An Analysis of Small Business Patents by Industry and Firm Size

by

Anthony Breitzman, Ph.D.
and
Diana Hicks, Ph.D.

Haddonfield, NJ 08033

for

SBA
Office of Advocacy
www.sba.gov/advo

Under contract no. SBAHQ-07-Q-0010

Release Date: November 2008

This report was developed under a contract with the Small Business Administration, Office of Advocacy, and contains information and analysis that was reviewed and edited by officials of the Office of Advocacy. However, the final conclusions of the report do not necessarily reflect the views of the Office of Advocacy.

SBA Office of Advocacy
www.sba.gov/advo

Small Business Research Summary

Advocacy: the voice of small business in government

November 2008 No. 335

An Analysis of Small Business Patents by Industry and Firm Size

Anthony Breitzman and Diana Hicks., Haddonfield, NJ 08033 2008 [60] pages.
Under contract no. SBAHQ-07-Q-0010

Background

This study is the third in a series that examines small business patent activity. The authors created a database of 1,293 technology firms with 15 or more patents issued between 2002 and 2006. These firms are designated as innovative firms because of their high level of patent activity. Using this database, the authors analyze the relative strengths of small and large technology businesses, including information such as the industry and technology within which the firm patents and the importance of the patent. The results demonstrate that small businesses that innovate are indeed special and that the technology they create helps define the cutting edge in a number of industries. The report presents a convincing case that small firms in emerging industries are one of the greatest engines of American economic growth.

Overall Findings

Small firms are a significant source of innovation and patent activity. Small businesses develop more patents per employee than larger businesses, with the smallest firms, those with fewer than 25 employees, producing the greatest number of patents per employee. Furthermore, small firm patents tend to be more significant than large firm patents, outperforming them in a number of categories including growth, citation impact, and originality. Finally, small firms tend to specialize in high tech, high growth industries, such as biotechnology, pharmaceuticals, information technology, and semiconductors.

Highlights

- Of the 1,293 firms reviewed in this study, 504 had 500 or fewer employees, 760 had more than 500 employees, and no size information could be obtained for 29 firms. That is, 40 percent of the firms with 15+ patents in the period are small firms. This compares favorably with the 33 percent figure from a previous study by the authors in 2003, and is just slightly below the 41 percent figure from a study by the authors in 2004.
- Small firms obtain many more patents per employee than do large firms. This result is quantified to show that this is not a small-firm large-firm phenomenon, but is actually a firm size issue at all levels. In particular, even within the small firm domain, companies with fewer than 25 employees will have a higher patent-to-employee ratio on average than firms with 50 employees, which will in turn have a higher patent-to-employee ratio than firms with 100 employees, and so on.
- Small firm patents outperform large firm patents on a number of impact metrics including growth, citation impact, patent originality, and patent generality. These metrics have been used for decades to measure the innovativeness of firms, labs, and agencies. The metrics have been validated and shown to correlate with increases in sales, profits, stock prices, inventor awards, and other positive outcomes. This suggests that the patents of small firms in general are likely to be more technologically important than those of large firms.
- Although small firms make up only 6.5 percent of all the patents in the database, they patent at

a higher rate in some technologies, particularly health-related. Ranked by broad technology areas, most small firms fall in health-related technologies (biotechnology, pharmaceutical, and medical devices) and information technology categories (communications/telecommunications, semiconductors, computer hardware and software).

Scope and Methodology

This project created a detailed database of 1,293 small and large technology firms, as well as more than 1 million patent records from these firms. The database was used to highlight differences between the patent activity of small and large firms, and also to test several hypotheses about emerging technologies and industries.

This report was peer reviewed consistent with the Office of Advocacy's data quality guidelines. More information on this process can be obtained by contacting the director of economic research at **advocacy@sba.gov** or (202) 205-6533.

References

This study is the third in a series that examines small business patent activity. The other are

- *Small Firms and Technology: Acquisitions, Inventor Movement, and Technology Transfer* **http://www.sba.gov/advo/research/rs233tot.pdf**
- *Small Serial Innovators: The Small Firm Contribution to Technical Change* **http://www.sba.gov/advo/research/rs225tot.pdf**

Ordering Information

The full text of this report and summaries of other studies performed under contract with the U.S. Small Business Administration's Office of Advocacy are available on the Internet at **www.sba.gov/advo/research.** Copies are available for purchase from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
(800) 553-6847 or (703) 605-6000
TDD: (703) 487-4639
www.ntis.gov
Order number: PB2009-101770
Paper A05 ($48.00)
Microfiche A05 ($33.00)
CD-ROM A00 ($30.00)
Download A00 ($15.00)

To receive email notices of Advocacy's newsletter, press, regulatory news, and research, visit **http://web.sba.gov/list.** For really simple syndication, visit **www.sba.gov/advo/rsslibrary.html**.

Executive Summary

This project created a detailed database of 1,293 small and large technology firms, as well as more than 1 million patent records from these firms. The database was used to highlight differences between the patent activity of small and large firms, and also to test several hypotheses about emerging technologies and industries.

The findings are separated into two groups—overall findings and findings from each major chapter.

Overall Findings

- The authors compiled a database of 1,293 top patenting firms consisting of any U.S. firm with 15 or more U.S. patents issued between 2002 and 2006. Of the 1,293 firms, 504 had 500 or fewer employees, 760 had more than 500 employees, and no size information could be obtained for 29 firms. That is, 40 percent of the firms with 15+ patents in the period are small firms. This compares favorably with the 33 percent figure from the previous study by the authors in 2003, and is just slightly below the 41 percent figure from the study by the authors in 2004.

- Thirty-eight percent of the small firms in the database are publicly listed firms. This is a rather surprising result given that less than one-tenth of 1 percent of all small firms are publicly listed. This suggests that large patent-holding small firms are much more likely to go public than nonpatenting firms.

- Small firms obtain many more patents per employee than large firms. This is not a new result, but it is quantified to show that it is not a small-firm / large-firm phenomenon, but is actually a firm-size issue at all levels. In particular, even within the small firm domain, companies with fewer than 25 employees will have a higher patent-to-employee ratio on average than firms with 50 employees, which will in turn have a higher patent-to-employee ratio than firms with 100 employees, and so on. This relationship appears to be governed by a power law or exponential law, but more research would be needed beyond the scope of this project to exactly determine the relationship.

- Small firm patents outperform large firm patents on a number of impact metrics including growth, citation impact, patent originality, and patent generality. These metrics have been used for decades to measure the innovativeness of firms, labs, and agencies. The metrics have been validated and shown to correlate with increases in sales, profits, stock prices, inventor awards, and other positive outcomes. This suggests that the patents of small firms in general are likely to be more technologically important than those of large firms. This is not to suggest that every small firm patent will be more valuable than its large firm counterpart, but that statistically the patents of small firms perform better on average than those of large firms.

- Although small firms make up 40 percent of all firms in the database, they make up just 6.5 percent of all the patents in the database. IBM, the largest patenting entity, with 34,700+ patents since 1997, actually has more patents than all of the small firms combined over the same period.

(Of course, IBM has more than 395,000 employees, while the small firms in the database have an average of 143, so that IBM is equivalent to more than 2,700 of these small firms.)

- Although small firms make up only 6.5 percent of all the patents in the database, they patent at a higher rate in some technologies, particularly health-related. Ranked by broad technology areas, most small firms fall in health-related technologies (biotechnology, pharmaceutical, and medical devices) and information technology categories (communications/telecommunications, semiconductors, computer hardware, and software). In the health-related areas small firms have a higher share of patenting than expected (that is, exceeding the 6.5 percent overall rate of small firm patenting) while in the information technology areas (other than telecommunications) large firms outpatent small firms more than expected.

Share of Small Firms in Database

- Small firms make up 40 percent of all U.S. companies with 15+ patents issued in the period 2002-2006. This is slightly below the 41 percent found in the previous study by the authors from 2004, but greater than the 33 percent found in the study from 2003. The leveling off of the small firm share tracks a similar leveling off of the share of industrial R&D accounted for by small firms.

- The small firms in this study are much younger than the large firms with 15+ patents. Ninety percent of the large firms are 15 or more years old, while only 43.5 percent of small firms are 15 or more years old. Moreover, one in five small firms has been in business less than 10 years, compared with only about one in 40 large firms.

- Of the small firms in this study, 306 were not in the previous study—that is, 306 small firms had 15 or more patents in 2002-2006, but fewer than 15 in 1998-2002. The large number of new entrants led to a hypothesis that the ratio of small to large firms would be greater than the 41 percent of the previous study. However, the number of new small firms entering the study is offset by a large number that fell out of the database. Six percent of small firms in the last study have now become large firms, while 17 percent have merged or been acquired. Most of the remaining small firms that dropped out did so because they fell below the 15-patent threshold, while another 4 percent dropped out because they became troubled or declared bankruptcy.

Small Firm Participation in Emerging Technologies

- Identifying emerging technologies is a difficult undertaking because it essentially involves predicting the future. Truly emerging technologies may not reveal themselves to be important for many years. However, as part of a project recently completed for the National Institute of Standards and Technology (NIST) and the Technology Administration, the authors were able to identify a number of real-time parameters useful for marking patent clusters as likely to contain emerging technologies. The same project identified the top 100 high-scoring emerging clusters.

  The project used these clusters to test the hypothesis that small firms would pursue different emerging technologies than would large firms—a reasonable hypothesis, given that small and large firms often patent in different technology areas in general. However, while most emerging clusters were dominated by either small or large firms, the patterns within clusters varied greatly. In some clusters where both large and small firms participated, they actually pursued different technologies, thus supporting the hypothesis. In other cases, the small and large firms competed directly in similar technologies, thus undermining the hypothesis. Finally, in one case where the emerging technology was clearly led by a small firm, and the emerging cluster contained only patents of that firm, there was still evidence that large firms were pursuing the same technology.

- Hence, while testing the hypothesis produced inconclusive results, it was useful in highlighting the following phenomena:

  1. Most emerging clusters are dominated by either small or large firms, but rarely both. A simulation conducted with similar parameters for 10,000 iterations could never get close to the distribution that occurs naturally. This suggests that small and large firms do in general work in differing emerging technologies and that the results are not random.
  2. In isolated cases, even though small and large firms appear to be working in the same space, there were differences.
  3. In other isolated cases, small and large firms competed directly with one another in a race to develop a technology.
  4. In still other isolated cases where it appears small firms have a dominant role in a technology, large firms are also pursuing the technology, perhaps at a smaller scale.

  On the whole, small and large firms tend to pursue different emerging technologies. However, the number of exceptions to this pattern means that there is no statistically significant difference. It does appear that small firms are more likely to attempt to build a business around a new emerging technology, whereas in general large firms work on emerging technologies in order to improve an existing product line or business unit.

- A more interesting result was independent of the hypothesis: small firms are much more likely to develop emerging technologies than are large firms. This is perhaps intuitively reasonable given theories on small firms effecting technological change, but the quantitative data here support this assertion. Specifically, although small firms account for only 8 percent of patents granted, they account for 24 percent of the patents in the top 100 emerging clusters. This means that they produce three times as many patents as one would expect in this special patent set. Put another way, approximately one in 31 small firm patents are contained in the top emerging clusters, compared with one in 117 large firm patents.

Custom Classification of Small Firms

- Some traditional industries figure prominently in these data. Unexpected innovators are found in batteries, gaming machines, packaging, and retail display (showcase) manufacture. These industries are not emerging, but they may be undergoing a technology-driven renaissance. In some cases, small firms may be able to play the role of disruptive innovators and enter such established industries.

- Two types of emerging industries were identified. The first type is concerned with commercializing new technology. The second, while innovative, is emerging because of innovations in business models. In both types there are industries in which small firms predominate.

  Candidates for new technology-based emerging industries include alternative energy, filtration equipment, and radio frequency identification (RFID). New technology-based emerging industries driven by small firm innovators also include manufacturing industries such as imaging and display, nanotechnology, photonics/optical components, and power supplies. Emerging industries driven by technology are also in the information sector and include digital security, electronic design automation (EDA), and positioning info/services/devices. These industries have small and large firm participants.

  Business model innovation principally involves separating manufacturing from all other business processes. Specialists in manufacturing seem to be large firms, presumably because of the capital required. Examples found here include semiconductor assembly and testing outsourcing, as well as contract manufacturing in a variety of types of production processes.

  Some firms contract out their manufacturing. Such firms may do research and development (R&D), product engineering, design, development, marketing, and distribution. However, if they do not manufacture, they are a different type of firm at the fundamental two-digit NAICS (North American Industry Classification System) level. Such firms are best classified as information firms. The contribution of this study is to acknowledge these categories of firms, and to point out that the new information industries tend to be populated by small firms with strong patent records. Presumably without manufacturing, capital requirements are low, allowing small firms to compete. In addition, since the firms must work with others to have their goods made, they cannot rely on secrecy to protect innovation. Therefore, innovation theory suggests such firms will aggressively pursue patent protection. Examples of these industries include: biomedical pipeline, biotechnology pipeline (nonbiomedical), communications technology design and marketing, and fabless semiconductor. Biomedical pipeline and fabless semiconductor are two of the largest categories in this study.

Contents

I. Introduction and Rationale

This report describes the key findings in a project to create a detailed database of 1,293 small and large technology firms. The firms were chosen because they have at least 15 patents in the last five years—in total, more than 1 million patents. The database was then used to highlight differences between the patent activity of small and large firms and to test several hypotheses about emerging technologies and industries.

The project, "SBA3," extended previous studies of small business patenting activities conducted by the authors for the U.S. Small Business Administration (SBA), Office of Advocacy. The first two studies, SBA1[1] and SBA2,[2] established the existence of a cohort of independent, nonbankrupt, for-profit small firms with 15 or more patents over a five-year period. Because small firms often find patenting too expensive and difficult, and thus make little use of the patent system,[3] few would even have guessed such firms exist. SBA1 and SBA2 were the first studies of small business patenting that were based on a large, rich, and well-defined dataset that encompassed the universe of significant patenting companies, rather being based on a sampling of a specialized patent set, or on the results of a survey.

In SBA3 the dataset again consists of all companies with 15 or more patents in the last five years. In addition to identifying the companies' 15+ patents, numerous extensions are included in this project, such as a mapping to NAICS codes, the inclusion of the contractor's (1790 Analytics) validated metrics, citation indicators, and emerging technology models.

Firms in emerging industries define America's industrial future. Economists have studied emerging industries ever since Joseph Schumpeter first asked what accounts for the genesis of new industries. New industries are associated with entrepreneurial activities, which combine knowledge in new and novel ways. Most important, emerging industries provide the platform for future economic growth. The gales of creative destruction, which Schumpeter so colorfully described as driving economic growth, are in part attributable to the emergence of new industries.

There is great academic and policy interest in identifying and tracking emerging technologies and industries. However, by their very nature, emerging industries do not fit neatly into existing classification schemes, making it difficult to identify them. Instead, researchers often focus on industries favored with extensive media coverage, such as biotechnology, nanotechnology, and the Internet. Although consequential, these industries likely are atypical, and the understanding of how new industries emerge would be greatly enhanced if there were a systematic quantitative method of identifying groups of firms engaged in similar new businesses.

Based on these ideas and others that developed while producing SBA1 and SBA2, the authors proposed testing three hypotheses in this project:

---

[1] Diana Hicks et al., *Small Serial Innovators: The Small Firm Contribution To Technical Change*, U.S. Small Business Administration, Office of Advocacy, Contract No. SBAHQ-01-C-0149, February 2003.
[2] Anthony Breitzman et al., *Small Firms and Technology: Acquisitions, Inventor Movement, and Technology Transfer,* U.S. Small Business Administration, Office of Advocacy, Contract No. SBAHQ-02-M-0491, January 2004.
[3] Judith Obermayer., *The Role of Patents in the Commercialization of New Technology for Small Innovative Companies,* final report for the U.S. Small Business Administration, Office of Advocacy, Research & Planning Inc., Cambridge MA, August 1981.

Hypothesis 1: The number and percentage of small firms in the database will increase substantially beyond the 41 percent measured in SBA2.

