costs as citations become more concentrated, accompanied by higher bargaining power on the part of cited assignees.

There may be concern that the number of patents in a market is endogenous with respect to the number of entrants. If incumbents are most likely to use patents as a deterrent when the threat of entry is strongest, and if we have not fully measured the threat of entry, our estimate of the effect of patents on entry could be biased upward (that is, it could be less negative than it would be if we could control for the threat of entry). We deal with this possibility using instrumental variables to identify the causal effect of patents on entry at the market level.

To estimate the parameters of the market-level entry model with instruments, we used a moment-based count data model in which the ratio of within-group means is used to

---

[31] We tried adding the concentration ratio squared, but neither CR coefficient was significant, nor were they jointly significant.

approximate a market-level effect.[32] This model is as follows, where y_{it} is the number of entrants to market i in year t:

$$y_{it} = \mu_{it} \frac{\bar{y}_{it}}{\bar{\mu}_{it}} + u_{it},$$

Where $\mu_{it} = \exp(x'_{it}\beta)$. We use GMM to solve the moment conditions:

$$\sum_{i=1}^{N} \sum_{i=1}^{T} x_{it} \left(y_{it} - \mu_{it} \frac{\bar{y}_{it}}{\bar{\mu}_{it}} \right) = 0$$

Because of the high level of serial correlation in the log of the cumulative stock of patents in a market in year t, the models perform better when we use the log of the number of patents *granted* in that year as our key RHS variable. Our first instrument is the number of non-software[33] patents granted in year t that are held by the assignees holding patents relevant to the SOF. This instrument is designed to pick up the fact that if a firm has experience navigating the patent system for technologies other than software, that firm may be more likely to patent software inventions. If a market is populated by several such firms, there will be more patents in the market for reasons unrelated to the threat of entry. The second instrument is the number of incumbent firms in the market with a primary SIC outside software, designed to capture the fact that software embedded in hardware was considered patentable earlier on, which should be associated with higher patent counts and is not obviously correlated with the threat of entry. The third instrument is a qualitative variable that captures variations in the legal strength of patent protection in software over

[32] This, the GMM counterpart to the Hausman, Hall and Griliches (1984) fixed-effects Poisson estimator, is the "mean scaling model" described in Blundell, Griffith and Windmeijer (2000), p. 5. We thank Bronwyn Hall for the use of TSP code used to estimate these models.

[33] We begin with the Graham-Mowery definition of a software patent, that is, patents in International Patent Classes G06F, G06K, and H04L. In order to be conservative about what we treat as a software patent, we augment this definition by dropping all of the 7XX USPTO classes. See Hall and MacGarvie (2006) for a discussion of different definitions of software patents.

time. It is equal to zero for all markets in 1994, reflecting the fact that the USPTO did not issue new guidelines on the patentability of software until 1996, when the dummy takes on a value of 1 for most markets for the rest of the sample period. It then takes on a value of 2 after 1998, reflecting the importance of the *State Street Bank & Trust Co. v. Signature Financial Group* ("State Street") decision in 1998.[34] We are not aware of a test for weak instruments in the non-linear GMM setting. Instead, we use the "rule of thumb" F-test statistic of 10 proposed by Staiger and Stock (1997) for instrument validity in two-stage least squares. The F-statistic on our instruments' coefficients is 26.33 in a first stage regression in which the dependent variable is the market's patents.[35] When the quality of the market's patents is the dependent variable, the F statistic is 301.46.

Columns 1-3 of Table 5 contain the results from estimating these models. The first column is from a specification in which all of the right-hand side variables in the entry equation are treated as exogenous, and the second is from a model in which the number of patents in the market is instrumented. The third column contains results obtained when both the market's patents and the quality of those patents are instrumented.

The results are (not surprisingly) close to those of the fixed effects negative binomial model in Table 4.[36] Instrumenting for the number of patents in a market actually increases the coefficient slightly, from -0.513 to -0.501, both of which are significant at the 1% level. When the quality of the market's patents is also instrumented, the coefficient falls again slightly to 0.569. The test of over-identifying restrictions passes.

---

[34] See *"State Street" Decision Causes "Boom" in Software Patent Filings*, http://library.findlaw.com/1999/Mar/1/128488.html (accessed Aug. 30, 2006). The patent in question covered a software system for mutual fund management, and though often discussed in the context of business method patents, this decision was viewed as a major expansion of patent protection for software in general.

[35] All pre-determined variables, including year and market fixed effects are included in this first stage.

[36] For ease of estimation, we include the modal lag and the modal lag squared instead of a full set of dummies for all the deciles of the modal lag.

When there is feedback from the dependent variable to the independent variables, however, the parameters of this mean scaling model are inconsistent (see Blundell et al (2000)). In this setting, the concern is that period *t*'s entrants may increase the stock of patents in a market in period *t+1*. This does not necessarily cause substantial bias in the estimates. Here, period *t*'s entrants make only small contributions to the total patent stock in a market in period *t+1*, because the average number of patents contributed by an entrant is small relative to the total number in the market (less than 0.1% on average). However, we are aware that this potential exists, and accordingly we also present estimates in Table 5 of the quasi-differenced model suggested by Chamberlain (1992) and Wooldridge (1997).

This model can be used for consistent estimation of the parameters in the presence of inter-temporal feedback between regressors and the dependent variable[37] and solves the moment conditions:

$$-\sum_{i=1}^{N}\sum_{i=2}^{T} x_{it-1}\left(y_{it-1} - \mu_{it-1}\frac{y_{it}}{\mu_{it}} \right) = 0$$

The results from estimating this model are found in columns 4-6 of Table 5. As was found with the mean scaling model, the coefficients on the market's patents are little changed by the use of instruments. They remain negative and statistically significant, suggesting that a ten percent increase in the number of patents in a market is associated with approximately a 3% decrease in the number of entrants. However, the hypothesis of over-identifying restrictions cannot be rejected at the 5% level, which means that we could

[37] Blundell et al. (2000) note that a problem of the quasi-differenced estimator is that when the variables of interest are highly persistent and lagged values are used as instruments, these instruments can be quite weak. As a solution to this problem, Blundell et al. propose an estimator which incorporates pre-sample information on the dependent variable into a linear feedback model. For our purposes, however, this estimator is not useful because we do not have pre-sample information on entry. Furthermore, our sample period is relatively short, especially when using lagged values of the variables, so using the 1994 entry counts as the pre-sample information would reduce our sample period to only four years: 1998, 2000, 2002 and 2004.

expect to obtain different parameter estimates if we varied the instruments that we use. However, we fail to reject the hypothesis even when we do not use our additional instruments (i.e., when we use only lagged values of the RHS variables as instruments), leading us to conclude that the reason the test fails is related to the lagged variables and not our additional set of instruments. (Column (7) of Table 5 gives fixed-effects negative binomial estimates for the same equation, to allow easy comparison.)

Net effect of patents on entry

While the marginal estimated effects of the patent variables are interesting, the net impact in changes in the overall level of patenting on entry may be more relevant to the debate about software patents and innovation. We therefore computed estimates of the net effect on entry under a variety of scenarios.

First, suppose the number of patents was increased by 10% across the board, holding constant the quality of patents in a market and the quality of patents held by entrants (and all other explanatory variables). The estimated coefficients in column (3) of Table 4 imply a net effect on entry of -2.9%, with a standard error of 2.2%.[38] Of course, without having estimated a structural model of patenting behavior and entry decisions, these reduced form coefficients can easily be misinterpreted. For example if this increase in patents represents a pure increase in the propensity to patent rather than an increase in innovation, it is unlikely that the average quality of patented innovations would remain constant, and as a result the net decrease in entry rates implied by these regressions may be an overestimate. As an extreme example, suppose the number of patents held by both

[38] This estimate is based on the results in column 3 of Table 4. A doubling of the market's patents leads to a decrease in entry of 37%, and a doubling of the patents held by entrants leads to an increase of 10% in the number of entrants. The standard errors described here are computed using the delta method (via STATA's "nlcom" command).

entrants and incumbents increased by 10%, but the additional patents are of no technological significance and receive no citations. Then, basing our projection on the coefficients in column 3 of Table 4, our point estimates imply an *increase* of 3.5% in the number of entrants (with a standard error of 2.4%). Here, the positive effect on entry of a decrease in the average quality of incumbents' patents overwhelms the negative net effect of increased numbers of patents. However, a more realistic scenario might be one in which the number the new patents receive only half as many citations as the existing patents. Under this scenario, our estimates imply that an across-the-board increase of 10% in the number of patents would result in a much smaller impact on the number of entrants (0.2% with a standard error of 1.2%).

It is also worth noting that our model will give different predictions if the increase in patents is asymmetric for incumbents and entrants. Since it is almost certainly easier for a small entrant with a handful of patents to double its patent stock than it is for e.g. IBM to do the same, a more realistic scenario to consider may be one in which entrants' patents increase by a larger percentage than the total number of patents in the market. As an example, suppose incumbents' patents are increased by 10% while entrants' increase by 20%. Holding patent quality constant, the net effect on the number of entrants is 2.2% with a standard error of 2.5%.

Repeating these thought experiments using the IV estimates in Table 5 gives similar results, though with slightly larger estimated net effects and somewhat smaller standard errors allowing us to reject the null hypothesis that the net effect in zero. Taking coefficient estimates from columns 3 and 6 of Table 5 and the associated variance-covariance matrices, we find that a 10% increase in the number of patents held by both entrants and incumbents, with no change in the mean number of citations per patent, leads to a 2.5% reduction in the number of entrants predicted by the quasi-differenced model (which, with a standard error of 1.2% is significantly different from zero at the 5% level) and a 5.0% reduction in the number of entrants predicted by the mean scaling model (with a standard error of 1.1%). As before, if the additional patents are of zero quality, the estimated net effect on entry is positive or zero in these models, however a 10% increase in patents with the additional patents receiving 50% fewer citations is associated with a 1.4% reduction in entry in the quasi-differenced model (with a standard error of 1.1%) and a 1.9% reduction in entry in the mean scaling model (with a standard error of 1.3%).

Firm level entry model

Having established these general patterns, we turn to the firm-level model to examine differences in the effects of patents on different types of entrants. We follow closely the empirical models developed by Scott Morton (1999) and Kyle (2006), and describe the probability that firm *i* enters market *j* as depending on potential profits Π_{ij} obtained from entering the market:

$$\Pi_{ij} = v_j + \phi_{ij}$$

Where v is variable profits and ϕ_{ij} is the fixed cost of entry. We assume that firm *i* will enter market *j* if the cost of entry is below some cutoff value F_j*, so that

$$Pr(enter_{ij}) = pr\ (F_j^* - \phi_{ij} > 0)$$

We expect the probability of entry to depend on the following factors : market-specific factors affecting F_j* which include demand and competition, denoted Z, firm-specific determinants of the cost of entry X, and market-specific determinants of the cost of entry M. We express the probability that firm *i* enters market *j* in year *t* as follows:

$Pr(enter_{ijt}) = Z_{jt}\delta + X_{ijt}\beta + M_{jt}\gamma$

Scott Morton (1999) estimates a probit model in which the dependent variable is equal to 1 if a generic drug producer enters a market, and 0 otherwise, and the independent variables include estimates of the size of the market and firm characteristics related to entry costs. Kyle (2006) estimates a discrete-time hazard model of the decision to enter foreign pharmaceutical markets as a function of market size and firm, drug, and market-specific characteristics. Here we present results from a continuous time Cox proportional hazard model. Results from a discrete-time hazard model are also included, and they are robust to the choice of functional form.

Our dependent variable $Enter_{ijt}$ equals 1 in the year that the firm enters a market, and 0 before. Firms are dropped from the regression once they have entered a market. The dataset is therefore an unbalanced panel, with 58,037 firm-year combinations and 27 markets, for a total of 1,566,999 observations.[39] Following Berry (1992) and Scott Morton (1999), we begin by treating all the software firms in our sample that have not previously entered a market as potential entrants. This is likely to be quite a strong assumption, and so we also experiment with more restrictive definitions of the set of potential entrants, for example by defining potential entrants as those firms that have not previously entered an "adjacent" market. We define adjacent markets as being in the same broad SOF category (for example markets "AI_N: Neural Networks" and "AI_L: Natural language"). As an alternative to restricting the sample in this way, we relax the "all firms are potential entrants" assumption by including presence in an adjacent market as an explanatory variable in the regression.

[39] There are 100,422 observations in which the firm has either previously entered the market and is thus dropped or has a missing value for age. Thus, the dataset on which the regressions in Table 6 are based contains 1,466,577 observations.

In addition to the explanatory variables included in the market-level regressions described above, we include the following determinants of the cost of entry: the log of the stock of relevant patents held by the firm[40] (or, alternatively, a dummy equal to 1 if the firm holds patents relevant to the market), the mean number of forward citations to the firm's patents[41], and the measures of bargaining costs. We control for firm size using a categorical measure of the level of the firm's revenues.[42] To control for the firm's prior experience in related markets, we include the number of adjacent markets in which the firm is active, and a count of the number of other markets (outside the broad SOF class) previously entered by the firm. Finally, we include controls for the age of the firm, the stage of the product cycle (through the modal citation lag in the SOF), and year and market fixed effects. We cluster the standard errors by firm to account for potential correlation across observations on the same firm. The parameters in Tables 6 and 7 are expressed as hazard ratios (or odds ratios in the case of the discrete-time hazard model).

The results in the first column of Table 6 show that (consistent with prior research) the probability of entering market j is increasing in the number of adjacent markets the firm has already entered. Similarly, the total number of markets in which firms are active is a significant positive predictor of entry. Although we do not report the coefficients due to space constraints, the effect of the firm's size (revenues) on the probability of entry is concave, with an inflexion point at the $100 million to $250 million range.

[40] When the log of the firm's patents is included on the right-hand-side, we also include a dummy equal to 1 if the firm has no patents.

[41] These firm-level patent and citation counts are substituted for the number of patents and average citations to patents held by entrants, which were independent variables in the market-level regressions.

[42] This is a set of dummies for each category of revenue: 0 = under $1m; 1 = $1m - $2.5m; 2 = $2.5m - $5m; 3 = $5m - $10m; 4 = $10m - $25m; 5 = $25m - $50m;6 = $50m - $100m; 7 = $100m - $250m; 8 = $250m - $500m; 9 = Over $500m

After controlling for the average quality of patents by the mean citation count per patent, the number of patents in a market is associated with a large and significant reduction in the probability of entry. A one-unit increase in Ln(Market's Patents) lowers the hazard ratio by almost one half. In elasticity terms, a 1% increase in the number of patents in a market is associated with approximately a 0.5% reduction in the hazard of entry.[43] Furthermore, if a firm holds patents in a market, the hazard of entry is approximately three times higher. The positive association between the firm's patents and entry persists after controlling for the average quality of the firm's patents (column 2).[44] It is also interesting to note that the coefficient on the dummy indicating that a firm has patents in a market is significantly larger than the one on the dummy indicating that the firm has other patents not related to that market (see column 5). This suggests that the former variable captures something about the benefits of having intellectual property related to a specific market, and not just the phenomenon that capable firms both have more patents and are more likely to enter new markets in general.

As in the results from the market-level model, our measure of the concentration of citations (the four-assignee concentration ratio) is not significantly associated with the probability of entry, after controlling for the number and quality of patents in a market. As in the market-level model, the number of cited assignees and the number of cited assignees squared (divided by 100) are significantly associated with the probability of entry (see

---

[43] This can be seen by recognizing that the Cox model for the hazard is $h(t) = h_0(t) \exp(x'\beta)$. If we take natural logs of both sides, and take the derivative with respect to a particular X variable X_k, we have $\delta \ln(h_i(t))/\delta X_k = \beta$. If X enters in logs, we can interpret the coefficient on the log of X as the percentage change in the hazard associated with a 1% change in X. Since the table reports hazard ratios, coefficients can be obtained by taking the log of the hazard ratios reported in the table.

[44] However, somewhat surprisingly, there is no significant difference in the effect of the number of patents in the market on firms that do or do not have patents, as evidenced by the coefficient on the interaction between the patent dummy and the market's patent count

column 6), again lending support to the view that entry is affected by bargaining costs that increase with the number of parties who could potentially sue an entrant for infringement.

When we break the results down according to whether or not the firm is active in an adjacent market (columns 5 and 6 of Table 6), the results are not dramatically different.[45] However, there are some small but interesting differences in the patent coefficients: potential entrants are more negatively affected by the number of patents in a market when they are not active in an adjacent market, and the firm's patent count and "no patent" dummy similarly have a larger effect on entry when the firm is not active in adjacent markets. This may reflect the fact that firms already active in adjacent markets have sources of bargaining power that help counteract any deterrent effect of patents.

Exit model

We also estimate the parameters of a model of exit, where exit is identified when a firm is no longer observed in a market after two sample periods of activity in that market, or when a firm leaves the sample altogether.[46] We use the same set of predictors that we use for the entry model, and again we include fixed effects for the market, year, level of revenues, and the modal lag in the market, and we cluster the standard errors by firm.[47] The results are reassuring in that the variables that are associated with increases in the probability of entry generally appear to be associated with lower exit probabilities. For example, just as increases in the scope of the firm's activity across markets are positively associated with entry, increases in scope are negatively associated with the probability of

[45] A potential entrant is active in an adjacent market for 12% of the firm-market-year observations.
[46] We only have data on firms that exit the sample after 1999, so we restrict the analysis to the 2000-2004 period.
[47] Because there are small numbers of firms in some markets, we aggregate the market-level fixed effects at a higher level, grouping all "adjacent" markets together (i.e. neural networks, natural language, and voice technology are grouped together as "artificial intelligence", etc.)

exit in a given year. Most importantly, having patents relevant to a market significantly increases the probability of survival in that market. This is seen in the coefficient on the firm's patent dummy, which shows that having patents relevant to a market reduces the odds of exiting that market by approximately 36%.[48] Having additional patents further reduces the probability of exit, as shown by the coefficient on the dummy equal to one if the firm has patents outside market *j* in year *t*.

Increases in the number of patents in a market are associated with a reduction in the probability of exit for the average firm, but this effect is only significant at the 10% level. The variables intended to capture patent-related bargaining costs faced by entrants—the number of cited assignees and the concentration ratio of citations over assignees—are not significantly associated with the probability of exit.

Conclusions

Controlling for the characteristics of the firm and market, we find that software firms are less likely to enter product markets in which there are more patents. All else equal, a 1% increase in the number of patents in a market is associated with approximately a 0.5% reduction in the hazard of entry for a typical firm. This result still holds after controlling for changes in the average number of citations received by incumbents' patents (as well as the size and the growth rate of the market), which suggests that patents have an entry-deterring effect above and beyond the degree to which they reflect the technological capabilities of the firms that generate them. However, it is difficult to draw unambiguous conclusions about the overall impact of increased patenting on entry from this result alone. For example, the negative impact of patents held by incumbents (and non-competitors) on

[48] The magnitude of this effect is consistent with Cockburn and Wagner's finding for survival of dot-com's.

entry is offset by the apparently large benefits to entrants from holding patents: we find that, all else equal, firms holding software patents associated with a market are three times more likely to enter that market, and 36% less likely to exit a market after entry.

The net impact of increased patenting on entry is therefore ambiguous, and we stress that the conclusions that can be drawn from the results presented here about the impact of increased numbers of software patents on entry into markets for software products are quite sensitive to assumptions about accompanying changes in the quality of patents and similarities or differences in the growth rates of patenting by incumbents and entrants. The table below summarizes the implications of our estimates from the different market-level models described above for the net effect on entry, using various sets of seemingly reasonable assumptions on these factors.