Hypothesis 2: Small and large firms differ in the emerging technologies they pursue.

Hypothesis 3: NAICS codes will not satisfactorily represent the businesses of the serial innovators. In the differences between NAICS codes and custom classification will be found emerging industries.

This report explores each of these hypotheses in detail.

II. Overview of Small Business Patent Database

A. Introduction

A key tool used in this project is a carefully constructed database of small and large technology firms, specifically, all U.S. firms that obtained 15 or more patents in the five-year period from 2002 to 2006. This is similar to the databases of patents through 2000 and 2002 that the authors built in previous projects for the SBA, referred to in this report as SBA1[4] and SBA2.[5] The current database is a unique resource, consisting of 1,293 firms and more than a million patent records. In addition to patent information, the database contains information on the number of employees, revenues, and industry information where available. This section discusses how the database was built and explores some unexpected results.

B. Method

The database built for this project leverages the existing 1790 Analytics patent database consisting of all organizations with 40 or more patents issued in the last five years. It is important to understand that the patent office records assignees and not necessarily companies. Therefore it is not trivial to locate all of the patents owned by Microsoft or General Motors. Patents owned by General Motors and others are a mixture of names of firms, establishments, subsidiaries, and variants of firm names. Mergers and acquisitions are also constantly changing the status of firms. As an example, large firms like General Motors and Procter & Gamble patent under more than 100 names. Extreme cases of firms that have a history of mergers, such as Glaxo-SmithKline, will have patents under more than 300 names.

The core database used by 1790 Analytics on a day-to-day basis, and licensed to information companies such as Thomson Scientific, tracks nearly 4,000 organizations in three patent systems and is made up of more than 50,000 individual subsidiary and variant assignee names. This database is maintained by a data manager with more than 20 years experience tracking and standardizing assignee names.

The 1790 database tracks U.S. firms, foreign firms, nonprofits, universities, and government agencies. The database used for this project is a subset of the main database consisting of U.S.-based companies. In addition, this project extends the database to include U.S. firms with 15 or more patents granted between 2002 and 2006. An additional extension made to the database includes a lookup of the number of employees for each of the 1,293 firms, as well as identification of revenues, line of business, and SIC (Standard Industrial Classification) and NAICS (North American Industry Classification System) codes where available. These data were identified using multiple sources including Mergent/Moody's International, Lexis/Nexis, annual reports, and Dun & Bradstreet.

Assembling the database was by far the most time-consuming part of the project, and the authors and 1790 staff were scrupulous in this task and well aware of the hazards of firm identification, particularly when it comes to small businesses. The story of Tether's reanalysis of Pavitt's work is worth mentioning here. Pavitt analyzed 4,278 innovations commercialized in the United Kingdom since 1945.[6] Tether reanalyzed the Pavitt

---

[4] Diana Hicks et al., *Small Serial Innovators: The Small Firm Contribution To Technical Change*, U.S. Small Business Administration, Office of Advocacy, Contract No. SBAHQ-01-C-0149, February 2003.
[5] Anthony Breitzman et al., *Small Firms and Technology: Acquisitions, Inventor Movement, and Technology Transfer,* U.S. Small Business Administration, Office of Advocacy, Contract No. SBAHQ-02-M-0491, January 2004.
[6] Keith Pavitt, M. Robson and J. Townsend. 1987. "The Size Distribution of Innovating Firms in the UK: 1945-1983." *Journal of Industrial Economics 35*: 297-316.

data in the 1990s and re-checked the classification of the firms as small or large at the time of the innovation. He found that some of the subsidiaries of large firms had been misclassified as small firms. The net result was that the reclassification by Tether eliminated the statistical significance in the headline result of Pavitt's study that small firms were becoming more important to innovation.[7] This cautionary tale points to the need to be very careful in assembling company data.

The cutoff date for the company structure is December 31, 2007. Any firms that merged after that date are as they were at the end of 2007. Similarly, while in general bankrupt companies were removed, any that became troubled after December 31, 2007, were not removed.

In general for this project, all subsidiaries are combined with their parent companies. For example, the patents of Ethicon and Cordis Corp. are combined with the ultimate parent company, Johnson and Johnson, in the database. Similarly the U.S. biotechnology company Genentech is removed completely because it is majority-owned by the foreign firm Roche Holdings, and foreign firms are not part of this study. However, private equity firms, which have become more prominent since the last project, are handled differently, since these firms may hold a variety of firms for a short period of time. In this project, if an equity firm holds a majority interest in one or more firms that run as independent companies, the companies are treated as independent companies within the database. For example Johns Manville, Waterpik, and Polaroid, among others, operate as companies but are majority-owned by holding companies like Berkshire Hathaway or private equity firms like the Petters Group Worldwide.

In summary, this project is built on a database of more than 1 million patents from 1,293 U.S. firms with 15 or more U.S. patents granted between January 1, 2002, and December 31, 2006. The firms run the gamut from the small business Rain Bird with 15 patents related to sprinklers in the period of study, to computer giant IBM with 16,700 U.S. patents in the period.

C. Results

The project tests three hypotheses using the database, with results presented in later chapters. However, it is worth exploring some basic results to give the reader an overview of the database.

Table II.1 reveals the breakdown of the 1,293 firms covered in the database: 760 are large firms, 504 are small, and no size information could be obtained for 29 firms. These latter firms are very likely to be small firms based on the dearth of information and the small number of patents. Since they represent only 2 percent of the total, including or excluding them from any analysis would not change the results in a significant way.

[7] Bruce S. Tether., I.J. Smith and A.T. Thwaites. 1997. "Smaller enterprises and innovation in the UK: the SPRU Innovations Database revisited." *Research Policy* 2: 19-32.

Table II.1 Summary Statistics for U.S. Company Patent Database

Company Size	Number of Companies	Percent Identifiable	Percent of Total	Number Publicly Listed	Percent Publicly listed	Average Number of Patents 2002-2006
Large	760	60	59	583	77	307
Small	504	40	39	191	38	38
Unknown	29		2			21
Total Known	1,264			774	61	
Grand Total	1,293					

Table II.1 shows that 40 percent of the firms with 15 or more patents in the five-year period are small.[8] This is a higher percentage of small firms than the 33 percent found in the SBA1 project and slightly lower than the 41 percent from the SBA2 project. A detailed discussion of the fates of firms entering and exiting the databases in these projects can be found in later in the report in section III.

Table II.1 also reveals that 60 percent of all the firms and 38 percent of the small firms in the database are publicly listed. Publicly listed firms here are those with ticker symbols actually traded on the major U.S. exchanges, and not companies that are technically public, but not traded or only traded over the counter. According to Bureau of the Census statistics, 6 percent of U.S. manufacturing firms with employees are publicly held.[9] Using the more stringent publicly traded criteria, fewer than 0.1 percent of all firms are publicly traded.[10] The large share of publicly traded firms in this dataset is notable. It suggests that firms of all sizes with patented technologies are more likely to become successful enough to go public than firms that do not patent.

Table II.2 shows additional summary statistics pulled from the database. For example, large firms in the database tend to be fairly large, with an average of $7 billion in sales and an average of 18,489 employees (28 firms have more than 100,000 employees and 114 firms have sales exceeding $10 billion, including nine firms with sales exceeding $100 billion). Not surprisingly, the large firms have more patents, on average, than the small firms, but the small firms obtain more patents per employee than the large firms.

Patents per Employee

The statistic from Table II.2 showing that small firms obtain more patents per employee than larger firms is not a new result and was discussed extensively in SBA1. What was not discussed, however, was that the relationship is actually more complex. It is tempting to take the two figures and conclude that small firms are more efficient in generating patents than large firms, but as Table II.3 shows, the advantage in patents per employee decreases with size even within small firms.

[8] Throughout this project a firm with 500 or fewer employees is considered a small firm.
[9] Source: U.S. Census Bureau, 2002 Survey of Business Owners, found using American FactFinder.
[10] The 0.1 percent calculation comes from the 3,647 publicly traded companies identified via Google Finance http://finance.google.com [accessed August 14, 2008] divided by the estimated number of employer firms 5,696,600 in 2003 obtained from U.S. Small Business Profile, U.S. Small Business Administration, Office of Advocacy, 2004, http://www.sba.gov/advo/stats/profiles/04us.pdf [accessed August 14, 2008]. Even though the company counts are five years old, the less than 0.1 percent statistic is reasonable. The figure would not change if the number of employer firms decreased by more than 2 million, and of course it would always be correct if there was any increase in employer firms.

Table II.2 Additional Summary Statistics for U.S. Company Patent Database

Company Size	Average Sales (Dollars)	Average Number of Employees	Average Sales per Employee (Dollars)	Median Sales per Employee (Dollars)	Average Number of Patents 2002-2006	2002-2006 Patents per Hundred Employees
Large	7,405,416,093	18,489	400,532	250,000	307	1.7
Small	39,420,941	143	275,018	105,971	38	26.5

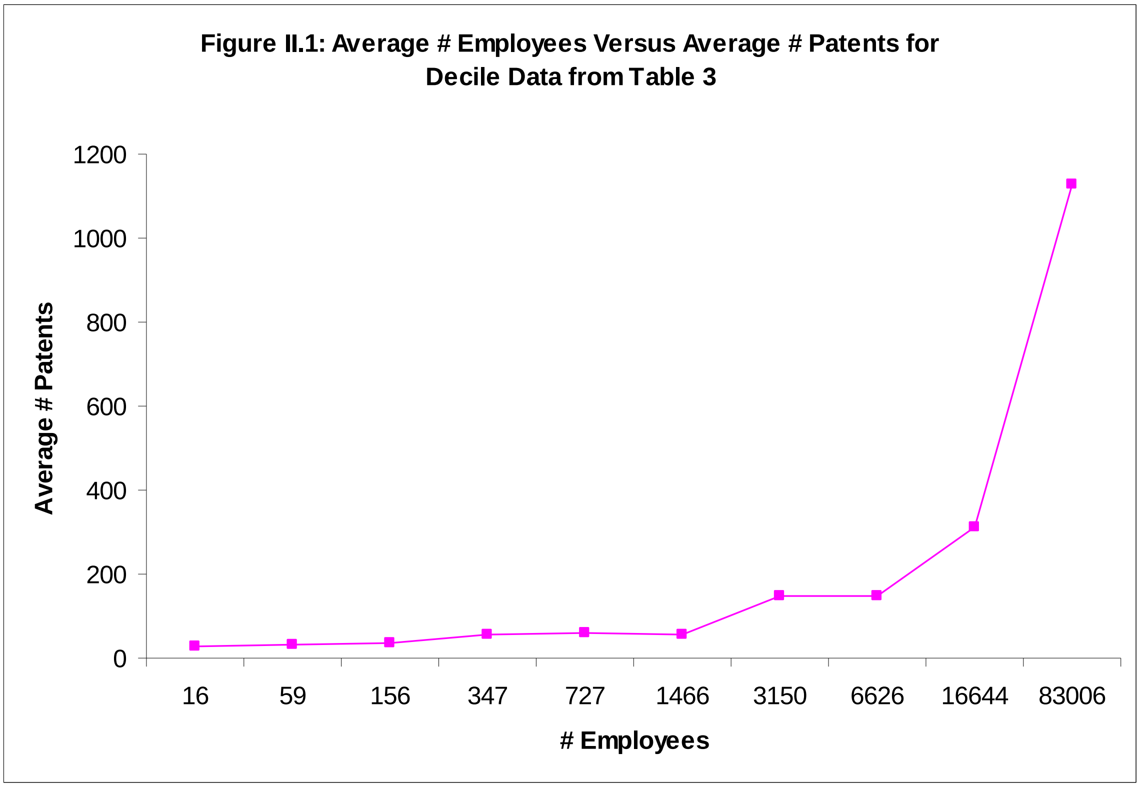
Table II.3 Patents per Employee by Size Percentile for Companies in the Database

Firm Size Decile	Average Number of Employees	Average Number of Patents 2002-2006	Patents per Employee
1	15.8	29.8	1.886
2	58.9	32.4	0.550
3	156.1	34.6	0.222
4	347.3	54.9	0.158
5	727.3	60.0	0.083
6	1,465.5	58.0	0.040
7	3,149.6	149.3	0.047
8	6,625.8	147.1	0.022
9	16,643.9	312.1	0.019
10	83,006.0	1,127.9	0.014

Table II.3 was constructed by taking the 1,264 firms with known employee counts and rank-ordering them by number of employees. This ranked list was then sliced into deciles. For example, the smallest 126 firms are put into the first decile. On average, these companies have 15.8 employees and 29.8 patents in the five-year period. The 127 firms in the second decile have 58.9 employees on average and 32.4 patents in the five-year period. As the employee count increases, the patents per employee decrease. This occurs even in the first four deciles, which all consist of small firms, defined as those with 500 or fewer employees.

Plotting the employee count against the patent count produces Figure II.1. It appears that the relationship between employees and patents is governed by a slow-growing power law or exponential function. Deriving such a relationship is well beyond the scope of this project, as much more time and data would be needed to identify and validate the relationship.[11] It is, however, an oversimplification to state that small firms are more efficient than larger firms. Instead, it appears that there is a nonlinear relationship between firm size and numbers of patents where, as the number of employees increases, the level of patenting increases at a slower rate. This relationship negatively affects the patents per employee output of all size firms as they start to grow.

[11] It would be imprudent to develop a model that did not take into consideration the millions of firms with fewer than 15 patents. Expanding the database to consider firms with one, five, or even 10 or more patents would be a huge undertaking to obtain a result that would likely have little or no policy implications.



Patent Scorecard

Patent scorecards statistically compare the patents of large firms versus small firms using a variety of metrics described below.

Number of Patents: The number of U.S. patents granted to a company in a given time period.

Percent Growth: The growth in U.S. patents from one time period to another—here defined as the percent change in patent activity from the five-year periods 1998-2002 to 2003-2007.

2007 Pipeline Impact: A measure of how frequently a set of patents is cited by subsequent 2007 patents. Highly cited patents (those cited by many subsequent patents) tend to contain technological ideas of particular importance. The 2007 pipeline impact is a slightly different citation indicator in that it looks at patents from the past five years and measures how frequently they are cited by recent patents. In other words, it shows the strength of an organization's (or set of organizations') patent pipeline.

The pipeline impact for each patent within a portfolio is calculated by taking the number of times the patent has been cited by 2007 patents, and dividing this number by the mean number of citations received by patents from the same U.S. Patent and Trademark Office (USPTO) classification and issue year. The mean pipeline impact for any patent is therefore 1.0. The pipeline impact for a patent portfolio is then calculated as the mean of the impact values for each patent within it. The expected pipeline impact value for any patent portfolio is therefore also 1.0.

A pipeline impact value above 1.0 shows that a patent set is cited more frequently than expected by recent patents. For example, a pipeline of 1.50 shows that an organization's patents are cited 50 percent more frequently than expected. A value below 1.0 shows that an organization's patents are cited less frequently than expected. For example, a pipeline impact of 0.80 shows that an organization's patents are cited 20 percent less frequently than expected. All normalizations here are calculated against the full USPTO database, which includes unassigned public sector and foreign patents (as well as patents from firms with fewer than 15 patents in the last five years). The patents in this study accounted for 31 percent of U.S.-issued patents in 2002-2006.

2007 Pipeline Generality: Whereas pipeline impact measures the level of impact of a company's patent portfolio, pipeline generality measures the breadth of this impact. It is calculated based on the range of USPTO classifications (USPOCs) that cite a company's patent portfolio.

The 2007 pipeline generality for a single patent is calculated based on the dispersion across USPOCs of the patents granted in 2007 that cite to it. If a patent is cited only by other patents in the same USPOC, its breadth of influence is regarded as relatively low. On the other hand, if a patent is cited by patents across a range of USPOCs, it is regarded as having a more general influence.

In the same way as the pipeline impact, the pipeline generality for a patent is divided by the expected value for a patent of the same USPOC and year, so the expected generality for any patent is 1.0. The pipeline generality for a patent portfolio is then calculated as the mean of the generality values of all patents within it. The expected value for pipeline generality is therefore 1.0. Values above this denote a patent portfolio with a higher than expected breadth of influence, while values below 1.0 show a portfolio whose patents have a relatively narrow influence.