Predicted change in the number of entrants associated with a ten percent increase in the number of patents held by both incumbents and entrants

Fixed Effects Negative Binomial model	Mean Scaling GMM model	Quasi-Differenced GMM model
Patent holdings increase by 10%, no change in mean cites per patent		
-2.89%	-4.95%***	-2.45%**
(2.2%)	(1.1%)	(1.2%)
Patents holdings increase by 10%, new patents receive zero citations		
3.54%	1.38%	-0.31%
(2.4%)	(2.9%)	(2.5%)
Patent holdings increase by 10%, new patents receive half as many citations as existing patents		
0.20%	-1.94%	-1.42%
(1.2%)	(1.3%)	(1.1%)

As the table shows, our estimated coefficient in the market-level entry model imply a wide range of effects that an increase of 10% in the total number of software patents in the economy would, all else equal, have on the net number of entrants into the average software market over two years. Though the majority of these estimates of the net effect are negative, several are positive, and it is important to note that that the standard errors on these net effects are quite large, with only the largest of this range of estimates statistically discernable from zero.

We also find evidence for similarly complex relationships between rates of entry and measures of bargaining costs faced by an entrant. The number of assignees cited by patents related to a market presumably correlates with the number of entities that an entrant would have to obtain licenses from in order to minimize expected costs of entering a market. Yet it may also be inversely correlated with the blocking power of incumbents' patents or the degree to which they "cover" technology space. Consistent with these offsetting effects we find a strong positive effect of increases in the number of cited assignees on entry when there are relatively few of them in the market (and a strong negative effect in markets with relatively high numbers of cited assignees. In the first case, we interpret a higher number of cited assignees as indicative of weaker or more fragmented patent rights, which more than offsets any increase in bargaining costs from having to negotiate with larger numbers of actors. In the second case, the increase in bargaining costs dominates.

Perhaps the most important conclusion to be drawn from this study is that the relationship between patents and the dynamics of industry structure is multifaceted, with a variety of offsetting effects at work. Our results suggest that, at least in this industry, these offsetting effects appear to be quite closely balanced, making the impact of patent policy changes on entry, the pace of innovation, and ultimately social welfare quite difficult to predict. The need for further research appears, as ever, to be clearly indicated. A more structural approach, with an explicit specification of patenting decisions and equilibrium among potential and actual market participants, may be able give more easily interpretable estimates of responses to policy changes. It would also be interesting to extend this research to cover other industries and technologies.

References

Acs, Z. and D. Audretsch (1990) *Innovation and Small Firms.* Cambridge, MA: MIT Press.

Adams, J., R. Clemmons, and P. Stephan (2006) "How Rapidly Does Science Leak Out?" *NBER Working Paper No. 11997.*

Allison, P. (1984) *Event History Analysis: Regression for Longitudinal Event Data*, Newbury Park: Sage Publications, Inc.

Bresnahan, T. and P. Reiss (1991) "Entry and Competition in Concentrated Markets." *Journal of Political Economy*, 99, pp. 977-1009.

Berry, S. (1992) "Estimation of a Model of Entry in the Airline Industry." *Econometrica*, 60(4), pp. 889-917.

Bessen, J. and E. Maskin (2000) "Sequential Innovation, Patents, and Imitation." Working Paper. http://www.researchoninnovation.org/patent.pdf.

Bessen, J. and R. Hunt (2003) "An Empirical Look at Software Patents." Working Paper. http://www.researchoninnovation.org/swpat.pdf.

Campbell-Kelly, M. (2003) *From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry*. Cambridge, MA: MIT Press.

Chamberlain, G. (1992) "Comment: Sequential Moment Restrictions in Panel Data." *Journal of Business and Economic Statistics,* 10, pp. 20-26.

Cockburn, I. (2005) "State Street Meets the Human Genome Project." Chapter in R. Hahn (ed.) *Intellectual Property Rights in Frontier Industries: Bioinformatics and Software.* Washington, DC: Brookings-AEI Joint Center.

Cockburn and Wagner (2006) "Patents and the Survival of Internet-Related IPOs." Working paper.

Ericson, R. and A. Pakes (1995) "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *The Review of Economic Studies*, 62(1), pp. 53-82.

Geroski, P. (1995) "What do we know about entry?" *International Journal of Industrial Organization,* 13, pp. 421-440.

Geroski, P. (1989) "Entry and the rate of innovation." *Economics of Innovation and New Technology* 1, pp. 203-214.

Gilbert, R. and D Newbery (1992) "Preemptive Patenting and the Persistence of Monopoly." *American Economic Review,* 73(2), pp. 514-526.

Gort, M. and S. Klepper (1982) "Time Paths in the Diffusion of Product Innovations." *Economic Journal,* 92(367), pp. 630-653.

Graham, S. J. H. and D. C. Mowery (2003) "Intellectual Property Protection in the U. S. Software Industry." Chapter in W. M. Cohen and S. A. Merrill (eds.), *Patents in the Knowledge-Based Economy*, Washington, DC: National Academies Press.

Greenstein, S. and J. Wade (1998) "The Product Life Cycle in the Commercial Mainframe Computer Market, 1968-1982, *RAND Journal of Economics,* 29(4), pp. 772-789.

Hall, B. (2004) "Exploring the Patent Explosion." *NBER Working Paper No. 10605.*

Hall, B. and M. MacGarvie (2006) "The Private Value of Software Patents." *NBER Working Paper No. 12195.*

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

Heller and Eisenberg (1998) "Can Patents Deter Innovation? The Anti-commons in Biomedical Research" *Science*, 280(5364): 698–701.

Jovanovic, B. (1982) "Selection and the Evolution of Industry." *Econometrica,* 50(3), pp. 649-670.

Jovanovic, B. and S. Lach (1989) "Entry, Exit, and Diffusion with Learning by Doing." *American Economic Review*, 79(4), pp. 690-699.

Klepper, S. (1996) "Entry, Exit, Growth, and Innovation over the Product Life Cycle" *American Economic Review*, 86(3), pp. 562-583.

Kyle, M. (2006) "The Role of Firm Characteristics in Pharmaceutical Product Launches." Forthcoming in *RAND Journal of Economics.*

Lanjouw, J. and I. Cockburn (2001) "New Pills For Poor People? Empirical Evidence After GATT." *World Development,* 29(2), pp. 265-289.

Lanjouw, J. and M. Schankerman (2004) "Patent Quality and Research Productivity: Measuring Innovation With Multiple Indicators." *Economic Journal*, 114(495), 441-465.

Lerner, Josh. (2002) "150 Years Of Patent Protection." *American Economic Review*, 92(2), pp. 221-225.

Lerner, J. and F. Zhu (2005) "Software Patent Shifts after Lotus v. Borland." *NBER Working Paper No. 11168.*

Mann, R. J. and T. Sager (2005) "Patents, Venture Capital, and Software Start-Ups." University of Texas Law School, Law and Economics Research Paper No. 57, September 2005.

Mazzeo, M. (2002) "Product Choice and Oligopoly Market Structure." *RAND Journal of Economics*, 33(2) pp. 1-22.

Merges, R. P. (2006) "Patents, Entry and Growth in the Software Industry." Working paper, UC Berkeley School of Law.

Olley, S. and A. Pakes (1996) "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica,* 64(6), pp. 1263-1297.

Prusa, T. J. and J. A. Schmitz (1991) "Are New Firms an Important Source of Innovation? Evidence from the PC Software Industry." *Economics Letters*, 35(3), pp. 339-342.

Prusa, T. J. and J. A. Schmitz (1994) "Can Companies Maintain Their Initial Innovation Thrust? A Study of the PC Software Industry"; *Review of Economics and Statistics*, 76(3), pp. 523-540.

Scott Morton, F. (1999) "Entry Decisions in the Generic Drug Industry" *RAND Journal of Economics,* 30(3), pp. 421-440.

Seim, K. (2004) "An Empirical Model of Firm Entry with Endogenous Product-Type Choices." Mimeo, Stanford Graduate School of Business.

Shapiro, C. (2001) "Navigating the Patent Thicket: Cross-licenses, Patent Pools, and Standard-Setting." In *Innovation Policy and the Economy, Vol. 1.* Adam Jaffe, Joshua Lerner, and Scott Stern, (eds.), MIT Press for the National Bureau of Economic Research.

Smith, B. L. and S. O. Mann (2004) "Innovation and Intellectual Property Protection in the Software Industry: An Emerging Role for Patents." *University of Chicago Law Review,* 71, pp. 241-264.

Staiger, D., and J. H. Stock (1997), "Instrumental Variables Regression with Weak Instruments." *Econometrica,* 65, p. 557-586.

Sutton, J. (1991) *Sunk Costs and Market Structure: Price Competition, Advertising and the Evolution of Concentration*. Cambridge, MA: MIT Press.

Toivanen, O. and M. Waterson (2005) "Market Structure and Entry." *RAND Journal of Economics*, 36(3), pp. 680-699.

Wooldridge, J. (1997) "Distribution-Free Estimation of Some Nonlinear Panel Data Models." *Journal of Econometrics,* 90, pp. 77-91.