Pipeline Originality: Pipeline originality measures the breadth of technologies cited by an organization's patents. It is based on the idea that patents that cite a wide range of technologies are more likely to contain original ideas than patents that build upon a narrow range of similar technologies.

Pipeline originality is calculated in a similar way to pipeline generality, except that it examines patents referenced by a portfolio, rather than the citations the portfolio receives from later patents. In the same way as pipeline generality, the mean pipeline originality for any patent, or patent portfolio, is 1.0. Values above this show a portfolio that builds on a wide range of technologies, and thus has more likelihood of containing original ideas. A value below 1.0 shows a portfolio that builds on a narrow set of previous technologies, and so may contain many patents that represent incremental improvements on previous technologies.

Citation Index: The citation index is a traditional citation measure used by analysts to measure the impact of papers and patents. It is similar to the pipeline impact in that it is a normalized citation measure with an expected value of 1.0 for an average portfolio. The main distinction is that the citation index examines all citations to a portfolio whereas the pipeline impact measure examines citations from 2007. The shortcoming of the citation index is that if a portfolio starts to age and to lose impact over time, it will not be reflected in the citation index in the way it would for the pipeline impact measure.

Table II.4 is a patent scorecard that measures the previously defined metrics for the set of all U.S. patents produced by small and large firms in the database. This table reveals that, other than in the overall count of patents, the small firms outperform the large firms in every measure. Specifically, the small firm output of

Table II.4 Patent Scorecard for 1,264 Firms in Patent Database

Firm Size	Number of Firms	1998-2002 Patents	2003-2007 Patents	Percent Growth	2007 Pipeline Impact	2007 Pipeline Generality	Pipeline Origin-ality	Citation Index
Small Business	504	12,303	18,838	53	1.73	1.79	1.07	1.69
Large Business	760	208,250	229,118	10	1.18	1.24	1.01	1.19

patents has increased 53 percent in the last five years compared with the previous five-year period. Although much of this growth rate has to do with the number of new entrants into the database, as well as small firms that did not exist in the first period, the growth rate is still quite impressive and much larger than the 10 percent growth in patenting for large firms.

The pipeline impact figure for both small and large firms is above average, but the patents of small firms are much more highly cited by 2007 patents than are the patents of large firms. The 1.73 pipeline impact figure suggests that the patents of small firms are cited 73 percent more by recent patents than is typical for patents of the same age and patent classification. This is comparable to the results using the standard citation index. By this measure, the patents of small firms are cited 69 percent more than average patents of the same age and class. Patents of large firms are cited more than expected, but not as often as patents of small firms.

Numerous validation studies have shown an association between highly cited patents and various positive outcomes. For example, patents that have won inventor awards tend to be highly cited. Also, firms with highly cited patents have shown increases in sales, profits, and stock price. The interested reader can find a review of validation studies related to patent citation analysis in Breitzman and Mogee, 2002.[12]

The small firms also outperform their larger counterparts in the generality and originality metrics. This suggests patents for small firms tend to combine a wider range of technologies in order to create new inventions, and they also, in turn, are built upon by a greater variety of technologies in general.

Technology Categories

As part of this general overview of the patent database it would be useful to give the reader an idea of the general technology categories in which small and large firms patent. In past projects, the U.S. Patent Office classifications and international patent classifications (IPC), which are very specific and useful for searching for patents at a very fine level, are not as good for categorizing large sets of patents at a policy level. Therefore, 1790 Analytics developed a concordance that maps the roughly 17,500 IPC subclasses to 10 broad technology categories and 74 subcategories.

Table II.5 shows the patent activity by small and large firms in each of the subcategories. It also shows the percentage of patents in the category produced by small and large firms. The total at the bottom of the table indicates that large firms produce 93.5 percent of the total and small firms 6.5 percent of the total. Therefore, in subcategories where small firm patenting exceeds 6.5 percent, the category has a small firm excess, and

[12] Anthony Breitzman and Mary Mogee, “The Many Applications of Patent Analysis,” *Journal of Information Science*, 28(3), 187-205, 2002.

Table II.5 Small and Large Firm Patenting by Technology Category and Subcategory (1997-2007)

Technology and Subcategory	Number of Large Firm Patents	Number of Small Firm Patents	Percent Large Firm	Percent Small Firm	Small Firm Excess	Significance (P<)
Agriculture/Husbandry/Food/Tobacco	6,977	813	89.56	10.44	3.94	0.001
Agriculture minus Agrigenetics	2,682	264	91.00	9.00	2.50	0.001
Agrigenetics/Agricultural Biotechnology	1,958	409	82.70	17.30	**10.80**	0.001
Animal Husbandry/Hunting/Trapping/Fishing	405	80	83.50	16.50	**10.00**	0.001
Food	1,695	57	96.70	3.30	-3.20	0.001
Tobacco	237	3	98.80	1.30	**-5.20**	0.001
Building/Constr./Housing MaterialsandFixtures	7,055	450	94.00	6.00	-0.50	NS
Building/Construction Materials	1,254	50	96.20	3.80	-2.70	0.001
Foundations	57	21	73.10	26.90	**20.40**	0.001
Furniture/House Fixtures	4,177	304	93.20	6.80	0.30	NS
Water Suppy/Plumbing/Pipes/Waste Treat.	1,567	75	95.40	4.60	-1.90	0.01
Chemical and Chemical Processes	49,353	3,435	93.49	6.51	0.01	NS
Centrifuges	141	5	96.60	3.40	-3.10	NS
Cleaning	1,298	70	94.90	5.10	-1.40	0.05
Other Chemical Processes	3,321	282	92.20	7.80	1.30	0.01
Waste Treatment	48	0	100.00	0.00	**-6.50**	NS
Dyes/Paints/Coatings	5,329	184	96.70	3.30	-3.10	0.001
Glass/Ceramic/Cement	1,696	76	95.70	4.30	-2.20	0.001
Metal Working	2,249	84	96.40	3.60	-2.90	0.001
Metallurgy	3,637	142	96.20	3.80	-2.70	0.001
Other Chemistry	5,959	254	95.90	4.10	-2.40	0.001
Other Organic Compounds	11,788	1,884	86.20	13.80	7.30	0.001
Petroleum/Gas/Coke	3,391	73	97.90	2.10	-4.40	0.001
Resins/Polymers/Rubber	10,496	381	96.50	3.50	-3.00	0.001
Electrical	33,619	2,338	93.50	6.50	0.00	NS
Electrical Devices	21,594	1,082	95.20	4.80	-1.70	0.001
Electrical Lighting/Displays	3,308	259	92.70	7.30	0.80	NS
Power Systems	8,717	997	89.70	10.30	3.80	0.001
Health	48,717	7,934	85.99	14.01	7.51	0.001
Biotechnology	7,077	2,663	72.70	27.30	**20.90**	0.001
Cosmetics/Health and Beauty Aids	1,127	29	97.50	2.50	-4.00	0.001
Dentistry/Dental Preparations	1,179	162	87.90	12.10	5.60	0.001
Diagnosis/Surgery/Medical Instruments	23,962	2,668	90.00	10.00	3.50	0.001
Nuclear and X-Ray	4,304	305	93.40	6.60	0.10	NS
Pharmaceuticals	11,068	2,107	84.00	16.00	9.50	0.001
Industrial Processes/Tools/Equipment	19,453	1,124	94.54	5.46	-1.04	0.001
Hand Tools/Machine Tools	9,708	501	95.10	4.90	-1.60	0.001
Bottle and Jar Filling/Dispensing	852	16	98.20	1.80	-4.60	0.001
Compressors and Pumps	1,738	58	96.80	3.20	-3.30	0.001
Filtration	1,338	72	94.90	5.10	-1.40	0.05
Pulverising/Milling	197	16	92.50	7.50	1.00	NS
Spraying Apparatus/Nozzles	890	124	87.80	12.20	5.70	0.001
Shaping/Extruding/Working of Plastics	3,217	244	93.00	7.00	0.60	NS

Wood and Paper Mfg.	1,513	93	94.20	5.80	-0.70	NS
Instruments/Measuring, Testing and Control	40,038	2,849	93.36	6.64	0.14	NS
Clocks/Watches/Time Pieces	224	9	96.10	3.90	-2.60	NS
Control Devices	2,187	114	95.00	5.00	-1.50	0.01
Counting/Sorting/Handling coins & currency	330	37	89.90	10.10	3.60	0.01
Measuring and Testing	20,005	1,436	93.30	6.70	0.20	NS
Optics/Photography/Electrophotography	17,292	1,253	93.20	6.80	0.30	NS
Mechanical	32,792	997	97.05	2.95	-3.55	0.001
Aerospace/AeroNautics	2,881	36	98.80	1.20	**-5.30**	0.001
Engines and Parts	11,345	209	98.20	1.80	**-4.70**	0.001
Motor Vehicles and Parts	8,817	346	96.20	3.80	-2.70	0.001
Fasteners	762	16	97.90	2.10	-4.40	0.001
Hydraulics	526	51	91.20	8.80	2.40	0.05
Locks/Hinges/Deadbolts	1,311	55	96.00	4.00	-2.50	0.001
Other Mechanical	3,849	135	96.60	3.40	-3.10	0.001
Other Transport	2,972	108	96.50	3.50	-3.00	0.001
Rail Transport	329	41	88.90	11.10	4.60	0.001
Other	41,947	2,065	95.31	4.69	-1.81	0.001
Earth Moving/Drilling/Mining/Blasting	5,659	199	96.60	3.40	-3.10	0.001
Heating/Ventilation/AC/Refrigeration	5,688	242	95.90	4.10	-2.40	0.001
Cranes/Winches/hoisting	269	16	94.40	5.60	-0.90	NS
Decorative Arts	389	10	97.50	2.50	-4.00	0.01
Education Systems/Teaching Aids	229	45	83.60	16.40	9.90	0.001
Life Saving/Fire Fighting	524	19	96.50	3.50	-3.00	0.01
Musical Instruments	185	38	83.00	17.00	**10.60**	0.001
Other	12,097	521	95.90	4.10	-2.40	0.001
Printing Presses/Printing (mechanical)	1,201	5	99.60	0.40	**-6.10**	0.001
Signs/Seals/Displays	242	27	90.00	10.00	3.50	0.05
Stationary /Binders/Labels/Writing Implements	542	13	97.70	2.30	-4.10	0.001
Containers	3,824	81	97.90	2.10	-4.40	0.001
Packaging/Labeling/Conveying	4,877	217	95.70	4.30	-2.20	0.001
Sports/Games/Amusements	3,169	407	88.60	11.40	4.90	0.001
Textiles and Apparel	3,052	225	93.10	6.90	0.40	NS
Semiconductors/Computers/Communication	185,702	10,299	94.75	5.25	-1.25	0.001
Communications/Mostly Telecom	54,040	4,316	92.60	7.40	0.90	0.001
Computer Hardware	37,626	1,389	96.40	3.60	-2.90	0.001
Computer Peripherals	11,992	646	94.90	5.10	-1.40	0.001
Computer Software	26,948	887	96.80	3.20	-3.30	0.001
Information Storage	20,544	1,326	93.90	6.10	-0.40	0.01
Semiconductors/Solid-State Dev./ Electronics	34,552	1,735	95.20	4.80	-1.70	0.001
Total	465,653	32,304	93.50	6.50	0.00	0.001

the value of that excess (consisting of the percentage of small firm patents in the category minus 6.5 percent) can be found in the second to last column. The large firm excess is similar and is shown as a negative value in the small firm excess field. All subcategories with small firm excess values above 2.5 percent are highlighted in green and subcategories with large firm excess values below -2.5 percent are highlighted in yellow. The top five excess values for small and large firms are also highlighted in boldface type.

For example, the 2,663 patents in biotechnology produced by small firms account for 27.3 percent of the category, which is well above the average value for patents in all categories. Thus, the category is highlighted in green and the small firm excess value of 20.9 percent is highlighted in bold. This suggests that, not surprisingly, the biotechnology industry has much small firm participation. Aerospace technology, has barriers to entry, such as the need for wind tunnels, large space factories, and teams of engineers that make it very difficult for small firms to enter and thrive. In this technology only 1.2 percent of all patents come from small firms, so it is highlighted in yellow and the large firm excess value is in bold type.

Note also that the last column contains a P value for a significance test using a chi-square test on the small and large firm distribution. Specifically, the question is whether the differences between small firm patenting and expected small firm patenting are attributable to random variance. In most cases the P value is less than 0.001, suggesting that the probability that these differences are random is less than 0.1 percent. In 61 of the 74 subcategories, small firms patent at a statistically different rate than would be expected if they patented in the same areas as large firms. This provides a robust demonstration that small firms tend to patent in different technologies than large firms.

Since Table II.5 is sorted so that the subcategories are contained in the main category hierarchy, it is difficult to spot the top subcategories of patenting by small and large firms. Therefore, a subset of Table II.5 showing the subcategories with the highest small firm patenting is shown in Table II.6. Notice that communications and telecommunications patenting is the most active area for small firms, with 4,316 patents from 1997 to 2007. Several columns have been taken from Table II.5 for Table II.6, and two columns have been added at the end.[13] The last column shows the share of all small firm patents in the subcategory. Thus the 4,316 small firm communications patents represent 13.4 percent, or nearly one in eight small firm patents. Communications is also an area of enormous interest for large firms, so although small firms produce many communications patents, the small firm excess is only 1 percent (i.e., 7.5 percent of the communications patents are produced by small firms).

The green and yellow shaded subcategories are carried over from Table II.5, so the set of green subcategories in Table II.6 represents technologies in which there is a higher than expected amount of small firm patenting compared with large firm patenting among the top subcategories of small firm patent activity. For example, in Table II.5, the subcategory related to foundations had a higher than expected level of small firm activity relative to large firm activity, but since it contains only 21 small firm patents, it does not make it into Table II.6. The main takeaway from Table II.6 is that small firms are very active in health-related fields. (Even the other organic compounds category, while technically a chemistry field, has patents related to pharmaceutical preparations.) Within these fields, small firms have a higher percentage of patents than expected compared with large firms.

Small firms are active in various information technology subcategories such as semiconductors, computer software, and computer hardware, but they have a negative value for small firm excess in every case, suggesting that small firms account for fewer than the 6.5 percent expected share of patents versus large firms in each subcategory.

Table II.7 is similar to Table II.6 except that Table II.6 showed the top subcategories of small firm patenting, while Table II.7 shows the top subcategories of large firm patenting. The top four subcategories in patent

[13] The statistical significance column has been removed since in every case where there is a notable amount of small firm or large firm excess, that excess is statistically significant.

Table II.6 Top Subcategories of Small Firm Patenting from Table II.5.