Table 1: Variable Definitions and Summary Statistics

1,566,999 Firm-Market-Year Observations

Variable name	Definition	Mean	Std. Dev.	Min.	Max.
Firm-level variables					
Age	t – year the firm was founded	16.056	12.619	0	229
Other markets	The number of other SOF classes in which firm i is active in year t	0.475	1.0120	0	23
D(firm has patents in market)	A dummy equal to 1 if firm i has any patents in market j in year t.	0.0069	0.0827	0	1
Firm's patents in market	The cumulative stock of firm i' patents relevant to market j in year t	0.0769	4.543	0	1619.46
Quality of firm's patents in market	The cumulative number of citations to firm i's patents in market j granted in year t, divided by *firm's patents in market*	0.0758	1.8776	0	382
Adjacent markets	The number of markets related to market j in which a firm is active in year t	0.182	0.643	0	22
Market-level variables					
Incumbents	The number of firms active in SOF market j in year t-1	129.698	155.361	0	728
Growth of Revenues	The change in the log of the sum of *Revenues* for the firms active in market j from year t-1 to year t	-6.694	2.540	-18.421	-2.452
Revenues	A categorical measure of the firm's total revenues (mean calculated by taking the mid-point of the range, or 500mil for the category >=500mil)	$46.9 million		<$1 million	>= $500 million
Market's patents	The cumulative stock of patents associated with market j in year t	1501.009	1599.792	4.03	5982
Quality of market's patents	The cumulative stock of citations to *Market's Patents,* divided by *Market's Patents*	13.860	12.650	1.579	94.896
Entrants' total patent stock	The sum of the patent stocks of firms entering market j in year t	1.568	9.358	0	113.531
Mean Quality of entrants' patents	Total citations per patent held by firms entering market j in year t	0.899	6.393	0	79.599
Number of cited assignees in market	Number of assignees that are cited by patents in this market	695.969	591.726	6	2307
Four-assignee citation concentration ratio	The share of all citations made in a market that are made to the top four most-cited assignees, by year	0.465	0.120	0.211	0.935

Table 2: Summary Statistics on Entry, by SOF and year

	Number of Incumbents						Number of Entrants					
SOF	**1994**	**1996**	**1998**	**2000**	**2002**	**2004**	**1994**	**1996**	**1998**	**2000**	**2002**	**2004**
Invoicing/Billing Software	372	400	452	458	526	551	23	33	28	18	10	35
Tax preparation and reporting software	99	99	111	107	124	122	5	7	10	4	0	11
Voice technology software	13	20	37	59	82	103	4	6	10	4	3	15
Natural language software	4	5	8	10	16	19	1	2	2	2	2	3
Neural network software	4	5	11	16	27	24	0	4	5	3	0	0
Automatic teller machine software	21	20	20	19	28	27	1	1	2	4	0	2
Fax software	0	24	68	91	114	116	12	21	18	8	8	8
Internet tools	0	0	41	234	504	728	0	23	90	78	35	71
Wide area network software	83	102	131	134	159	142	22	22	14	12	2	6
Local area network (LAN) software	8	16	45	58	88	84	7	13	11	13	5	9
File management software	140	167	230	262	340	505	35	36	26	37	31	92
Hierarchical DBMS software	14	20	32	34	42	47	2	8	5	4	0	5
Relational DBMS software	135	133	155	172	185	167	13	20	21	7	1	14
Database query language software	51	47	72	108	117	120	5	18	39	11	5	17
Robotic software	9	10	12	13	12	12	2	2	2	0	0	0
Quality control software	30	38	53	71	90	92	6	10	9	8	2	19
Three dimensional representation software	36	41	71	116	166	155	4	14	22	8	4	8
Electronic message systems software	46	50	70	92	152	206	7	12	16	16	10	29
Desktop publishing software	37	34	45	47	64	55	3	3	7	6	1	7
Artificial intelligence R&D	19	21	33	33	40	39	0	4	1	0	0	0
Geographic information systems software	38	48	81	94	136	131	6	15	12	7	1	14
Peripheral device drivers	34	41	50	58	98	108	5	5	9	16	1	8
Disaster recovery software	18	20	31	51	59	70	6	8	15	5	3	8
Security/auditing software	75	87	130	174	297	396	13	26	24	26	21	56
Performance measuring software	38	56	86	127	198	271	12	16	31	17	25	70
Inventory management software	416	442	549	557	651	661	30	53	38	26	16	48
Order entry/processing software	268	290	336	375	470	501	29	36	42	23	13	39
Mean	74.4	82.8	109.6	132.2	177.2	201.9	9.4	15.5	18.9	13.4	7.4	22.0
Median	36	41	68	92	117	120	6	13	14	8	3	11
Std. Dev.	108.9	115.5	134.1	137.5	173.9	208.6	9.9	12.8	18.5	15.7	9.9	24.9

Table 3: Means of market-specific patent variables, 1994-2004

SOF	Patents in market	Citations per patent in market	Number of cited assignees	Four-assignee CR	Modal Citation lag
Invoicing/Billing Software	47.91	38.742	74.67	0.53	5.33
Tax preparation and reporting software	9.31	6.497	12.67	0.70	4.67
Voice technology software	2080.88	8.785	674.50	0.43	6.50
Natural language software	745.22	9.232	361.00	0.50	7.17
Neural network software	442.68	7.771	245.00	0.47	7.17
Automatic teller machine software	269.03	14.668	356.00	0.45	5.33
Fax software	2761.70	15.822	1342.00	0.45	4.67
Internet tools	2761.70	15.822	1342.00	0.45	4.67
Wide area network software	2761.70	15.822	1342.00	0.45	4.67
Local area network software	2761.70	15.822	1342.00	0.45	4.67
File management software	2031.60	15.095	897.50	0.49	5.00
Hierarchical DBMS software	2155.44	14.430	1027.00	0.48	5.50
Relational DBMS software	2155.44	14.430	1027.00	0.48	5.50
Database query language software	2155.44	14.430	1027.00	0.48	5.50
Robotic software	216.21	8.801	182.17	0.45	5.75
Quality control software	55.58	9.643	76.67	0.41	5.88
3D representation software	1399.67	7.676	757.33	0.41	4.67
Electronic message systems software	130.07	26.880	190.67	0.53	5.86
Desktop publishing software	748.35	11.100	410.17	0.50	6.50
Artificial intelligence R&D	658.85	8.232	325.83	0.47	5.33
Geographic information systems software	3080.80	8.505	1085.67	0.43	4.50
Peripheral device drivers	3242.56	10.719	1077.83	0.45	4.33
Disaster recovery software	1767.84	12.099	713.33	0.44	5.67
Security/auditing software	657.29	21.701	479.67	0.47	7.00
Performance measuring software	3984.35	9.387	1023.67	0.42	6.33
Inventory management software	339.70	13.275	452.17	0.45	7.50
Order entry/processing software	1106.24	18.839	945.67	0.40	7.17
Total	1501.01	13.86	695.97	0.46	5.67

Table 4: Market-level Entry Model

Conditional Fixed-Effects Negative Binomial Regression with market fixed effects

Dependent variable = Number of entrants in market *j* in year *t*

	(1)	(2)	(3)	(4)	(5)
Incumbents	0.004***	0.004**	0.003*	0.002	0.003*
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Incumbents squared	-0.000**	-0.000**	-0.000*	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Growth of revenues	0.155***	0.177***	0.181***	0.218***	0.180***
	(0.050)	(0.053)	(0.053)	(0.055)	(0.053)
Ln(Market's patents)	-0.081	-0.469**	-0.388*	-0.248	-0.387*
	(0.097)	(0.215)	(0.226)	(0.255)	(0.229)
Ln(Quality of market's patents)		-0.858**	-0.789**	-0.564	-0.789**
		(0.387)	(0.400)	(0.432)	(0.401)
Ln(Entrants' total patent stock)			0.081	0.105*	0.080
			(0.060)	(0.058)	(0.060)
Ln(Quality of patents held by entrants)			0.116*	0.080	0.116*
			(0.063)	(0.063)	(0.063)
Number of cited assignees /100				0.076*	
				(0.045)	
Number of cited assignees/100 squared				-0.004**	
				(0.002)	
Four-assignee CR					0.018
					(0.577)
Modal lag decile=2	-0.177	-0.184	-0.106	-0.093	-0.106
	(0.133)	(0.128)	(0.131)	(0.128)	(0.132)
Modal lag decile=4	-0.035	-0.064	-0.024	0.067	-0.024
	(0.143)	(0.139)	(0.139)	(0.141)	(0.139)
Modal lag decile=6	-0.243	-0.166	-0.210	-0.056	-0.209
	(0.187)	(0.180)	(0.183)	(0.185)	(0.184)
Modal lag decile=7	-0.097	-0.135	-0.074	-0.058	-0.075
	(0.188)	(0.183)	(0.183)	(0.181)	(0.185)
Modal lag decile=9	-0.172	-0.236	-0.198	-0.220	-0.199
	(0.238)	(0.230)	(0.227)	(0.223)	(0.230)
Modal lag decile=10	0.272	0.279	0.230	0.073	0.230
	(0.203)	(0.195)	(0.195)	(0.208)	(0.197)
Constant	3.584***	9.000***	8.398***	7.255**	8.384***
	(0.737)	(2.669)	(2.769)	(2.963)	(2.808)
Observations	162	162	162	162	162

162 observations

Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: IV Estimates of Market –level Entry Model

GMM Estimates of Count Data Regression

Variables treated as endogenous:	Mean scaling model: None	Mean scaling model: Market's patents	Mean scaling model: Market's patents & quality	Quasi-differenced model: None	Quasi-differenced model: Market's patents	Quasi-differenced model: Market's patents & quality	Fixed-effects Negative Binomial Model
Modal lag	-0.228***	-0.245***	-0.241***	0.030	0.019	0.047	-0.119
	(0.064)	(0.065)	(0.069)	(0.055)	(0.058)	(0.060)	(0.077)
Modal lag squared	0.014***	0.015***	0.015***	-0.004	-0.003	-0.004	0.009**
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.005)
Incumbents	-0.002	-0.001	0.000	-0.004***	-0.004***	-0.003***	0.003
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Incumbents squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Growth of revenues	0.394***	0.340***	0.337***	0.254***	0.250***	0.258***	0.173***
	(0.086)	(0.074)	(0.078)	(0.044)	(0.044)	(0.045)	(0.052)
Ln(Market's patents)	-0.513***	-0.501***	-0.569***	-0.274***	-0.296***	-0.298**	-0.338*
	(0.105)	(0.111)	(0.117)	(0.110)	(0.113)	(0.124)	(0.202)
Ln(Quality of market's patents)	-0.999***	-1.027***	-0.764**	-0.204	-0.256	-0.292	-0.747**
	(0.281)	(0.290)	(0.348)	(0.287)	(0.298)	(0.332)	(0.363)
Ln(Entrants' total patent stock)	0.054	0.040	0.036	0.052	0.058	0.038	0.052
	(0.039)	(0.041)	(0.043)	(0.040)	(0.041)	(0.043)	(0.061)
Ln(Quality of patents held by entrants)	0.091**	0.097**	0.088**	0.089*	0.085*	0.065	0.140**
	(0.040)	(0.039)	(0.041)	(0.048)	(0.048)	(0.051)	(0.061)
Over-identification test statistic	58.524	53.9909	44.9491	108.55	106.222	93.491	
p-value	[.248]	[.195]	[.272]	[.101]	[.032]	[.038]	

162 observations.