Subcategories	Number of Large Firm Patents	Number of Small Firm Patents	Large Firm Excess (Percent)	Small Firm Excess (Percent)	Share of all Large Firm Patents (Percent)	Share of all Small Firm Patents (Percent)
Communications/Mostly Telecom	54,040	4,316	-1.0	1.0	11.6	13.4
Diagnosis/Surgery/Medical Instruments	23,962	2,668	-3.7	3.7	5.1	8.3
Biotechnology	7,077	2,663	-21.0	**21.0**	1.5	8.2
Pharmaceuticals	11,068	2,107	-9.6	9.6	2.4	6.5
Other Organic Compounds	11,788	1,884	-7.4	7.4	2.5	5.8
Semiconductors/Solid-State Dev./Electronics	34,552	1,735	1.6	-1.6	7.4	5.4
Measuring and Testing	20,005	1,436	-0.3	0.3	4.3	4.4
Computer Hardware	37,626	1,389	2.8	-2.8	8.1	4.3
Information Storage	20,544	1,326	0.3	-0.3	4.4	4.1
Optics/Photography/Electrophotography	17,292	1,253	-0.4	0.4	3.7	3.9
Electrical Devices	21,594	1,082	1.6	-1.6	4.6	3.3
Power Systems	8,717	997	-3.9	3.9	1.9	3.1
Computer Software	26,948	887	3.2	-3.2	5.8	2.7
Computer Peripherals	11,992	646	1.3	-1.3	2.6	2.0
All Remaining Categories	158,448	7,915	1.7	-1.7	34.0	24.5
Total	465,653	32,304	0.0	0.0	100.0	100.0

Table II.7 Top Subcategories of Large Firm Patenting from Table II.5

Subcategories	Number of Large Firm Patents	Number of Small Firm Patents	Large Firm Excess (Percent)	Small Firm Excess (Percent)	Share of all Large Firm Patents (Percent)	Share of all Small Firm Patents (Percent)
Communications/Mostly Telecom	54,040	4,316	-1.0	1.0	11.6	13.4
Computer Hardware	37,626	1,389	2.8	-2.8	8.1	4.3
Semiconductors/Solid-State Dev./Electronics	34,552	1,735	1.6	-1.6	7.4	5.4
Computer Software	26,948	887	3.2	-3.2	5.8	2.7
Diagnosis/Surgery/Medical Instruments	23,962	2,668	-3.7	3.7	5.1	8.3
Electrical Devices	21,594	1,082	1.6	-1.6	4.6	3.3
Information Storage	20,544	1,326	0.3	-0.3	4.4	4.1
Measuring and Testing	20,005	1,436	-0.3	0.3	4.3	4.4
Optics/Photography/Electrophotography	17,292	1,253	-0.4	0.4	3.7	3.9
Other	12,097	521	2.2	-2.2	2.6	1.6
Computer Peripherals	11,992	646	1.3	-1.3	2.6	2.0
Other Organic Compounds	11,788	1,884	-7.4	7.4	2.5	5.8
Engines and Parts	11,345	209	**4.6**	-4.6	2.4	0.6
Pharmaceuticals	11,068	2,107	-9.6	9.6	2.4	6.5
Resins/Polymers/Rubber	10,496	381	2.9	-2.9	2.3	1.2
Hand Tools/Machine Tools	9,708	501	1.5	-1.5	2.1	1.6
Motor Vehicles and Parts	8,817	346	2.6	-2.6	1.9	1.1
Power Systems	8,717	997	-3.9	3.9	1.9	3.1
All Remaining Categories	113,062	8,620	-0.6	0.6	24.3	26.7
Total	465653	32304	0.0	0.0	100.0	100.0

activity for the large firms are all related to information technology. This compares with Table II.6, where four of the top five subcategories were in health-related subcategories.

D. Conclusion

This section described the methodology used to build the database of small and large firm patents, which is the key building block for the remainder of the research project. Some basic overview results from the database are described in detail. Highlights include the fact that small firms outperform large firms in performance metrics such as the citation index, and that small and large firms tend to patent in different technology categories. The latter result is statistically significant.

III. Share of Small Firms Compared with Previous Projects

A. Introduction

To review, this project is built upon a carefully constructed database containing all U.S. firms that obtained 15 or more patents in the five-year period from 2002 to 2006. This is similar to databases of patents through 2000 and 2002 that the authors built in previous projects for the SBA, referred to in this report as SBA1[14] and SBA2.[15] The current database consists of 1,293 small and large firms and more than 1 million patent records. In SBA1, 33 percent of the firms in the database had 500 or fewer employees. In SBA2, 41 percent of the firms in the database were small firms. Based on this trend the authors hypothesize that the share of small firms in the current project will exceed the 41 percent found in SBA2.

In testing this hypothesis the authors also explore some related facts, for example firm age for small and large firms and the reasons firms are dropped from and added to the database over time.

B. Discussion

The first hypothesis was that the number and percentage of small firms in the database would increase substantially beyond the 41 percent measured in the SBA2 project because the percentage of small firms increased from 33 percent in SBA1 to 41 percent in SBA2. Instead, the authors found that the share of small firms held steady at 40 percent.

This result may be explained by trends in small firm R&D spending as a share of all corporate R&D spending. Figure III.1 reports National Science Foundation figures for the share of industrial R&D spending accounted for by firms with fewer than 500 employees[16]. Beginning in about 1989 small firms accounted for a rising share of industrial R&D. There was a spike in 2001 and a leveling off thereafter. The figure also shows the time periods covered by the three studies. The R&D share growth was strong between the periods covered in SBA1 and SBA2, so the share of patenting firms increased. However, between SBA2 and this study, the small firm share of industrial R&D leveled off, as did the share of small firms in the patenting database.[17]

The leveling off of firms within the dataset may thus be explained by the corresponding leveling off of the share of R&D expenditures produced by small firms in recent years. However, given all of the advantages large firms have over small firms in getting into the database, it is interesting that 40 percent of the firms with 15+ patents in the last five years are small firms. It is therefore instructive to take a deeper look into some of the properties of the small firms in the study.

[14] Diana Hicks et al., *Small Serial Innovators: The Small Firm Contribution To Technical Change*, U.S. Small Business Administration, Office of Advocacy, Contract No. SBAHQ-01-C-0149, February 2003.
[15] Anthony Breitzman et al., *Small Firms and Technology: Acquisitions, Inventor Movement, and Technology Transfer*, U.S. Small Business Administration, Office of Advocacy, Contract No. SBAHQ-02-M-0491, January 2004.
[16] National Science Board, "Science and Engineering Indicators," National Science Foundation. Various volumes: 1986-2006.
[17] The correlation between R&D expenditures and patent activity is well known and has been studied extensively. See, for example, Igor Prodan, "Influence of Research and Development Expenditures on Number of Patent Applications: Selected Case Studies in OECD Countries and Central Europe, 1981-2001." *Applied Econometrics and International Development*, 5, 4, 205.

Figure III.1 Share of Industrial R&D Spending Accounted for by Firms with Fewer than 500 Employees

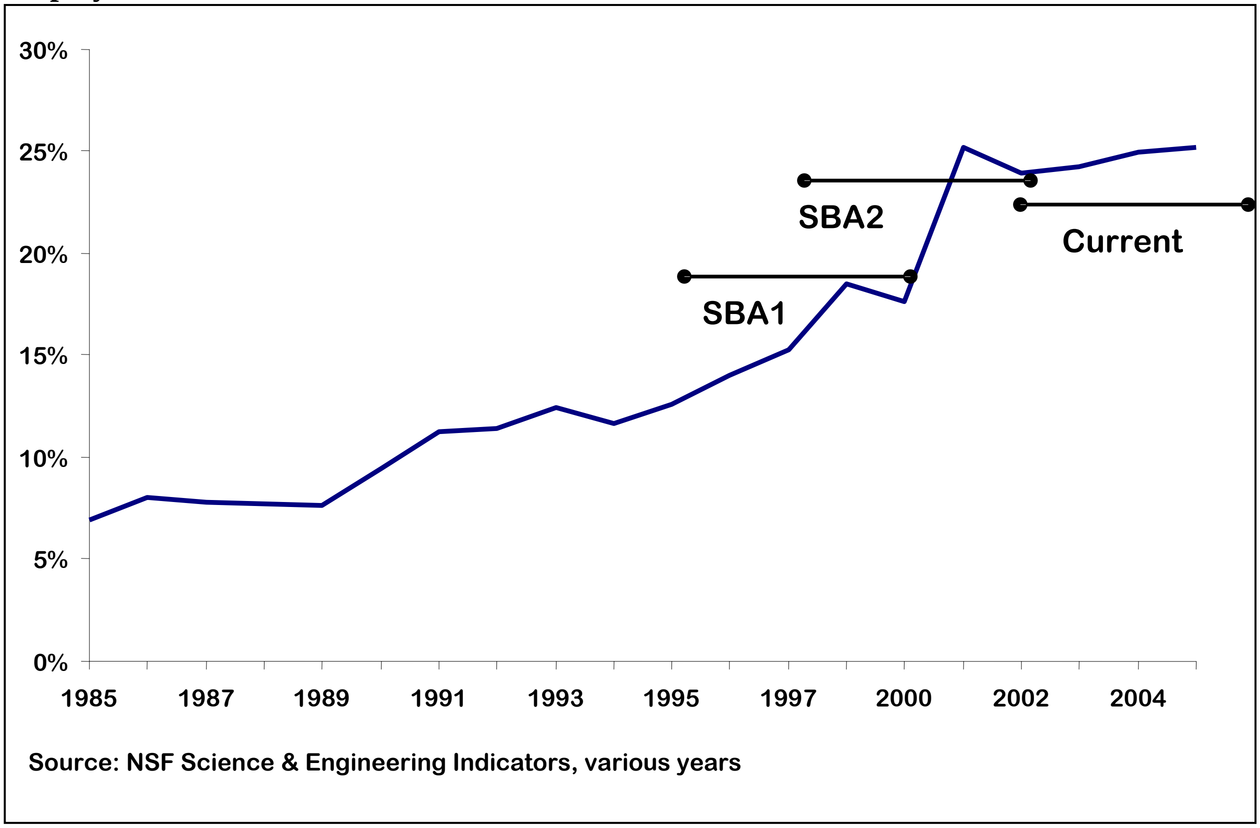
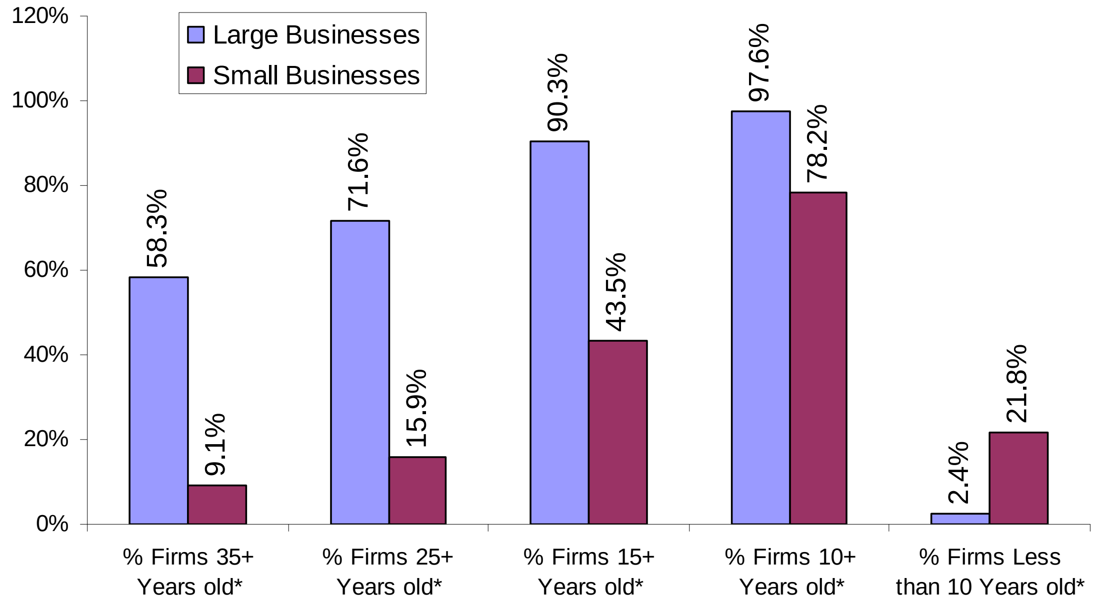


Figure III.2 shows that the small firms in the database are much younger, on average, than the large firms. In particular, one in five of the small firms did not exist 10 years previously while only one in 50 of the large firms are less than 10 years old. Similarly, only 43 percent of the small firms had existed for 15 or more years at the end of the period, while more than 90 percent of the large firms had been in existence for 15 or more years. A similar finding was observed in the SBA2 study, which led to the hypothesis above. In other words, given the trend of a large fraction of active patenting small firms entering the database, it can only be expected that the percentage of small firms in the set would increase. Thus, for the hypothesis to be wrong, there must be a large number of small companies dropping out of the database as well.

Before exploring this latter point it is important to note where the age statistics come from. It is fairly labor-intensive, and in some cases impossible, to identify the actual age of a firm. As a proxy for firm age, the authors used the age of the first patent application. This proxy was tested based on a lookup of the actual age of approximately 130 firms in the database. Of the 130 firms, 90 were founded after 1969 and these were used to test the patent application age proxy. Absolute differences were computed for the 90 firms between actual founding date and date of first application. The average difference was five years, but that was inflated by a few firms such as Stratus Technologies that have existed for 20 years, but just recently started patenting. In most cases the differences are smaller. In particular the median difference for the 90 firms is just two years. The results of Figure III.2 would thus be unlikely to change substantially if actual firm ages could be obtained.

Figure III.2: Percent of Small and Large Firms (by Age) with 15+ U.S. Patents 2002-06



*Age of firms is a difficult number to determine (especially for non-public companies). As a proxy we use the age of the first patent application date as a starting point. This is a reasonable proxy for patent intensive firms such as these.

Table III.1 shows the small and large firms in the newly created database that were also in the previous study. Of the 504 small firms in this study, only 198 (39 percent) were in the previous study. In other words 306 firms are new entrants in this study. The number of new entrant small firms outpaces the number and percentage of large firm new entrants. Again, this seems to support the hypothesis that there should be an increasing percentage of small firms in the study.

Recall from above that there were 516 small firms in the previous study and 504 in this study, including 306 new small entrants. Thus 300+ small firms in the previous study are either no longer small firms or no longer in the study. To be more precise, Table III.2 shows the fates of the 516 small firms in the previous study. Eighty-seven of the small firms have been acquired since that study, suggesting that these firms developed technology that was of interest to larger firms. A typical example is Stratagene, which was a small biotechnology firm with 460 people and $85 million in revenues whose technology caught the attention of Agilent. It was acquired by Agilent in August 2007 for $245 million.

In Table III.2, another 29 small firms have become large firms since the last study. Thus 116 or 22 percent of the small firms from the last study have graduated to become larger firms or have been acquired by larger firms. The question remains, what became of the 211 firms that dropped out of the database in this round? Researching 211 firms from a former study is well beyond the scope of this project. However, the results for a random sample of 30 of the firms that were in the previous study but were dropped in this study because they had too few patents to be included are summarized in Table III.3.

Table III.1 Overlaps between Firms in Last Study (SBA2) and Current Study

Firm Size	Firms in Database	Also in Previous Study (SBA2)	Percent in Previous	New Entrants	Percent New Entrants
Large	760	539	71	221	29
Small	504	198	39	306	61
Total	1264	737	58	527	42

Table III.2 Fate of Firms in Previous Study (SBA2)

Fate	Number of Firms	Percent of Firms
Acquired	87	17
In new Study—Large Firm	29	6
In new Study—Small Firm	189	37
Dropped for this study	211	41
Total	516	100

Table III.3 Fate of Random Sample of 30 Small Firms from Previous Study but Not in Current Study

Fate	Number of Firms	Percent of Firms
Small Firms Dipped below 15 Patents issued 2002-2006	22	73
Acquired	4	13
Troubled/Bankrupt	3	10
Now Large Firms but fewer than 15 Patents in 2002-2006	1	3
Total	30	100

Note that the majority of the firms (73 percent) that dropped out of the database are still operating, but they have fewer than 15 patents in the last five years. A typical example is Zircon Corporation, which makes levels and stud sensors. It had a flurry of patenting in the late 1990s related to short-range radar that could go through walls and determine stud locations. Now this established business files only a few patents per year to protect the business. Thus they are not in the current study because they do not pass the 15-patent threshold over the five-year period. A similar example is Second Chance Armor, a small firm that makes bulletproof vests. The key patents that established the business came in the late 1990s. They now file only a few per year to maintain their technological edge over competitors.

A slight dip in patenting like the Zircon example is not uncommon. Both large and small firms will vary their patent output over the years. Moreover, because of the backlog of pending patents within the Patent Office, sometimes a dip in patenting is beyond the control of the companies. Note that the 15-patent cutoff tends to favor large firms. Within the current database, 160 of the 504 small firms had 20 or fewer patents issued in 2002-2005, while only 68 of the 760 large firms in the database had 20 or fewer patents issued in the period. This suggests that small firms are much more likely to drop out of the database if their patenting falls by only a few patents over a five-year period.