Asymptotic standard errors in parentheses. Standard errors in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

See text for estimation method. All models also include year fixed effects.

Table 6: Firm-level entry model

Year and market fixed effects included
Coefficients expressed in exponentiated form.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Discrete-time hazard	Cox Proportional Hazard						
	Full Sample						Not active in adjacent mkt	Active in adjacent mkt
Incumbents	1.003***	1.002**	1.002*	1.002*	1.002*	1.001	1.004***	1.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Incumbents squared	1.000***	1.000**	1.000**	1.000**	1.000**	1.000	1.000***	1.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Age	0.981***	0.963***	0.963***	0.964***	0.964***	0.963***	0.961***	0.966***
	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Other markets	1.296***	1.251***	1.251***	1.250***	1.251***	1.250***	1.403***	1.036
	(0.020)	(0.019)	(0.019)	(0.020)	(0.020)	(0.019)	(0.027)	(0.024)
Growth of revenues	1.173***	1.185***	1.218***	1.219***	1.219***	1.249***	1.239***	1.189***
	(0.026)	(0.026)	(0.028)	(0.028)	(0.028)	(0.031)	(0.039)	(0.043)
Adjacent markets	1.636***	1.252***	1.253***	1.225***	1.229***	1.253***		1.201***
	(0.038)	(0.048)	(0.048)	(0.050)	(0.049)	(0.047)		(0.033)
Ln(Market's patents)	0.918	0.948	0.510***	0.510***	0.511***	0.540**	0.373***	0.428***
	(0.082)	(0.084)	(0.107)	(0.106)	(0.107)	(0.134)	(0.111)	(0.134)
Ln(Firm's patents in market)	1.327**	1.388***	1.392***			1.388***	1.381	1.225**
	(0.162)	(0.148)	(0.164)			(0.165)	(0.324)	(0.110)
No patent dummy	0.456***	0.455***	0.452***			0.438***	0.405**	0.515***
	(0.105)	(0.108)	(0.101)			(0.099)	(0.169)	(0.108)
Ln(Quality of market's patents)			0.300***	0.305***	0.303***	0.325***	0.208***	0.218***
			(0.106)	(0.108)	(0.107)	(0.134)	(0.106)	(0.113)
Ln(Quality of firm's patents in market)			0.994		0.856	0.986	1.100	1.096
			(0.110)		(0.102)	(0.110)	(0.148)	(0.115)
D(firm has patents in market)				2.842***	3.397***	1.129***		
				(0.440)	(0.609)	(0.038)		
D(firm has patents outside market)				1.448***	1.437***	0.994***		
				(0.149)	(0.151)	(0.001)		
Number of cited assignees/100						1.120***		
						(0.042)		
Number of cited assignees/100 squared						0.994***		
						(0.001)		
Observations	1466577	1466577	1466577	1466577	1466577	1466577	1300582	165995

Robust standard errors, clustered by firm, in parentheses. * significant at 10%; ** significant at 5%; ***significant at 1%
All regressions also include year and market fixed effects, and a set of categorical variables measuring firm revenue

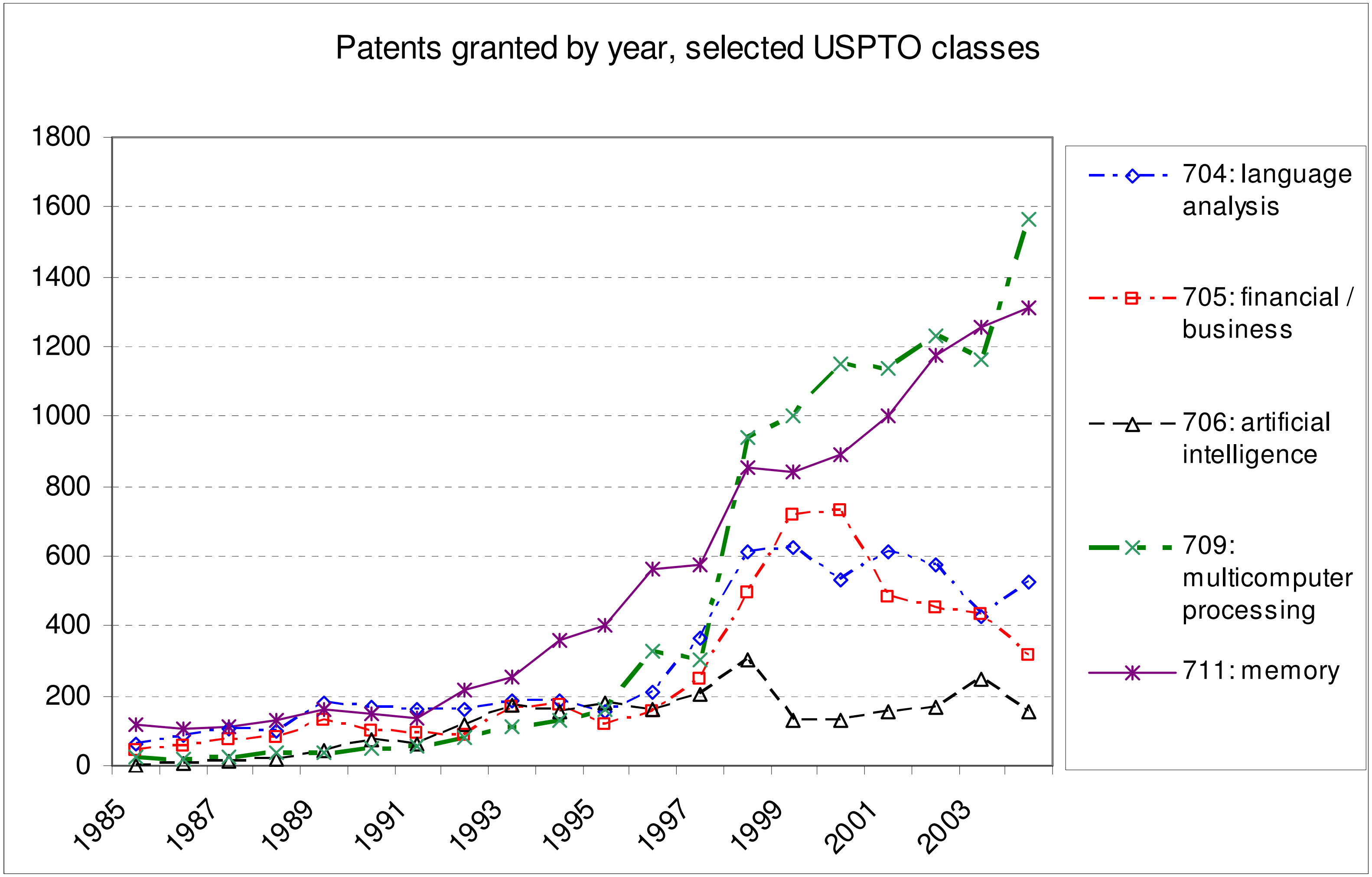
Table 7: Firm-level exit model

Coefficients expressed in exponentiated form.

	(1)	(2)	(3)	(4)	(5)	(6)
	Discrete-time Hazard	Cox Proportional Hazard				
Incumbents	1.004 (0.003)	1.003 (0.002)	1.002 (0.003)	1.002 (0.003)	1.002 (0.003)	1.003 (0.003)
Incumbents squared	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
Firm's age	0.985** (0.006)	0.988** (0.005)	0.988** (0.005)	0.988** (0.005)	0.988** (0.005)	0.988** (0.005)
Growth of revenues	0.887 (0.085)	0.928 (0.074)	0.947 (0.078)	0.950 (0.079)	0.937 (0.076)	0.921 (0.078)
Other markets	0.840*** (0.031)	0.840*** (0.024)	0.841*** (0.024)	0.841*** (0.024)	0.841*** (0.024)	0.840*** (0.024)
Adjacent markets	0.937*** (0.022)	0.952** (0.019)	0.952** (0.019)	0.950** (0.019)	0.950** (0.019)	0.951** (0.019)
Ln(Market's patents)	0.926 (0.051)	0.931 (0.043)	0.930 (0.045)	0.925 (0.044)	0.912* (0.048)	0.848 (0.094)
Ln(Firm's patents)	0.735 (0.139)	0.848 (0.129)	0.957 (0.195)			
D(firm has no patents in market)	0.912 (0.461)	0.949 (0.402)	0.897 (0.400)			
D(firm has patents in market)				0.638*** (0.106)	0.639*** (0.106)	0.641*** (0.107)
D(firm has patents outside market)				0.502*** (0.088)	0.502*** (0.088)	0.503*** (0.088)
Number of cited assignees/100						1.051 (0.095)
$(\text{Number of cited assignees}/100)^2$						0.999 (0.003)
4-assignee CR					0.681 (0.469)	
Ln(Quality of market's patents)			1.167 (0.155)	1.173 (0.156)		
Ln(Quality of firm's patents in market)			0.897 (0.074)			
Observations	14241	12257	12257	12257	12257	12257

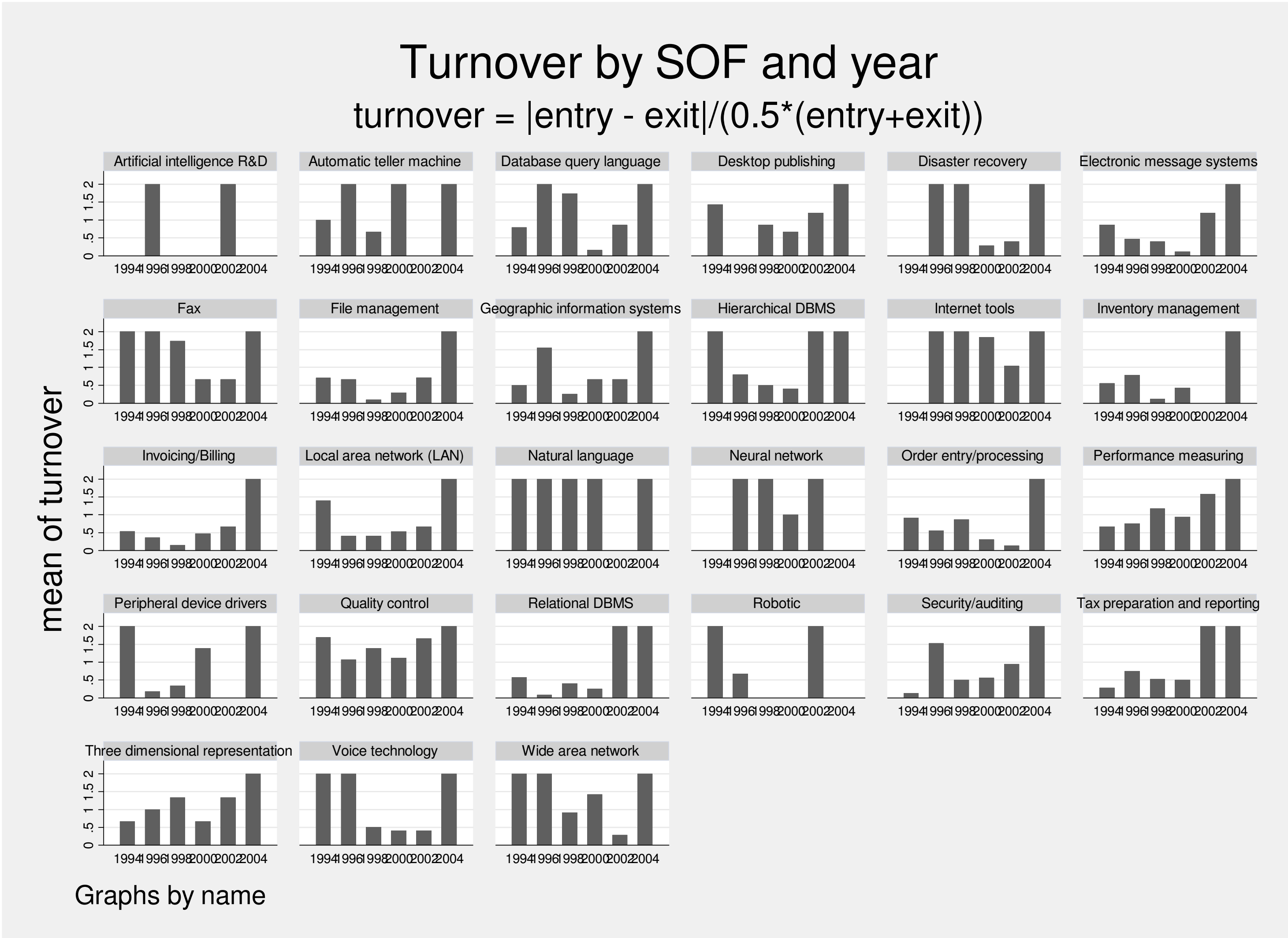
Year and market fixed effects included.
Robust standard errors, clustered by firm, in parentheses.
* significant at 10%; ** significant at 5%; *** significant at 1%

Figure 1



Patents granted by year, selected USPTO classes
1800
1600
1400
1200
1000
800
600
400
200
0
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
704: language analysis
705: financial / business
706: artificial intelligence
709: multicomputer processing
711: memory

Figure 2



Note: exit statistics reported here include only firms that exit a particular market but remain in the sample. They exclude firms that exit from the sample altogether, due to inconsistent reporting of these exits over time

Figure 3

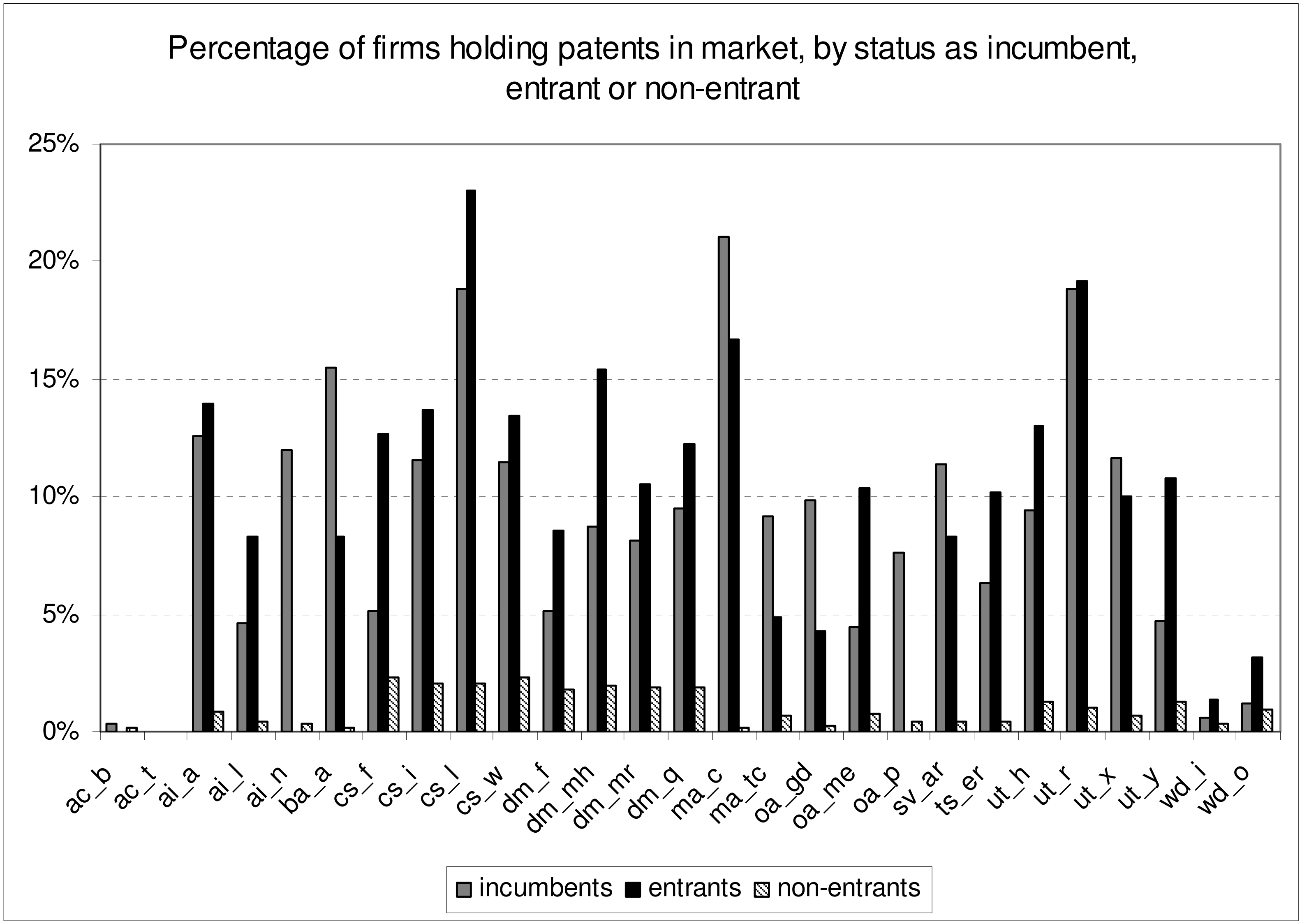
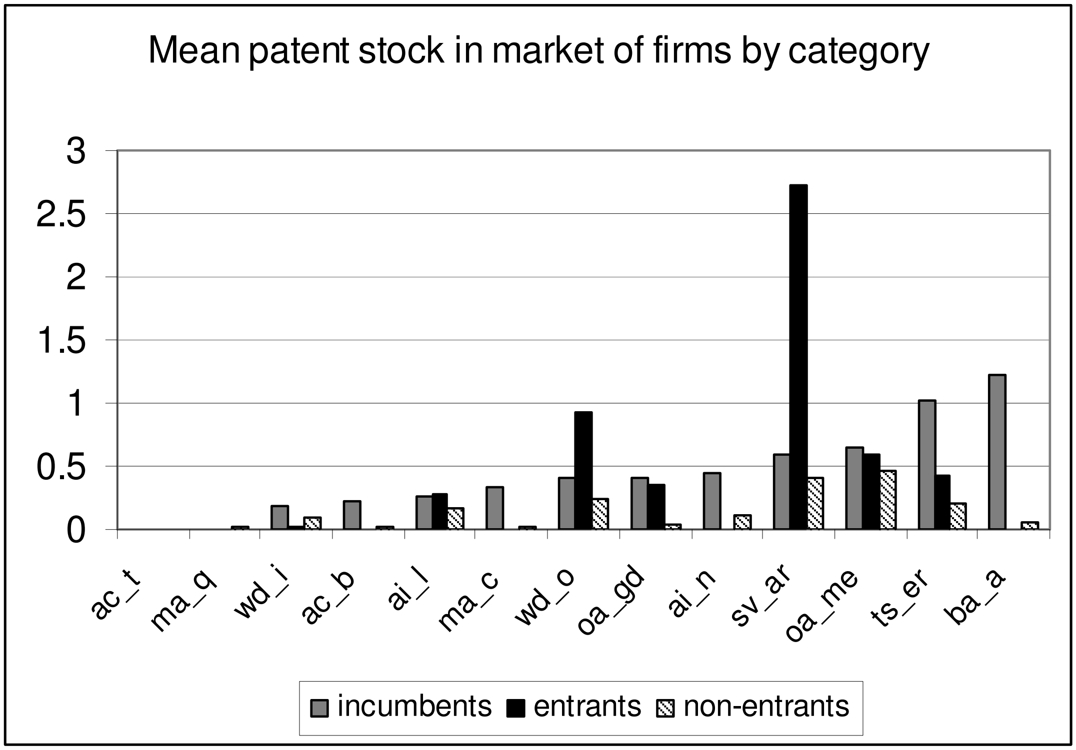
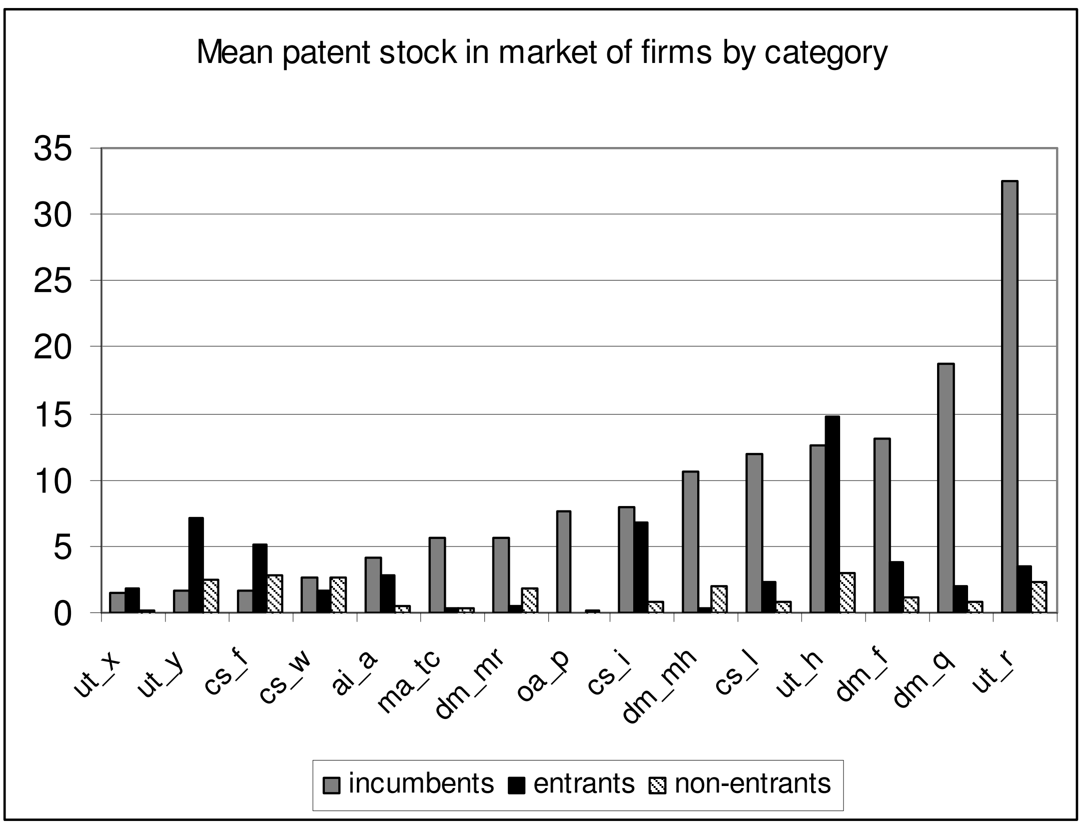