Table III.4 Projected Fates of Small Firms from Previous Study but Not in Current Study

Fate	Number of Firms	Percent of Firms
Acquired[1]	115	22
In new Study—Large Firm	29	6
In new Study—Small Firm	189	37
Still Small but Dipped below 15 Patents 2002-2006[2]	155	30
Now Large but Dipped below 15 Patents 2002-2006[2]	7	1
Troubled/Bankrupt[2]	21	4
Total	516	100

[1] Projection based on 87 known acquisitions plus projection based on random sample of 30 small firms out of 211 that dipped below 15 patents 2002-2006.

[2] Projection based on random sample of 30 small firms out 211 that dipped below 15 patents 2002-2006.

A slightly different example is Cybex International, a small firm that made exercise equipment. This firm had a number of patents to establish its business in the late 1990s and now obtains only a few per year to maintain the business. Cybex is different from the previous examples because, although it has slowed its patenting, it has grown from a small business in the SBA2 study to a large business now.

Also in Table III.3, three of the 30 firms sampled are now troubled or bankrupt. While “bankrupt” is a well-known term, the term “troubled” is used here to describe a firm that is not technically bankrupt, but appears to be in trouble. For example one firm whose stock price has fallen from $80 per share to less than 10 cents per share is now down to the founder and just a few employees.

The projections of Table III.3 allow a better estimate of the fates of the 516 small firms in SBA2, shown in Table III.4. Although there were 306 new small firm entrants in this study, this number was offset by an estimated 155 firms that slowed their patenting and 151 firms that either became larger firms or were acquired by other firms. Also, an estimated 21 firms dropped out because of bankruptcy.

C. Conclusions

Small firms make up 40 percent of all U.S. companies with 15+ patents issued in the period 2002-2006. This is slightly below the 41 percent found in the previous study by the authors from 2004, but greater than the 33 percent found in the study from 2003. The leveling off of small firm share tracks a similar leveling off of the share of industrial R&D accounted for by small firms.

Small firms are at greater risk than large firms of falling short of every criterion for entry into the database. There is almost no risk that a large firm will become small, but 6 percent of small firms become large. The risk that a large firm will lose independence is presumably smaller than the small firms’ 22 percent risk of acquisition and 4 percent risk of becoming troubled. Episodic patenting is also characteristic of those with low patent output, and 32 percent of the small firms are very close to the 15 patent threshold, in that they have 20 or fewer patents (9 percent of the large firms do). Thus, small firms are far more likely to fall below the threshold. Given all of these factors it is impressive that 40 percent of the participating firms in this study are small firms.

The good news is that, despite the barriers to gain entry and stay in the database, 198 small firms in the current study were also in the previous study, and 306 small firms are new entrants. The 61 percent of small firms that are new entrants into the database is more than double the 29 percent of new-entrant large firms. Moreover, of the small firms that dropped out, all but about 4 percent are still around in some form. The majority of small firms that have dropped out have continued patenting at a slower rate, while others have been acquired or grown into larger firms.

IV. Small Firm Participation in Emerging Technologies

A. Introduction

Over the past five plus years, Breitzman has been working with the National Institute of Standards and Technology (NIST) and the Technology Administration (TA) in a series of research projects aimed at developing methods of identifying clusters of emerging technology.[18] In this section the authors leverage these studies and databases and match them to the SBA database created for this project in order to analyze small firm participation in emerging technologies.

B. Summary

Identifying emerging technologies is a notoriously difficult undertaking because it essentially involves predicting the future. Truly emerging technologies may not reveal themselves to be important for many years. However, as part of a project recently completed for NIST and the TA, the authors were able to identify a number of real-time parameters useful for marking patent clusters as likely to contain emerging technologies, as well as the top 100 high-scoring emerging clusters.

Some key findings are explored in detail:

- Most of the emerging clusters are dominated by either small firms or large firms, but rarely both within the same technology cluster.
- Small firms are much more likely to develop emerging technologies than are large firms. In particular about 1 in 31 small firm patents are found in the emerging technology clusters, while only one in 117 large firm patents are found in this special set.

C. Method

The emerging clusters methodology is based on patent citations, but it overcomes the lag time generally associated with citation analysis. The basic idea behind citation analysis is that patents referenced or cited by many later patents are referenced for a reason. That is, the later references are a direct result of a key idea being built upon by later patents. It has been shown that highly cited patents are correlated with a number of positive outcomes, such as increases in sales, profits, and stock prices, as well as in the licensing value of patents, renewal rates, and inventor awards.[19] As mentioned, the problem with citation analysis is that it takes several years for citations from later patents to accumulate, so that traditional citation measures are not appropriate for identifying emerging technologies.

---

[18] For an exhaustive review of the method and results of these research projects, see Anthony Breitzman and Patrick Thomas, *The Emerging Clusters Project Final Report*, U.S. Department of Commerce, Office of the Under Secretary: Technology Administration, October 12, 2007. Available at http://www.ntis.gov/pdf/Report-EmergingClusters.pdf . A summary version of this report can be found in the Summer 2008 ASTRA Brief, pages 10-15, available at http://www.usinnovation.org/files/ASTRABriefsSummer08.pdf . [Accessed August 11, 2008]. The project built upon previous work for the NIST Advanced Technology Program which is summarized in Anthony Breitzman. *ATP Hot-Spots Project Briefing, Final Project Summary*: National Institute of Standards and Technology, Advanced Technology Program, Gaithersburg, MD, February 25, 2004.

[19] For a review of validation studies related to citation analysis see Anthony Breitzman and Mary Mogee, "The Many Applications of Patent Analysis," *Journal of Information Science*, 28(3), 187-205, 2002.

Instead, the emerging clusters method identifies a subset of recent patents called “next-generation” patents that build upon highly cited, high-impact patents called “hot patents.” The next-generation patents are then clustered by like technologies and scored based on a number of metrics such as science intensity, public institution links, and patent originality, to identify the patents most likely to represent emerging technologies. These latter metrics differ from citation metrics in that they can be applied to patents in real time; that is, when the patents are published, rather than several years later.

The method has been developed and validated over numerous projects in the last five years. Two key validations are worth mentioning. A set of known high-risk, emerging patents related to NIST’s Advanced Technology Program (ATP) were used as a control set. These are patents developed as part of ATP projects where the applicant technologies were picked as being high-risk emerging technologies by panels of technology experts. In one validation study, the ATP patents were twice as likely to be found among the emerging patent clusters as in a similar-sized sample of the general population of patents. Moreover, even in cases where ATP patents were not found, patents covering technologies similar to the ATP patents were found in the emerging clusters. Hence, emerging technologies related to ATP funding were more than twice as likely to be found among the emerging clusters as patents in the general population. The second key finding is that emerging technologies like those found in the ATP patents have characteristics that can be identified with real-time indicators, thus allowing a scoring method to identify early-stage, emerging technologies at the time the patents are published. Again, this is significant because other well-known patent indicators such as citation measures generally need a significant lag before they can be applied.

One major output of the research that will be leveraged in this project is a set of the top 100 emerging clusters of patents with issue dates between January 2005 and August 2006 that were identified in the most recent emerging clusters project. Specifically these are clusters of seven to 62 patents (the average cluster has 18 patents) in related technologies that scored highest using the emerging scoring method.[20]

The distinction between emerging technology clusters and emerging industries should be pointed out at the outset. This section discusses clusters of technology that might be quite narrow such as “digital watermarking” or “efficient waveguide coupling.”[21] These technologies might be a small part of a company’s business but significant technology breakthroughs. Emerging industries are discussed in a later chapter. These represent the focus or products of an entire company. For example waveguide coupling is a technology within the telecommunications industry. Large companies such as IBM or General Motors may not be in emerging industries, but may well take part in developing emerging technologies.

D. Results

The main hypothesis to be tested here is whether small and large firms differ in the types of emerging technologies they pursue. Before that though, it is worth investigating small firm participation in emerging technologies.

[20] The interested reader can find a detailed methodology on the hot-patent, next-generation, and clustering method in Appendix A. Details on the scoring method and validation can be found in Appendix B. These appendices were modified from the most recent Technology Administration report from the emerging clusters project. Breitzman, op cit 15.

[21] Digital watermarking is a way of encoding information within a movie or audio so that it will not degrade the movie or audio but so that it will be traceable. This is one way to deal with piracy issues for DVDs or streaming audio and video. Waveguide coupling is related to joining fiber-optic coupling. Splicing, splitting, and combining fiber-optic networks in a low-cost way without signal degradation is a huge problem.

Table IV.1 lists some basic statistics about how many small and large firm patents are among the top 100 emerging clusters. As a comparison, the table also contains the absolute numbers of U.S. patents granted to small and large firms in the same period. Not surprisingly large firms dominate the overall list, with 76,586 patents issued, of a total of 83,128. That is, large firms account for 92.1 percent of patents issued and small firms account for 7.9 percent. For the patents contained in the top 100 emerging clusters, the story is quite different. This very select set of patents represents only about 1 percent of all issued patents and has characteristics that make them likely to contain important technologies in the future. Small firms account for 24.5 percent of the 868 patents in this select set of the top 100 emerging clusters. In other words, small firms have more than three times as many patents in the emerging clusters as they would be expected to have, given their overall participation in the patent system.

Put another way, about one in 31 small firm patents and about one in 117 large firm patents are in the top emerging clusters. Small firms, even with fewer resources and lower expenditures on R&D than large firms, are better at developing emerging technologies. This finding supports the theories of Schumpeter's early work, in which he popularized the idea of "creative destruction" and suggested that small firms are the primary source of technological change.[22]

In Section II of this report, and in SBA1[23] and SBA2,[24] it has been shown that small firms tend to specialize in certain industries where barriers to entry are small. For example, Section II.D discussed the fact that small firms have relatively few patents in aerospace, where wind tunnels, large spaces, and teams of engineers are required. Small firms have a large number of patents in medical devices, biotechnology, and other health-related fields where individuals or small teams can make a contribution. Based on these observations the authors hypothesized that small and large firms will differ in the emerging technologies they pursue. The set of the top 100 emerging clusters of patents is used to explore this hypothesis.

Recall that the emerging clusters were identified as part of projects for NIST and the Technology Administration. In those projects patents were not limited to those of U.S. firms, so the clusters contain patents from U.S. universities, government labs, nonprofits, and foreign corporations and universities. Although the U.S. small and large firms in this project participate in 94 of the top 100 emerging clusters, this study will concentrate on the 45 clusters where U.S. firms have a majority of the patents. This is done to avoid basing conclusions on clusters where there is a significant foreign firm participation, as size information is not readily available for these firms.

If the hypothesis is correct, some clusters would be expected to be full of patents owned by large firms and others to be full of patents owned by small firms, but no (or very few) clusters should have patents from both small and large firms. Table IV.2 distributes the emerging clusters based on large versus small firm ownership of the patents. For example, three clusters have 100 percent of their patents owned by small firms and 27 clusters have 100 percent of their patents owned by large firms. That is, 30 out of 45 (67 percent)

[22] Joseph Schumpeter (1912): *Theorie der Wirtschaftlichen Entwicklung*, Leipzing, Duncker & Humboldt, English translation: *The Theory of Economic Development*, Harvard 1934. Also Joseph Schumpeter (1942): Capitalism, Socialism and Democracy, Harper and Brothers, New York.
[23] Diana Hicks et al., *Small Serial Innovators: The Small Firm Contribution to Technical Change,* U.S. Small Business Administration, Office of Advocacy, Contract No. SBAHQ-01-C-0149, February 2003.
[24] Anthony Breitzman et al., *Small Firms and Technology: Acquisitions, Inventor Movement, and Technology Transfer,* U.S. Small Business Administration, Office of Advocacy, Contract No. SBAHQ-02-M-0491, January 2004.

Table IV.1 Number of Patents in Top 100 Emerging Clusters Divided by Expected Patents in Same

Firm Size	Number of U.S. Patents Issued Jan-05-Aug-06	Percent Patents Issued Jan-05-Aug-06	Number of Patents in Top 100 Emerging Clusters	Percent Patents in Top 100 Emerging Clusters	Expected Patents in Top 100 Emerging Clusters	Actual Patents in Top 100 Emerging Clusters / Expected Patents
Large	76586	92.1	655	75.5	800	0.82
Small	6542	7.9	213	24.5	68	3.12
Total	83128	100.0	868	100.0	868	1.00

Table IV.2 Distribution of Clusters by Large V. Small Ownership of Patents

Percent of Patents Owned by Small and Large Firms	**Number of Clusters**	**Percent of Clusters**
100% v 0%	3	6.7
90-99% v 1-10%	2	4.4
80-89% v 11-20%	1	2.2
70-79% v 21-30%	1	2.2
60-69% v 31-40%	1	2.2
50-59% v 41-50%	2	4.4
40-49% v 51-60%	2	4.4
30-39% v 61-70%	1	2.2
20-29% v 71-80%	2	4.4
10-19% v 81-90%	0	0.0
1-9% v 91-99%	3	6.7
0% v 100%	27	60.0
All	45	100.0

of the clusters have no overlap between small and large firms. If the threshold is 90 percent, five clusters have 90 percent or more of their patents owned by small firms and 30 clusters have 90 percent of their patents owned by large firms (see yellow shaded cells in Table IV.2). In other words, 35 of 45 clusters (78 percent) are dominated by small or large firms.

Most clusters are thus dominated by either small or large firms with little overlap. To show that this is a nonrandom occurrence, a simulation in which 45 clusters of varying sizes between five and 32 patents was set up and populated with random patents, where 76 percent would be from large firms and 24 percent from small firms. These are the same parameters as in the set of emerging clusters. The authors ran the simulation 10,000 times and never got close to having 35 clusters where 90 percent of the members were from small or large firms. In four of 10,000 cases there were 13 clusters with 90 percent of their members coming from small or large firms, but there were no cases with more than 13. In every one of the 10,000 simulations, most of the clusters had a mixture of small and large firm patents. This suggests that there is something nonrandom in the clusters leading to clusters containing small firm patents or large firm patents but rarely a mixture of both. This supports but does not prove the hypothesis.

Mixed Clusters

The next step in the analysis is to examine the emerging clusters where there is overlap among small and large firms. The first cluster has 10 patents—five from a small firm named Spraycool, Inc., three from Raytheon; and two from IBM. The patents became clustered together because they all reference similar prior art, and all are related to cooling electronic components with liquid. The clustering of large and small firm patents would seem to contradict the hypothesis, but it is necessary to look deeper into the patented technology and the companies behind it. The IBM patent is related to keeping electronic components cool within a computer. This method was used in supercomputers in the 1970s and is being looked at again as a means to keep laptops and cellphones cool as they become smaller and as fans become problematic because of noise and size constraints. The Raytheon patents are related to a cooling system for an antenna array. In both of these cases the large firms are solving a product-related problem with a potentially interesting technology, and these patents make up a very small part of the companies' patenting strategies and product development.

Spraycool, on the other hand, has its entire business built around its cooling patents. The specific technology Spraycool is developing is slightly different from the IBM and Raytheon patents. Spraycool develops enclosures for maintaining a constant temperature in harsh environments, on land, underwater, at high altitudes, etc. In many cases this means keeping electronics cool, but it could also mean keeping components warm in cold environments. Spraycool essentially works with military contractors and other companies to develop custom enclosures for use in airplanes and boats as well as in combat conditions on land. It sells most of its enclosures to the military, but has recently started developing enclosures for the computer industry and data centers.

Hence, to the extent that both large and small firms are developing similar technologies, this cluster does contradict the hypothesis. However, Spraycool is actually developing a technology from which it is launching an industry, whereas IBM and Raytheon are solving a problem to support their historical industries—so it is not so clear that this cluster violates the hypothesis.

A second cluster where both small and large firms have patents is an 11-patent cluster with six large-firm patents and five small-firm patents. This cluster has five patents from the small firm Roy-G-Biv Corp, five patents from Cray Computer Corporation and one patent from IBM. In this case the small firm patented technology is completely different from the large firm patents. The small firm Roy-G-Biv Corporation has a number of patents related to software for motion control. Robotics companies for the most part develop machines such as robotic arms that can be used in industrial settings. These machines have very simple low-level languages for controlling movement. Roy-G-Biv specializes in software systems that allow for combining robotic machines and harmonizing the simple command languages so that a set of machines can be controlled as a system such as an assembly line.

The Cray and IBM patents, on the other hand, are related to synchronizing programs in a multi-threaded computer environment. In this case, it is clear that the small and large firms in this cluster are working on different technologies, so the question is why they are in the same emerging cluster. The answer is that all of the patents build upon the same prior art. This prior art was related to an operating system that could simultaneously deal with diverse computer languages all running at once. This is the key idea upon which the Roy-G-Biv industrial control patents are built. The way the prior art handled dealing with different computer languages coexisting simultaneously was with the notion of maintaining multiple "stack-frames."

This idea led to simultaneous threads running on a single computer. So this cluster has distinct technologies from small and large firms built on two aspects of a key prior patent. This example supports the hypothesis that small and large firms will pursue different technologies.

The last mixed cluster relates to silicon photonics, which is related to making optical components using silicon semiconductor production techniques. Optical technology has become pervasive in the communication industry, since photons running through fiber optic cables at light speed allow for transmission over longer distances than electrons passing through copper wires. However, the materials used for optical communications tend to be expensive and difficult to produce. Silicon, which has long been the material of choice for the electronics industry, has been ignored by the photonics industry since it does not manipulate photons as well as materials such as indium phosphide or gallium arsenide.[25]

Silicon photonics is an emerging technology related to overcoming the shortcomings of silicon in optics to leverage all the advancements related to manufacturing silicon components. As Simon Sherrington, research analyst for *Light Reading Insider*, notes:

> The silicon photonics market is potentially enormous. Silicon photonics has the potential to be cheaper than the electronics-based solutions for micro-, short-, and medium-range data transmission. It promises to cost less in terms of materials, components, and ongoing power consumption. At the same time, silicon photonics offers a potential increase in communications bandwidth at cost-effective price points.[26]

The cluster of interest is related to silicon photonics and it contains patents from Intel and a small firm named Lightwire (formerly known as SiOptical). This is the strongest example of a direct contradiction of the hypothesis. A detailed review of the technology reveals that large firms such as IBM, Intel, Lucent, and others are pursuing the technology alongside small firms such at Lightwire and Luxtera. Interestingly, these two small firms have roughly the same number of patents overall as the giants Intel and IBM.

Lightwire claims to be the first firm to produce a commercially available standards-compliant silicon-photonics product to market.[27] The larger firms seem to have been dabbling in the technology longer, with patents that go back farther than those of Lightwire and Luxtera. It will be interesting to see if the small firms end up being purchased by the large firms.

So here is a technology in which both small and large firms are competing and it has not yet been decided who will win in the marketplace. While this seems to be a clear contradiction of the hypothesis that small and large firms pursue different technologies, it is not really a typical case.

A Small Firm Only Example

The examples above looked in detail at the few cases in which patents of small firms were mixed with patents of large firms within the same emerging technology cluster, and offered support for and against the

[25] Kate Green, "A Record Breaking Optical Chip," *MIT Technology Review*, June 25, 2008.
[26] Press Release: "Silicon Photonics Fast Becoming a Commercial Reality, Report Finds," PR Newswire, August 13, 2008. Accessed from: http://ubmtechnology.mediaroom.com/index.php?s=43&item=2094 [Last accessed August 25, 2008].
[27] Press Release: "Lightwire Introduces First CMOS Photonics 10GbE SFP+ LRM Optical Module," Lightwire Inc. February 25, 2008. Accessed from: http://www.lightwire.com/press_release.cfm?RecordID=3 [Accessed August 25, 2008].

hypothesis that small and large firms would pursue different emerging technologies. In most clusters where there is either small or large firm participation, the case for the hypothesis should be much clearer. An interesting example that shows that things are not simple even in these clusters is the case of digital watermarking. This cluster consists of 32 patents, all of which are assigned to the small firm Digimarc. Digimarc's technology is related to digital watermarking and can be used to create secure drivers' licenses with anti-counterfeiting measures. It can also secure electronic media with "watermarks."

In 2005, for the first time, Digimarc's technology was embedded in DVDs sent to Academy Award screeners, in an effort to reduce piracy. This was in response to a 2004 incident in which an Academy of Motion Picture Arts and Sciences member was ordered to repay a movie studio $600,000 after it was revealed he shared his screener copies of movies such as *The Last Samurai* with a known pirate.[28] Since the technology can also be used to filter out content and prevent it from being downloaded, Digimarc has been asked by the Recording Industry Association of America to see how digital watermarking can be used to curb illegal distribution of music on peer-to-peer networks.[29]

However it is not the only company, or even the oldest company working in the space. Large firms like IBM, Hitachi, Sony, and Xerox, among others, have dabbled in the technology for years. Digimarc seems to have the most patents and probably the leading technology, but it faces competition from large firms.

E. Conclusion

This section studied a specialized set of emerging technologies identified in a previous project for NIST and found that small firms tend to pursue differing emerging technologies from those of large firms and that small firms have much greater participation levels in emerging technologies than would be expected.

[28] Lisa DiCarlo, "Digimarc's Growing Pains" Forbes.com, February 14, 2005. http://www.forbes.com/2005/02/14/cx_ld_0214digimarc.html [Accessed August 25, 2008].
[29] Ibid.

V. Identification of Emerging Industries by Classifying Firms

A. Introduction

In this analysis, emerging industries are identified using a custom industry classification scheme focused on a set of innovative firms. This custom classification is then compared against the general purpose NAICS classification scheme. The custom scheme is attuned to principles of innovation theory, while NAICS is production process-oriented. The discrepancies between the two schemes, as well as a heightened presence of small firms in a category, suggest where technology-based industries may be emerging.

Some key results are discussed in more detail. Some traditional industries figure prominently in these data. Unexpected innovators are found in: batteries, gaming machines, packaging, and retail display (showcase) manufacture. These industries are not emerging, but they may be undergoing a technology-driven renaissance. In some cases, small firms may be able to play the role of disruptive innovators and enter such established industries. Two types of emerging industries are identified. The first involves commercializing new technology. The second, while innovative, is emerging because of innovations in business models. Within both types are industries in which small firms predominate.

B. Method

For the database used in this project, NAICS and SIC codes were obtained for each firm where available. The NAICS codes come from Mergent⁄Moody's International for most of the public firms and Lexis⁄Nexis for the smaller firms.[30] Dun & Bradstreet reports provided SICs for firms not in the other two databases. Codes were found in these databases for all but 39 firms, or 97 percent of the 1,293 firms. SIC codes were converted to NAICS where possible. The SIC to NAICS mapping is not always unique, and the authors chose to convert only when the SIC to NAICS relationship was unique. Converting SIC to NAICS at the three-digit level in cases where this relationship was unique increased the number of firms. The final result was database-assigned NAICS codes for 1,150 firms, or 89 percent of the total.[31]

A qualitative method was used to classify the firms. In qualitative research, codes are attached to phrases, sentences, or paragraphs in text (usually interview transcripts). The codes represent the concept being discussed in the text segment. In this project, the line of business of each firm was coded from text describing each firm. The corpus of text was obtained from Google Finance where available.[32] For firms not covered in Google Finance, text was obtained from firm websites or as a last resort from other websites. No text was found for 3 percent of the firms (see Table V.1).

In qualitative research, ideally several people code the same texts and the reliability of the coding is checked by assessing inter-coder agreement. In this project, Hicks essentially checked the reliability of the database coding, working from assigned NAICS codes and assigning each firm a code in agreement or disagreement with the database coding. It is in the coding disagreements combined with the distribution of small versus large firms that the results are found.

[30] The codes appear to be NAICS 2002, not the latest NAICS 2007.
[31] Codes exist for 98 percent of the public firms and 75 percent of the private firms as well as 95 percent of the large firms and 84 percent of the small firms.
[32] Google Finance seems to get its information from Hoovers.

Table V.1 Source of Text Used in Corpus

Text Source	Firms	Percent	
Google Finance	997	77	94 percent of the large firms and 55 percent of the small firms 98 percent of the public firms and 46 percent of the private firms.
Website	219	17	
Other	42	3	
NA	35	3	
Total	1293	100	

C. About NAICS

The North American Industry Classification System (NAICS) is:

> An industry classification system used by statistical agencies to facilitate the collection, tabulation, presentation, and analysis of data relating to establishments. NAICS is erected on a production-oriented conceptual framework that groups establishments into industries according to similarity in the process used to produce goods or services. Under NAICS, an establishment is classified to one industry based on its primary activity. NAICS was developed jointly by Canada, Mexico, and the United States to provide comparability in economic statistics. It replaced the Standard Industrial Classification (SIC) system in 1997.[33]

In this project, NAICS is taken to be the canonical firm classification scheme. The scheme is very well founded, as a great deal of expertise was marshaled and effort was expended in its design and updating. The authors argue that NAICS is not used here as a straw man, but as a hard case against which to compare the authors' classification. In designing NAICS and in subsequent revisions, special attention was paid to developing production-oriented classifications for (a) new and emerging industries, (b) service industries in general, and (c) industries engaged in the production of advanced technologies. Because NAICS strives to do precisely what the authors aim to do here, one might expect all innovative firms to be easily classified into NAICS industries.

However, to qualify for a NAICS code, an industry must:

- Be consistent with classification principles of mutual exclusivity, exhaustiveness, and homogeneity of units within classes.
- Have empirical significance; that is, classes should produce gross revenues of $500 million (Canadian), be collectable, and be linked to a funded program for data collection.
- Be relevant; that is, they must be of analytical interest, result in data useful to users and be based on appropriate statistical research and subject matter expertise.[34]

While NAICS strives to keep up with emerging industries, and industries based on advanced technology, achieving those goals is difficult when the review cycle lasts five years and when industries must be of a

[33] U.S Census Bureau, 2002 Economic Census, Glossary of Terms for the Economic Census, http://bhs.econ.census.gov/econhelp/glossary/, accessed July 22, 2008.
[34] Statistics Canada, Revision of the North American Industrial Classification System, http://www.statcan.ca/english/concepts/naics_consultation.htm, accessed July 22, 2008.

certain size before they are assigned a code. The search here for emerging industries takes advantage of these characteristics to look for firms that do not fit well in the current NAICS. Because NAICS seeks to include emerging industries, well known emerging industries (search engines, for example) are included. Therefore the authors' conclusions will not be obvious. However, because NAICS by its nature cannot be as nimble as the authors can, there will be exclusions in NAICS. These are the source of the results here.

There are differences between the design principles of NAICS and the design principles here, which use a very generous lower limit on the size of an industry to which a code is assigned. Two firms were the minimum necessary to assign a code in the scheme. Second, while NAICS judges industry size based on revenue, here it is judged on patenting. Third, NAICS is designed to classify establishments. According to the Census Bureau, an establishment is defined as "a single physical location where business is conducted or where services or industrial operations are performed," and "a company may have one or many establishments." This project classifies enterprises or firms. It also adds a code for "diversified manufacturing" for firms whose businesses were too heterogeneous for a single NAICS code. This category is mainly composed of large firms and is not analytically interesting.

In principle, NAICS defines industries by the similarity in processes used to produce goods or services. This means that there is one category for windows with wooden frames and another for windows with metal frames. This principle seemed to clash with the business philosophies employed by a number of firms. In the past, firms may have specialized in working wood, plastic or metal, and no doubt a number still do. However, many firms in this set seem to have moved to providing "solutions"—a fuzzy word that often seems to mean bringing together whatever is necessary to solve the customer's problems.

For example, plastic manufacturing is nicely dissected in NAICS. If a firm manufactured the noodles used as packing material, it would be easily classified as a plastic foam producer. However, one of the unexpected areas of patenting activity seen in this database is in packaging. A typical packaging firm in this dataset might have branched out from being a noodle maker to become a "packaging solutions" provider, perhaps through acquisition. The firm would now produce noodles, other types of plastic, including sheet plastic, and also kraft-paper-based packaging.[35] This has advantages for customers such as Amazon, in that they do not have to put together the noodles with the bubble wrap with the kraft paper. The packaging firm can come in and advise on the best use of different types of packing material and develop a system for Amazon. NAICS has paper packaging, plastic packaging, etc., but no packaging industry.

The alternative to a process-based classification would be a product-based classification. An effort has been under way for some time to devise such a scheme, but it is not near completion. In the meantime, NAICS admits to being not entirely consistent with the process-based principle. For example, in many information-based industries, work consists of sitting in front of a computer screen all day—companies have similar requirements for capital and labor. Yet banks are classified into many categories, as are insurance agents and investment advisors. Seemingly, the classification relates to the type of service or product they provide. The noodle example suggests how the process-based principle at the firm level may be becoming outmoded, even in manufacturing, because of innovations in business models. Note that while the NAICS is well founded, the codes obtained from databases are less so, since employees of the database companies assign the codes used in this study.[36]

[35] Presumably the noodles and kraft paper packaging are produced at different establishments.
[36] Census assigns codes to establishments on the basis of information submitted by firms, but these codes are not public. Therefore, there is no "official" code for a firm. In some cases the coding for this study disagrees with the available NAICS code.

Table V.2 Distribution of Firms across NAICS

NAICS two-digit	Description	Percent of establishments in economy	Percent of firms in patent database	Percent of sales, receipts or shipments in NAICS category	Percent of patents in NAICS category
	Source	Census	This study	Census	This study
21	Mining	0	**1**	1	**1**
22	Utilities	0	0	2	0
23	Construction	10	0	6	0
31-33	Manufacturing	5	**78**	18	**89**
42	Wholesale trade	6	2	22	1
44-45	Retail trade	16	1	14	0
48-49	Transportation and warehousing	3	0	2	0
51	Information	2	**4**	4	**5**
52	Finance and insurance	6	1	13	0
53	Real estate and rental and leasing	5	0	2	0
54	Professional, scientific, and technical services	11	11	4	2
55	Management of companies and enterprises	1	0	1	0
56	Administrative and support and waste management and remediation service	5	0	2	0
61	Educational services	1	0	0	0
62	Health care and social assistance	10	0	6	0
71	Arts, entertainment, and recreation	2	0	1	0
72	Accommodation and food services	8		2	
81	Other services (except public administration)	8	0	1	0
		100	100	100	100

The distribution of firms is uneven across NAICS categories, both in this dataset and in the economy as a whole (Table V.2). In this table, Census data on the distribution of establishments are compared with the authors' distribution of patenting firms, and Census data on sales are compared with their data on distribution of patents. Industries overrepresented in the patenting data compared with the economic data are highlighted in bold. Most industries are underweighted in the patent data compared with the economic data, except mining, information, and especially manufacturing.[37]

Patenting is focused in manufacturing. Table V.3 reports the same data for manufacturing at the three-digit level. Here the disparities between the economic and patenting data are not so extreme. Computer manufacturing is the most heavily overweighted sector in the patenting data.

This may be due to miscoding by employees of the database companies, data entry errors in this project, or disagreements with the NAICS scheme itself.

[37] A key question here is whether the patent data are indicative of the distribution of innovation in the economy, or simply indicates the propensity to patent. European innovation survey data suggest that propensity to patent correlates well with firm tendency to innovate in house (specifically, across industries, share of firms that patent is highly correlated with share of firms that report innovating in house within the last year). The two sectors whose innovativeness is relatively underrepresented by patent data are financial services and information services. H. Hollanders and A. Arundel, *European Sector Innovation Scoreboards: European Trend Chart on Innovation*, European Commission, December 2005.