Percentage of firms holding patents in market, by status as incumbent, entrant or non-entrant
25%
20%
15%
10%
5%
0%
ac_b
ac_t
ai_a
ai_l
ai_n
ba_a
cs_f
cs_i
cs_l
cs_w
dm_f
dm_mh
dm_mr
dm_q
ma_c
ma_tc
oa_gd
oa_me
oa_p
sv_ar
ts_er
ut_h
ut_r
ut_x
ut_y
wd_i
wd_o
incumbents
entrants
non-entrants

Figure 4



Mean patent stock in market of firms by category
3
2.5
2
1.5
1
0.5
0
ac_t
ma_q
wd_i
ac_b
ai_l
ma_c
wd_o
oa_gd
ai_n
sv_ar
oa_me
ts_er
ba_a
incumbents
entrants
non-entrants



Mean patent stock in market of firms by category
35
30
25
20
15
10
5
0
ut_x
ut_y
cs_f
cs_w
ai_a
ma_tc
dm_mr
oa_p
cs_i
dm_mh
cs_l
ut_h
dm_f
dm_q
ut_r
incumbents
entrants
non-entrants

DATA APPENDIX

SOF-Patent concordance

This section describes the process used to develop a mapping between SOF categories and patents. Our initial approach was to look at specialists -- firms that produced in only one of the aggregate categories (i.e.: AI: "Artificial Intelligence software", DM "Database/file management software", etc.). We created a concordance based on the three most common USPTO primary classes associated with specialists in these fields. However, this approach proved unsatisfactory for several reasons.

First, the concordance is based on the patents of small, young firms with few patents. This creates potential for bias because the firms most actively engaged in patenting are the ones that have products in several areas. By focusing on specialists, we may miss an important part of patenting in the sector. Second, firms could be deterred from entering a market by the existence of patents held by firms that are not competitors in product markets but that hold key upstream patents and insist on costly licenses. So focusing only on patents held by the firm's direct competitors may also ignore important areas of the relevant intellectual property landscape.

Finally, some of the aggregate classes contain sub-classes that are quite heterogeneous. For example, "MA - manufacturing software systems" contains sub-classes MA_C "robotic software", MA_E "machine vision software", MA_Q "quality control software", and MA_F "factory data collection software", all of which are fairly distinct from each other. Focusing on the sub-classes makes it much easier to pick out a handful of class-subclass combinations that seem to map directly to the SOF category in question. For example, subclasses 245-264 (Robot control) of class 700 (DATA PROCESSING: GENERIC CONTROL SYSTEMS OR SPECIFIC APPLICATIONS) seem to map directly into SOF category MA_C. Similarly, subclasses 108-115 (performance monitoring for product assembly or manufacturing) of class 700 seem closely related to category MA_Q. Indeed, subclass 109 is called "quality control."

We identify the class-subclass combinations in the US Patent Classification that map into SOF sub-categories in the following way. First, we search the abstracts of our set of software patents for the key words used to describe the sub-category in the CorpTech codebook. We began by searching for the description of each SOF category in the patent abstracts. Since some of the key words are more specific than others, this method will obviously work well for some sub-categories (i.e.: "voice recognition software") and less well for others (i.e.: "operating systems").

Using these patents as a base, we then searched for words that co-occur with the key words. We calculate the frequency with which these words are observed in the patents containing key words, and divide it by the frequency with which the words are observed in all software patents, to obtain how many more times the word is observed in key word-matching patents than in random patents. We then examined the words in the top decile of this distribution, and selected the ones that were the best candidates for identifying relevant patents[49]. We then repeat the key word search including these words.

Once we have a set of patents that contain key words or words extremely likely to co-occur with key words, we looked at the citations made by these patents. We selected the most often-cited classes and subclasses, and then examined the PTO's description of

[49] This step is necessary to weed out idiosyncratic and misspelled words.

these classes. After a careful reading of the classification manual, we selected the classes that are both highly prevalent in the word-matching patents and clearly related to the sub-category in question. It is important to note that, because software is an area in which many of the patents have been re-classified following their grant dates, we also had to look up the current classifications of these patents. To do this, we used a script that downloads patents and their current classification from the USPTO website.

Table A1 lists the SOF-patent class concordance we obtained using this methodology. The concordance is currently restricted to 27 SOF categories.

Table A1: Mapping between CorpTech SOF codes and USPTO patent classes

CorpTech SOF code	CorpTech definition	Most commonly cited USPTO class	Subclasses and other class/subclass combinations used in mapping
ac_b	Invoicing/Billing Software	705 (DATA PROCESSING: FINANCIAL, BUSINESS PRACTICE, MANAGEMENT, OR COST/PRICE DETERMINATION)	001-045 (AUTOMATED ELECTRICAL FINANCIAL OR BUSINESS PRACTICE OR MANAGEMENT ARRANGEMENT) and 4XX (FOR COST/PRICE); 235 (REGISTERS) sub 375-385 (SYSTEMS CONTROLLED BY DATA BEARING RECORDS)
ac_t	Tax preparation and reporting software	705 (DATA PROCESSING: FINANCIAL, BUSINESS PRACTICE, MANAGEMENT, OR COST/PRICE DETERMINATION)	esp 019 (Tax processing) and 031 (Tax preparation or submission)
ai_a	Voice technology software	704 (DATA PROCESSING: SPEECH SIGNAL PROCESSING, LINGUISTICS, LANGUAGE TRANSLATION, AND AUDIO COMPRESSION/DECOMPRESSION)	All subclasses up to 278 are represented, esp 275, 9, 251, or 243
ai_l	Natural language software	704 (DATA PROCESSING: SPEECH SIGNAL PROCESSING, LINGUISTICS, LANGUAGE TRANSLATION, AND AUDIO COMPRESSION/DECOMPRESSION)	subclasses 8 and 9 esp (Multilingual or national language support; Natural language)
ai_n	Neural network software	706 (DATA PROCESSING: ARTIFICIAL INTELLIGENCE)	15-44 (Neural Networks)

ba_a	Automatic teller machine software	235 (REGISTERS)	379 and 380 (Banking systems and Credit or identification card systems); 705/41-43 (AUTOMATED ELECTRICAL FINANCIAL OR BUSINESS PRACTICE OR MANAGEMENT ARRANGEMENT); 700/231-238 (article handling/dispensing or vending)
cs_f	Fax software	709 (ELECTRICAL COMPUTERS AND DIGITAL PROCESSING SYSTEMS: MULTICOMPUTER DATA TRANSFERRING)	217-219 (REMOTE DATA ACCESSING) and 201-206 (DISTRIBUTED DATA PROCESSING and COMPUTER CONFERENCING)
cs_i	Internet tools	709 (ELECTRICAL COMPUTERS AND DIGITAL PROCESSING SYSTEMS: MULTICOMPUTER DATA TRANSFERRING)	all 2XX subclasses (deals with computers talking to each other) also 705, esp subclasses 026 (Electronic shopping (e.g., remote ordering) and 705/014(Distribution or redemption of coupon, or incentive or promotion program); and 707/10 (Database or file accessing, distributed or remote access)
cs_l	Local area network (LAN) software	709 (ELECTRICAL COMPUTERS AND DIGITAL PROCESSING SYSTEMS: MULTICOMPUTER DATA TRANSFERRING)	all 2XX subclasses
cs_w	Wide area network (WAN) software	709 (ELECTRICAL COMPUTERS AND DIGITAL PROCESSING SYSTEMS: MULTICOMPUTER DATA TRANSFERRING)	all 2XX subclasses
dm_f	File management software	707(DATA PROCESSING: DATABASE AND FILE MANAGEMENT OR DATA STRUCTURES)	sub 1-10 (DATABASE OR FILE ACCESSING) and 200-206 (FILE OR DATABASE MAINTENANCE)
dm_mh	Hierarchical DBMS software	707(DATA PROCESSING: DATABASE AND FILE MANAGEMENT OR DATA STRUCTURES)	sub 1-10 (DATABASE OR FILE ACCESSING) and 100-104.1 (DATABASE SCHEMA OR DATA STRUCTURE)
dm_mr	Relational DBMS software	707(DATA PROCESSING: DATABASE AND FILE MANAGEMENT OR DATA STRUCTURES)	sub 1-10 (DATABASE OR FILE ACCESSING) and 100-104.1 (DATABASE SCHEMA OR DATA STRUCTURE)
dm_q	Database query language software	707(DATA PROCESSING: DATABASE AND FILE MANAGEMENT OR DATA STRUCTURES)	sub 1-10 (DATABASE OR FILE ACCESSING) esp 002-006 (Query processing (i.e., searching)) and 100-104.1 (DATABASE SCHEMA OR DATA STRUCTURE)
ma_c	Robotic software	700 (DATA PROCESSING: GENERIC CONTROL SYSTEMS OR SPECIFIC APPLICATIONS)	sub 245-264 (**Robot** control)

ma_q	Quality control software	700 (DATA PROCESSING: GENERIC CONTROL SYSTEMS OR SPECIFIC APPLICATIONS)	108-115
oa_gd	Three dimensional representation software	class 345 (COMPUTER GRAPHICS PROCESSING AND SELECTIVE VISUAL DISPLAY SYSTEMS)	418-427 (Three-dimension) and 700/98 (3-D product design (e.g., solid modeling)); 118(three dimensional product forming) 119 (rapid prototyping); 120 (lithography)
oa_me	Electronic message systems software	709 (ELECTRICAL COMPUTERS AND DIGITAL PROCESSING SYSTEMS: MULTICOMPUTER DATA TRANSFERRING)	sub 206 (computer conferencing/Demand based messaging); 705/008 and 009 (Allocating resources or scheduling for an administrative function and Staff scheduling or task assignment)
oa_w	Word processor/text editor software	715 (DATA PROCESSING: PRESENTATION PROCESSING OF DOCUMENT, OPERATOR INTERFACE PROCESSING, AND SCREEN SAVER DISPLAY PROCESSING)	subclasses 5XX (PRESENTATION PROCESSING OF DOCUMENT)
oa_p	Desktop publishing software	715 (DATA PROCESSING: PRESENTATION PROCESSING OF DOCUMENT, OPERATOR INTERFACE PROCESSING, AND SCREEN SAVER DISPLAY PROCESSING)	500-542 (PRESENTATION PROCESSING OF DOCUMENT)
sv_ar	Artificial intelligence R&D	706 (DATA PROCESSING: ARTIFICIAL INTELLIGENCE)	15-62 (all subclasses)
ts_er	Geographic information systems software	701 (DATA PROCESSING: VEHICLES, NAVIGATION, AND RELATIVE LOCATION)	2xxx (NAVIGATION); 702/005 (Topography (e.g., land mapping))
ut_h	Peripheral device drivers	710 (ELECTRICAL COMPUTERS AND DIGITAL DATA PROCESSING SYSTEMS: INPUT/OUTPUT)	classes 1-74 (INPUT/OUTPUT DATA PROCESSING) esp sub 008-019 (Peripheral configuration/peripheral monitoring)
ut_r	Disaster recovery software	714 (ERROR DETECTION/CORRECTION AND FAULT DETECTION/RECOVERY)	sub 1-57 (DATA PROCESSING SYSTEM ERROR OR FAULT HANDLING) esp 006 (Redundant stored data accessed (e.g., duplicated data, error correction coded data, or other parity-type data)), also class 707(DATA PROCESSING: DATABASE AND FILE MANAGEMENT OR DATA STRUCTURES) 200-206(FILE OR DATABASE MAINTENANCE) esp sub 202 (Recoverability)
ut_x	Security/auditing software	726 (Information Security) all subclasses	also 705/50-79 (DATA PROCESSING: FINANCIAL, BUSINESS PRACTICE, MANAGEMENT, OR COST/PRICE DETERMINATION/BUSINESS PROCESSING USING CRYPTOGRAPHY)

ut_y	Performance measuring software	714 (ERROR DETECTION/CORRECTION AND FAULT DETECTION/RECOVERY)	all subclasses
wd_i	Inventory management software	705 (DATA PROCESSING: FINANCIAL, BUSINESS PRACTICE, MANAGEMENT, OR COST/PRICE DETERMINATION)	esp sub 28 (**Inventory management**) and 10 (Market analysis, demand forecasting or surveying)
wd_o	Order entry/processing software	705 (DATA PROCESSING: FINANCIAL, BUSINESS PRACTICE, MANAGEMENT, OR COST/PRICE DETERMINATION)	esp sub 1-45(AUTOMATED ELECTRICAL FINANCIAL OR BUSINESS PRACTICE OR MANAGEMENT ARRANGEMENT) including 26 (Electronic shopping (e.g., remote ordering))

How accurate and comprehensive is this concordance? Obviously, we need to balance type I errors associated with a too-narrow definition of the relevant set of patents against type II errors from a too-inclusive definition. We attempt to answer this question by determining what share of patents held by firms specializing in a category are picked up by the patent classes assigned to that category, and how many of the patents in those classes are assigned to specialist firms in the CorpTech database that *do not* operate in the category in question. A preliminary analysis of a selection of SOF codes well populated by specialist patents is found in Table A2.[50]

[50] We exclude patents held by firms that specialize in one SOF code, but that have a primary two-digit SIC code other than 73. We do this because these firms are not true specialists – they just appear as specialists in the Corptech dataset, which is restricted to software. These firms are likely to have patents in fields other than the software market in which they are active, and thus their patent portfolios are not a good indicator of state of the art in that particular software market.

Table A2: Validation of SOF-patent concordance for a selected set of SOF codes, using specialist patents

	Total specialist patents	True positives	Sensitivity	Positive predictive value
Billing/Invoicing software	9	4	0.444	0.571
Neural Network software	5	4	0.800	0.364
ATM software	14	10	0.786	1.000
Internet tools	62	17	0.274	0.218
WAN software	173	5	0.029	0.076
File maintenance software	26	4	0.154	0.053
Relational DMBS software	223	103	0.462	0.715
Quality Control software	7	0	0.000	0.000
Three-dimensional imaging software	45	5	0.111	0.417
Electronic message systems software	11	1	0.091	0.500
Geographical Information Systems software	25	9	0.360	0.474
Peripheral device drivers	117	41	0.350	0.410
Disaster recovery software	11	7	0.636	0.079
Security/auditing software	108	22	0.204	0.815
Performance measuring software	8	2	0.250	0.044

Sensitivity = share of specialist pats identified.
Positive predictive value = share of patents identified by mapping as belonging to that SOF that actually belong to a specialist in that SOF.

Because surely not all patents held by specialists are for technologies related to the firm's main product, we have also read the patents held by specialists to estimate how many such patents we should expect our concordance to (correctly) miss. We read the abstracts of all the specialist patents in a handful of categories, chosen because they are both narrowly-defined and populated by a significant number of specialist patents. These categories are invoicing/billing, automatic teller machines, geographic information systems, three-dimensional representation, and security/auditing. We found that a significant fraction of patents held by firms specializing in these fields were not strictly speaking covering technologies in the field. Table A3 lists the share of patents held by specialists in a SOF category that actually relate to technologies in that category.

As an example, consider the patents held by firms specializing in automatic teller machine software. A number of these patents are for software used to track and dispense medical items (5,912,818, 5,971,593, and 5,993,046). Others are for digital cash systems like smart cards (6,032,135). Others are simply not software patents (6,042,003: "lighting system for automated banking machine"), despite the fact that they are classified in IPC G06F.

Table A3: Specialist Patents

SOF	Number of specialist patents read	Share of specialist patents in SOF
invoicing/billing	10	60%
automatic teller machines	13	46%
geographic information systems	45	85%
three-dimensional representation	20	20%
security/auditing	26	65%

As a result, we should not necessarily expect our SOF-patent mapping to pick up all specialist patents, and the sensitivity of the mapping should be evaluated with this fact in mind. These findings, based admittedly on a small sample of SOFs, might suggest a rule of thumb like the following: if at least 50% of the patents held by specialists in a given area are picked up, the mapping can be considered successful.

Selectivity

Nothing about the 27 SOF classes for which we have established a patent concordance strikes us as being a source of serious selection bias. As noted in the main body of the paper, firms active in these markets tend to have more patents than firms in the markets we omit, but we feel this is an inevitable fact arising from the way the SOF categories are defined. Firms in sampled markets have on average sales of $50 million[51] and an age of 14.89 years. Firms in other markets have on average sales of $44 million, and an age of 14.53 years. The average entry rate of markets in the sample is 0.21, and the average exit rate is 0.12. Markets excluded from the sample have an average entry rate of 0.16 and an exit rate of 0.14 (the difference in exit rates is statistically insignificant), The high entry rate of the sample comes from the fact that it includes internet-related markets. When these are excluded, the average entry rate is 0.17, which is insignificantly different from the rate in the excluded markets.

[51] This calculation is based on a weighted average of the categorical revenue measures at the mid-point of the range. Because 23% of the observations in our CorpTech dataset have missing revenue data, this number may be inflated if the missing values tend to be firms with lower revenues.

Table A5: Truncation correction for citations in computer-related patents, based on Hall, Jaffe, and Trajtenberg (2005)

Citation weight	Application year
5.28194	1999
3.97728	1998
3.19610	1997
2.68650	1996
2.33322	1995
2.07697	1994
1.88451	1993
1.73594	1992
1.61867	1991
1.52441	1990
1.44748	1989
1.38391	1988
1.33079	1987
1.28600	1986
1.24793	1985
1.21536	1984
1.18731	1983
1.16304	1982
1.14194	1981
1.12352	1980
1.10738	1979
1.09320	1978
1.08070	1977
1.06966	1976
1.05989	1975
1.05122	1974
1.04351	1973
1.03666	1972
1.03055	1971
1.02509	1970