Table V.3 Distribution of Firms across Manufacturing NAICS codes

NAICS two-digit	Description	Percent of Establishments	Percent of Firms	Percent of sales, receipts or shipments	Percent of patents
31-33	Manufacturing	100	100	100	100
311	Food	8	1	12	0
312	Beverage and tobacco product	1	0	3	0
313	Textile mills	1	1	1	0
314	Textile product mills	2	0	1	0
315	Apparel	4	0	1	0
316	Leather and allied product	0	0	0	0
321	Wood product	5	0	2	0
322	Paper	2	1	4	1
323	Printing and related support activities	11	1	2	0
324	Petroleum and coal products	1	1	5	0
325	Chemical	4	**16**	12	**13**
326	Plastics and rubber products	4	2	4	1
327	Nonmetallic mineral product	5	1	2	1
331	Primary metal	1	1	4	0
332	Fabricated metal product	18	4	6	1
333	Machinery	8	**12**	6	**8**
334	Computer and electronic product	5	**36**	9	**51**
335	Electrical equipment, appliance, and component	2	**5**	3	**6**
336	Transportation equipment	4	**7**	16	11
337	Furniture and related product	6	1	2	0
339	Miscellaneous	9	9	3	**5**

Reflecting the reality of these distributions, the custom classification scheme varies in its level of detail from two-digit to six-digit. Finer coding was used where more firms were located. Note also that innovating firms, as represented in patent data, also underweight small firms. According to Census data, 99.7 percent of all firms, and 99.6 percent of manufacturing firms, have fewer than 500 employees. Forty percent of the firms are small. This is because a smaller percentage of small and medium-sized enterprises (SMEs) than of large firms are active in innovation[38] and smaller firms have fewer resources to devote to obtaining patents.[39]

[38] B. Stockdale, *UK Innovation Survey 2001*, UK Department of Trade and Industry.

[39] It has been hypothesized that patenting is very expensive and that larger firms have more resources to obtain and defend patents. For example, Scherer, F.M. (1983), "The Propensity to Patent," *International Journal of Industrial Organization*, 1, 107-128, discusses the correlation between R&D expenditures and patenting and shows that R&D expenditures increase with size. More recently, Anna Norman. (2001), "The patenting process; for SMEs, does size really matter?" Derwent Information. [online]. Available from: http://scientific.thomsonreuters.com/free/ipmatters/bti/8199610/ [accessed Aug. 11, 2008] showed that less than 10 percent of firms with five or fewer employees engaged in patent activities, with a monotonic increase in such activities with the growth in number of employees. In particular, 50 percent of firms with 101-250 employees engaged in patent activities. The argument that small firms have fewer resources for patent activity is also bolstered by Mary Mogee, *Foreign Patenting Behavior of Small and Large Firms: An Update*, U.S. Small business Administration, Office of Advocacy, Contract No. SBAHQ-01-M-0357, April 2003. Mogee established that although the patents of small firms are as important as those of larger firms, the patented inventions are less likely to be filed worldwide, presumably because of the high costs associated with increased foreign patent protection.

D. Results

The custom classification contains 87 codes. Sixty-one of these of these (69 percent) are standard NAICS codes; 26 are custom. Fifteen of the codes are dominated by small patenting firms. The dominant firm type was determined by a T-test using two categories of firms: small and large, or having fewer and more than 500 employees. Firms were counted as small if no employee numbers were found for them. If the P value on the T-test was less than or equal to 0.10, the industry was counted as large or small firm dominant. This means that large or small firms were more numerous than expected in the industry, given the number of large and small firms in the database as a whole.

Combining the two dimensions—NAICS and custom codes, and small or large firm dominance (or neither)—creates six categories to examine for emerging industries. Four of the categories capture established industries: regular NAICS codes and custom codes dominated by large firms.[40] Two categories isolate potential emerging industries. These are custom codes with no large or small firm dominance or with small firm dominance. Table V.4 summarizes this with the interesting categories highlighted in bold.

1. Standard Industry Codes

The 25 NAICS codes dominated by large firms are listed in Table V.5. Industries such as mining and oil; food, beverage, and tobacco; chemicals; pharmaceuticals; semiconductor device manufacturing; aerospace; telecommunications; and software are found here. The technique has successfully identified and isolated well-established innovative industries. One note of caution is warranted concerning finance and insurance. It is likely that, compared with other industries, innovation in this industry is underrepresented by patenting. Here are only very large firms: American Express, Citigroup, Goldman Sachs, Lehman Brothers, Mastercard, and Visa. Perhaps only the largest firms are beginning to dabble in patenting here, and so the numbers may not indicate the relative innovativeness of large and small firms in this industry. Patenting in the financial and retail industries consists mainly of a new species of patents known as "business method patents," which came into prominence only after a court decision in 1998.[41]

Table V.6 lists the custom codes dominated by large firms. Here the diversified manufacturing code is needed because NAICS is designed for establishments, while firms are classified here. The "unique" code was added to prevent the addition of a dozen or so detailed codes with one or two large firms in them. Firms found here include several musical instrument makers and Amazon, Gemstar–TV Guide, Hallmark, Quest Diagnostics, Radioshack, Sharper Image, Smith & Wesson, Time Warner, UPS, Walt Disney, and Weyerhaeuser. The consumer products code was added to capture firms that design and produce a range of household products such as gas and electric grills (W.C. Bradley), personal care products (Homedics), and electric hair clippers (Wahl). By producing a variety of low-cost products sold through retail outlets whose manufacturing might be outsourced, the firms seemed to have more in common with each other than with, for example, other metal working firms (Bradley). The best NAICS code alternative to the custom category would be the more limited: *335211—Electric Housewares and Household Fan Manufacturing*.

---

[40] Large firm custom codes were necessary because the authors classified firms whereas NAICS is designed for the more focused establishments. Although these custom codes are not in NAICS, the reason for that is not interesting here.

[41] The allowance of patents on computer-implemented methods for doing business was challenged in the 1998 *State Street Bank v. Signature Financial Group, Inc.*, (47 USPQ 2d 1596 (CAFC 1998)). The court affirmed the position of the USPTO and rejected the theory that a "method of doing business" was excluded subject matter.

Table V.4 Types of Codes in Custom Classification Scheme

Dominant Firm Type	NAICS	Custom	Total
Large	25	4	29
None	31	**12**	43
Small	4	**11**	15
Total	60	27	87

Table V.5 NAICS Codes Dominated by Large Firms

Code	Industry	Firms	Small	Large	Chi-square	P value
21	Mining, Oil and Support	12	2	10	2.7	0.10
311	Food Manufacturing	11		11	7.3	0.01
312	Beverage and Tobacco Product Manufacturing	4		4	2.6	0.10
313	Textile Mills	5		5	3.3	0.07
322	Paper Manufacturing	4		4	2.6	0.10
324	Petroleum and Coal Products Manufacturing	6		6	4.0	0.05
3251	Basic Chemical Manufacturing	20	2	18	7.5	0.01
3254	Pharmaceutical and Medicine Manufacturing	12	2	10	2.7	0.10
3255	Paint, Coating, and Adhesive Manufacturing	7		7	4.6	0.03
3256	Soap, Cleaning Compound, and Toilet Preparation Manufacturing	12		12	8.0	0.01
3259	Other Chemical Product and Preparation Manufacturing	5		5	3.3	0.07
331	Primary Metal Manufacturing	5		5	3.3	0.07
332	Fabricated Metal Product Manufacturing	22	5	17	2.7	0.10
3333	Commercial and Service Industry Mach	19	4	15	2.8	0.09
3336	Engine, Turbine, and Power Transmission Equip	6		6	4.0	0.05
33411	Computer and electronic equip	13	2	11	3.3	0.07
334413	Semiconductor and Device Manufacturing	29	7	22	3.0	0.08
334513	Inst and Related for Measuring, Displaying, and Controlling Industrial Process Variables	8		8	5.3	0.02
336	Transportation Equipment Manufacturing	50	10	40	8.5	0.00
3364	Aerospace Product and Parts	12		12	8.0	0.01
337	Furniture and Related Product Manufacturing	9	1	8	3.1	0.08
33920	Sporting and athletic	16	2	14	5.0	0.03
511210	Software	31	6	25	5.5	0.02
517	Telecommunications	12	2	10	2.7	0.10
52	Finance and Insurance	7		7	4.6	0.03

Table V.6 Custom Codes Dominated by Large Firms

Industry	Firms	Small	Large	Chi-square	P value
Consumer Products	10	1	9	3.7	0.05
Diversified Manufacturing	52	3	49	26.2	0.00
Packaging	20	3	17	5.2	0.02
Unique	26	3	23		

Packaging appears as a custom code because of the fine distinctions made in NAICS in manufacturing. Manufacturing accounts for 5 percent of establishments and 18 percent of sales, but 36 percent of NAICS codes. Thus there are two categories for heavy and light gauge springs. Similarly in packaging there are categories for:

322214—Fiber Can, Tube, Drum, and Similar Products Manufacturing
322221—Coated and Laminated Packaging Paper Manufacturing
322225—Laminated Aluminum Foil Manufacturing for Flexible Packaging Uses
326112—Plastics Packaging Film and Sheet (including Laminated)
332431—Metal Can Manufacturing
333993—Packaging Machinery Manufacturing
339991—Gasket, Packing, and Sealing Device Manufacturing

Some of the firms in the custom packaging category here would fit NAICS codes because they make packaging machinery or glass bottles. However, more sophisticated companies engage in more complex production not accurately represented by the micro-fine NAICS categories. For example:

- Innovative protective packaging products and packaging systems. Products include packaging made on-site, bubble, air cushions, kraft-paper cushioning and ready-to-use products, polyethylene foam, loosefill made of 100 percent recycled polystyrene or cornstarch, and kraft/bubble mailers.

- Manufacture and supply of standard, specialty, and custom packaging for high-definition discs (Blu-Ray, HD DVD and combination discs), DVD, CD, VHS, software, PC and video console game formats, and complete custom solutions including 3D modeling, prototyping, design engineering, and production manufacturing.

- Dispensing systems for the personal care, fragrance/cosmetic, pharmaceutical, household, and food/beverage markets—focusing on providing value-added dispensing systems (pumps, closures, and aerosol valves) to global consumer product marketers.

- Packaging and performance-based materials and equipment systems that serve an array of food, industrial, medical, and consumer applications. Food packaging is driven by developments in technologies that enable food processors to package and ship fresh and processed meats and cheeses through their supply chain. The food solutions segment focuses on case-ready packaging, ready meals, and vertical pouch packaging. Protective packaging includes technologies and solutions aimed at traditional industrial applications while increasing emphasis on consumer-oriented packaging solutions.

From the perspective of innovation theory, such companies are united by an innovation model in which they innovate and custom engineer each product to meet the needs of their large firm customers. Therefore, they are classified into one category that does not favor small firms, so it is a “traditional industry undergoing change,” probably both in technology (indicated by the strong presence in patenting) and in business model (indicated by the misfit with NAICS categories).

Table V.7 lists the NAICS codes not dominated by either large or small firms. Many contain too few firms to assess reliably whether one or the other type of firm truly dominates. Again the technique used here has isolated established industries, such as building materials, plastic, web portals, many types of machinery manufacturing, etc. Note that emerging industries do not always require new NAICS codes. For example, in recent years new companies have emerged to sell GPS devices and services; these fit very well into the existing category: *334511—Search, Detection, Navigation, Guidance, Aeronautical, and Nautical System and Instrument Manufacturing*. The three small firms working on carbon nanotubes fit in category *335991—Carbon and Graphite Product Manufacturing*. These firms were also given the nanotechnology code.

Table V.7 NAICS Codes Not Dominated by Either Large or Small Firms

Code	Industry	Firms	Small	Large	Chi-square	P value
23	Building materials and supplies	15	4	11	1.1	0.30
323	Printing and Related Support Activities	3		3	2.0	0.16
3252	Resin, Synthetic Rubber, and Artificial Synthetic Fibers and Filaments Manufacturing	2	1	1	0.1	0.77
3253	Pesticide, Fertilizer, and Other Agricultural Chemical Manufacturing	2	1	1	0.1	0.77
325413	in vitro diagnostics	4	2	2	0.2	0.68
325414	biol (not diagnostic)	5	1	4	0.8	0.37
326	Plastics and Rubber Products Manufacturing	14	3	11	2.0	0.16
327	Nonmetallic Mineral Product Manufacturing	3	1	2	0.1	0.82
333	Machinery Manufacturing	7	3	4	0.0	0.87
3331	Ag, Const, and Mining Machine	28	9	19	0.7	0.41
333293	Printing Machine and Equip Manuf	7	1	6	1.9	0.17
333294	Food Prod Machine	3	2	1	0.9	0.34
333295	Semic Manuf machines, materials	48	15	33	1.5	0.22
3334	Ventilation, Heating, Air Conditioning, and Commercial Refrigeration Equi	3	1	2	0.1	0.82
3335	Metalworking Mach	7	4	3	0.9	0.35
334112	Computer Storage	16	6	10	0.0	0.90
3342	Communications Equip Manuf	37	14	23	0.1	0.81
334310	Audio and Video Equip Manuf	10	6	4	1.7	0.19
3344	Semic and Other Electronic Component	10	4	6	0.0	0.99
3345	Navigational, Measuring, Electromedical, and Control Inst Manuf	8	2	6	0.7	0.39
334510	Electromedical and Electrotherapeutic Apparatus Manuf	18	9	9	0.8	0.37
334511	Search, Detection, Nav Guidance, Aeronautical, and Nautical System and Inst	4	2	2	0.2	0.68
334515	Inst for Test Electricity and Electrical Signals	2	1	1	0.1	0.77
334516	Analytical Laboratory Instrument Manufacturing	32	16	16	1.4	0.23
335	Electrical Equipment, Appliance, and Component Manufacturing	34	12	22	0.3	0.59
33591	Battery Manuf	4	3	1	2.1	0.15
335991	Carbon and Graphite Product	5	3	2	0.9	0.35
3391	Medical and Surgical Equip and Supplies Manuf	72	35	37	2.5	0.11
33993	Doll, Toy, and Game	5	1	4	0.8	0.37
518	Data Processing, Hosting and Related Services	3	1	2	0.1	0.82
519130	Internet Publishing and Broadcasting and Web Search Portals	2		2	1.3	0.25
5415	Computer Systems Design and Related Services	5	1	4	0.8	0.37

Battery manufacturing is not a new industry, but it is a place of considerable interest at the moment, as there is strong demand for new battery technology. Three of the four patent-intensive firms here are small, suggesting that the advantages small firms have in trying new approaches to a technology are being pursued here because of the expected high payoff for success.

Table V.8 reports the final set of standard NAICS categories—those dominated by small firms. Three industries are identified here, one manufacturing and two information industries.

Agriculture. This category contains two seed companies.

Showcase, Partition, Shelving, and Locker Manufacturing. This industry contains three small firms. The surprising thing is that it appears on a list of firms with a large number of patents. This industry

Table V.8 NAICS Codes Dominated by Small Firms

Code	Industry	Firms	Small	Large	Chi-square	P value
11	Agriculture	2	2		3.0	0.08
337215	Showcase, Partition, Shelving, and Locker Manuf	3	3		4.6	0.03
5413	Arch, Eng, and Related Services	18	13	5	8.0	0.01
5417	Scientific R and D Services	20	18	2	21.4	0.00

would not be expected to be generating patentable innovation, and this was the case in the past. This industry, like battery manufacturing, points to the emergence of innovative opportunities favoring small, nimble firms in areas of the economy not normally considered innovative.

Professional, Scientific and Technical Services. These two segments differ in emphasizing either engineering or scientific research. These are classic areas in which small firms dominate. They are well established and have a NAICS code. What seems to be changing among these firms is a shift from rhetoric emphasizing engineering and scientific prowess to an emphasis on commercialization. Firms are trying to spin off firms to manufacture their innovations, following the trend among universities.

2. Categories Representing Potentially Emerging Industries

Table V.9 reports the first set of categories with potential to reveal emerging industries: custom codes not dominated by either large or small firms. Eleven candidate industries are identified here, eight manufacturing and three information industries.

1. **Fluid handling machines**: pumps, compressors, valves, hydraulics in filtration. NAICS contains two relevant categories: fluid power cylinder and actuator, and fluid power pump and motor. These are found under miscellaneous machine manufacture.[42] One or two firms included here would fit into existing NAICS categories, such as the maker of hydraulic valves for cars. However, most would not; they make fluid handling solutions, high-precision, ultra-pure fluid handling systems for harsh environments; lubrications systems; spray nozzles, etc.[43]
2. **Alternative energy.** This is a small category that was included because it is currently the subject of much venture capital investment and innovation. One firm is working on hydrogen gas delivery systems, the other on photovoltaics, using its technology also in nickel-metal hydride (NIMH) batteries.
3. **Defense contractor.** A custom code was added for a clearly unique firm category not covered in NAICS. There were not enough firms to be significant in a statistical test, but the category was obviously dominated by large firms.
4. **Filtration equipment.** NAICS contains codes for water and sewage treatment facilities, but nothing for the manufacture of equipment to treat water. Several of the high-tech firms in this study specialized in manufacturing filtration equipment.

[42] It is not uncommon to find categories featured here under miscellaneous in NAICS. This makes sense because industries discussed here are as yet too small to deserve a NAICS category of their own and so are lumped together by NAICS essentially under not elsewhere classified.
[43] The full list of firm descriptions is in Appendix C.

Table V.9 Custom Codes Not Dominated by Either Large or Small Firms

Industry	Firms	Small	Large	Chi-square	P value
31-33 – Manufacturing					
Fluid handling mach—pumps, compressors, valves, hydraulics in filtration	13	5	8	0.0	0.92
Alternative energy	2	1	1	0.1	0.77
Defense contractor	3		3	2.0	0.16
Filtration equipment	4	1	3	0.4	0.55
Gaming machines	7	3	4	0.0	0.87
RFID	5	2	3	0.0	0.99
Semiconductor assembly and testing outsourcing	5	1	4	0.8	0.37
Contract manufacturing, various	2		2	1.3	0.25
51 – Information					
Digital security	13	8	5	1.7	0.11
Electronic Design Automation (EDA)	14	7	7	0.6	0.43
Positioning info, services, devices	3	1	2	0.1	0.82

5. **Gaming machines.** Coin-operated gambling machines are included by NAICS under *339999—All Other Miscellaneous Manufacturing*, along with fly swatters, candles and bone novelties. This seems strange in a classification so detailed that spring manufacture has its own four-digit code and two six-digit codes, one for heavy gauge and one for light gauge. The databases most often classified these firms into *71—Arts, Entertainment, and Recreation*, which is for theatres, national parks, and casino services. This classification is inappropriate for firms that design, make, and sell machines. This strange and inconsistent treatment is remedied here. The data suggest that gaming machines are an area of some innovative attention at the moment, although this is a new industry.
6. **Radio frequency identification (RFID).** The six companies in this category can most broadly be described as producers of automated identification and data collection products or solutions. This includes wireless or radio frequency identification. This category is clearly a new technology-driven industry segment, one which shows potential for small firms.
7. **Semiconductor assembly and testing outsourcing.** These firms manufacture and test devices under contract for other large firms. This type of activity is not represented in NAICS. It is a newer type of industry, but its emergence reflects changing business models, rather than new technology or manufacturing processes. The issue here is that the firms provide a service, and service firms are separated from manufacturing firms at the two-digit level. However, the firms manufacture, and so should be classified in manufacturing. The firms present NAICS with a paradox. The industry no doubt has large requirements for capital and so seems inhospitable to small firms.
8. **Contract manufacturing, various.** This category presents like those of the previous one, although in a broader set of products.

51 Information

9. **Digital Security.** This category encompasses identity protection, biometrics, authentication, anti-piracy protection, digital rights management, etc. Large firms include Verisign and Symantec. The products can combine both software and hardware, for example, biometrics, making them difficult to classify in NAICS. NAICS is not particularly detailed on software, focusing primarily on the distinction between custom and published software and software reproduction. Digital security looks very much like an emerging industry, although it does not favor small firms. Perhaps this is because the segment has seen strong growth, and relatively young firms have grown large quickly.

10. **Electronic design automation (EDA).** EDA is a specialized type of software for the design of integrated circuits. The data reveal a specialized software segment whose firms are most often coded as semiconductor manufacturers by databases. The segment appears to offer opportunities for small players.
11. **Positioning information, services, devices.** While GPS device companies fit comfortably into the navigation instrument category, several firms seem to specialize in positioning services, software and/or devices, and selling to large companies. It was quite difficult to determine what they actually sell; "solutions" may be the best description.

Table V.10 reports the second set of categories with potential to reveal emerging industries: custom codes dominated by small firms. Ten candidate industries are identified here, five manufacturing and five information industries.

1. **Photonics, optical components.** This segment again leads to questions about why a classification scheme with two categories for heavy- and light-gauge spring manufacture conflates photonic and optical components with *334413—Semiconductor and Device Manufacturing*. In this dataset are 30 firms focused on photonics in a way distinct from electronic integrated circuit firms. The production processes of photonics firms often use different technologies than semiconductor firms, for example, involving mirrors or fiber optical cable. Photonics is identified here as an emerging industry, and one that provides opportunities for small firms.
2. **Imaging and display.** NAICS contains a category for AV equipment: *334310—Audio and Video Equipment Manufacturing*, defined to include manufacture of items such as: amplifiers, camcorders, DVD players, speakers, microphones, and TVs. The category appears here as well and includes mostly audio equipment manufacturers and a set top box manufacturer. The imaging and display category was created to capture the set of firms that define themselves as developing imaging technology. These firms create technology for: in-vivo imaging, ferroelectric liquid-crystal-on-silicon (FLCoS) microdisplays, optical storage and digital imaging, electronic ink, holographic imaging, digital projection, etc. Kodak and Polaroid are both classified here, as Kodak is now described as producing imaging technology products. This is a technology-based emerging industry, and one in which serial innovators play a large role.
3. **Nanotechnology.** Nanotechnology is probably the most discussed emerging industry at the moment, and it is represented in this version of the serial innovators dataset for the first time. However, the industry is unstable, and a number of these firms disappeared through acquisition or bankruptcy this year, after the cutoff for construction of this dataset.
4. **Power supplies.** Fuel cells, microturbines, power adapters; this category is probably too heterogeneous for NAICS use. The category points to the emergence of technology-based industries in alternative power sources—fuel cells and microturbines as well as power supplies for electronic equipment. The two categories are related, however, as fuel cells may find an early application powering electronics.
5. **Quirky Technology Seeking Market** (QTSM). Spherical semiconductors, holographic data storage, laser peening, supercritical fluid technology for surface modification, 3D printers for rapid prototyping, lasing in cholesteric liquid crystals, polycrystalline diamond compact (PDC)—the firms in this category were formed to commercialize somewhat unique technologies. The QTSM self-description typically states that the firm was "formed to commercialize technology X." Perhaps

Table V.10 Custom Codes Dominated by Small Firms

Industry	Firms	Small	Large	Chi-square	P value
31-33 – Manufacturing					
334413/2 photonics, optical components	31	21	10	10.4	0.00
Imaging and display	30	21	9	11.7	0.00
Nanotechnology	20	17	3	17.4	0.00
Power supplies	12	9	3	6.3	0.01
Quirky Technology Seeking Market (QTSM)	12	12		18.4	0.00
51 – Information					
Biomed pipeline – drug and device	113	91	22	86.2	0.00
Biotech pipeline—non biomedical	2	2		3.0	0.08
Communications technology design and market	5	4	1	3.4	0.07
Fabless semiconductor	66	42	24	16.6	0.00
Pure play licensing	5	5		7.6	0.01
Unclassified	29	20	9	10.6	0.00

"technology commercializers" might seem a better title, but there is a sense in which these firms are in the wilderness with respect to an industry, and therefore a market. There is a fine line here. For example, what distinguishes a QTSM and a nanotechnology firm? Simply more firms are being formed to commercialize nanotechnology and there is more investor interest in them, so such firms were not classified as QTSM. Many biomedical pipeline firms have the same self-description, but again, their market goal is more clear, many firms are being formed in the same way, investors are interested and so such firms were classified as biomedical pipeline, not QTSM. QTSM firms are distinguished from scientific R&D services, which embody a larger spread of expertise and market themselves as service providers.

NAICS 51: Information

6. **Biomedical pipeline: drug and device.** This category, the largest in the study, includes what are commonly known as biotechnology firms. However, it was noted here that these firms are often classified by databases into a manufacturing category when most of them are precommercialization. This means that they undertake research, clinical trials, for example, and may license technology, but they do not make anything (laboratory-scale facilities to make prototypes would not count as manufacturing here). Manufacturing versus not manufacturing is a fundamental (two-digit) distinction in NAICS and a fundamental distinction in innovation theory concerning how technology is commercialized. A firm that manufactures has different capital and skill requirements than one that does not; therefore the distinction is fundamental economically as well. Firms that do not manufacture are properly classified as information firms based on their production activities. A great deal of care was put into clarifying whether these firms manufactured or not. The assumption that these firms are destined to transition into manufacturing firms seems rash. More accurately, such firms can be tagged as an emerging and permanent category of information firms whose existence is possible because of increased acceptance of technology trading as a business model in recent years.
7. **Biotech pipeline: nonbiomedical.** This category was added to capture two small firms that look very much like pipeline firms, except their application goal is outside biomedicine—for example, in biofuels. This is a candidate for an emerging industry.
8. **Communications technology design and market.** This category is analogous to the pipeline firms, but in a different technology. The firms deal in communication equipment, wireless phones, and

such, but they design and license technology. This industry is a candidate for a new industry, less for its technology, although this is cutting edge, than for its business model, which the data suggest may favor small, nimble firms.

9. **Fabless semiconductor.** This is the third largest category in the study and again similar to pipeline firms in conception. These firms are classified as manufacturing firms by the database when they do not manufacture anything. They design integrated circuits (ICs) and license their designs or contract for their manufacture and market the ICs. Designing, licensing, and marketing are the hallmarks of innovative sectors of the information industry. Even more care was needed to identify these firms than the pipeline firms as this industry is particularly coy about admitting to a lack of manufacturing facilities. A few pioneering firms in this area proudly announce on their websites that they are fabless, but for the rest it was necessary to search the firm's SEC filings for admissions of the truth that all manufacturing is contracted out. The emergence of manufacturing-related information sectors points to a broader concept than technology trading, called "task trading." It is task trading that enables a small firm to engineer an innovative product, contract for manufacture, and control marketing and distribution. This presents a fundamental opportunity for small innovative firms to succeed and control their technology. Innovation theory needs to be reworked to accommodate this phenomenon. The shift in business models also challenges the NAICS scheme at a fundamental level. Note that the nonmanufacturing business model may apply to other firms here classified into manufacturing categories. Extreme care in identifying nonmanufacturers was applied only to the biomedical pipeline and fabless semiconductor firms. The shifting business models represented here probably affect a broader set of sectors, although less systematically. Much more work is required to understand the intersection of changing business models and innovation.
10. **Pure play licensing.** This category captures firms whose only business appears to be technology licensing, without R&D, service, consulting, or any other activities that mark pipeline or fabless firms. Research Corporation, which licenses university patents, and iBiquity Digital, which licenses HD radio technology, are classified here.
11. **Unclassified** – These firms could not be classified. Often their business seems unique. Sometimes they are firms for which the authors do not have the correct description. They are listed in Appendix D.

E. Conclusion

This section of the study created a custom classification of innovative firms to identify emerging industries. Among the firm classifications were candidates for emerging industries, as well as standard industries in which innovative opportunities appear to favor the nimbleness of innovative small firms.

VI. Closing Summary

This report consists of four somewhat self-contained sections, each of which examines a different aspect of small firm patenting. While the four sections can be read and considered independently, they are actually related along two common themes. First, each section examines an issue that is testable via a patent database of small firms constructed specifically for this project. The second major theme covered in multiple sections is the notion of “emerging.” The study tests hypotheses related to whether small firms are emerging overall, as well as the participation of small firms in emerging technologies and in emerging industries.

The key section that binds all others together is Section II, which provides an overview of the small business patent database. This section describes in detail how the database used in each of the remaining sections was built. The database consists of all firms with 15+ patents issued between the years 2002 and 2006. This database of 1,293 firms was further researched to identify the small firms with 500 or fewer employees and the large firms with more than 500 employees. In total, there were 504 small firms and 760 large firms, plus 29 firms for which no employee information could be identified (these latter firms are very likely to be additional small firms).

Also Section II explores a number of key aspects of the database, including the relationship between numbers of employees and patent rates. An unexpectedly high percentage of small firms in the database are public companies. (Thirty-eight percent of small firms in the database are public, while less than 0.1 percent of all small firms in the general population are publicly listed.) Also, small firm patents are found to perform better than large firm patents on a number of performance metrics.

Section III explores the hypothesis that the percentage of small firms would rise above the level of 41 percent, found by the authors in a previous study. The level of small firms in this study is 40 percent, about the same as the 41 percent found in 2004 (SBA2) but more than the 33 percent found in 2003 (SBA1). This leveling off of small firm share tracks a similar leveling off in the share of industrial R&D accounted for by small firms based on NSF statistics. Section III also explored the relative youth of small firms and the reasons for firms falling out of the database over time.

Section IV examines small firm participation in emerging technologies. Identifying emerging technologies is of course difficult, but this study leveraged a set of 100 emerging patent clusters developed in a previous project for NIST. The key result here is that small firms have three times as many patents in emerging technologies as would be expected based on the overall percentage of patents for which they are responsible. Specifically, small firms account for just 8 percent of all patents in the database, but 24 percent of the patents of U.S. firms in the emerging technology clusters. Another key finding of this section is that small firms tend to patent in different areas of emerging technology than large firms. This is not a surprising result, since Section II showed that small and large firms tend to patent in different technologies in general.

Section V examines the analog of the emerging technologies discussed in Section IV, exploring emerging industries and the role of small firms within them. A key part of this section involves a custom classification of firms that enhances the existing NAICS and SIC industrial classifications. The problem with existing classifications is that their fixed nature means that, as new industries start to emerge, the firms within those industries need to be forced into the existing classifications, even if they do not fit very well.

The key results of this section involve identifying emerging industries and then showing that small firms ae very active within them. Two types of emerging industries were identified. The first type involves commercializing new technologies, such as alternative energy, filtration equipment, and RFID. The other type involved innovations in business models, such as companies that are separating R&D from testing and manufacturing. Some examples are small firms involved in biomedical pipeline and fabless semiconductor processes. In the former, small firms are developing drugs or treatments but are pre-clinical trial and pre-commercialization. In many cases these firms generate the pipeline and then sell or license the technology to large firms with the means to bring the drugs to market. The fabless semiconductor example is also an interesting development, in that it allows small firms to compete in an area from which they would normally be shut out because of cost. Several firms have carved out a niche by designing semiconductor chips and then outsourcing the manufacture or outlicensing the design to other firms.

Appendix A – The Emerging Clusters Method

I. Hot Patent Methodology

A. Key Concepts

The following key concepts will help elucidate the process by which emerging clusters of technology are identified.

1. Citation Analysis

Patent citation analysis is based on the examination of citation links between different generations of patents. When an inventor applies for a patent, the inventor must demonstrate that the invention is novel, useful, and nonobvious to someone with expertise in the same technology. To achieve this, the inventor cites to earlier patents and papers as prior art, and explains how the new patent improves on the earlier inventions. The patent examiner may also add citations to earlier patents that limit the scope of the new invention.

The idea behind citation analysis is that patents cited by many later patents tend to contain important ideas upon which numerous later inventors have built. This does not mean that every important patent is highly cited, or that every highly cited patent is important. However, numerous validation studies have revealed the existence of a strong positive relationship between citations and technological importance.[44] For example, it has been shown that positive citation indicators have been correlated with higher sales, profits, and increasing stock prices. Other studies have found links between inventor awards and citations.[45]

2. Citation Classic

On average, a patent receives approximately three citations over its first five years, but a small minority receive many more than this during this period. Furthermore, small numbers of patents receive hundreds, or even thousands, of citations in their lifetime. The patents that have received many citations over a long lifetime and continue to be cited many years later are known as citation classics.

Some examples of citation classics include the first patents for reliable ink-jet printing, the ethernet, the LCD panel, the stent, programmable logic arrays, and airbags. These inventions were not only groundbreaking for their time, but also continue to be cited by new patents.

The Palmaz stent, Patent #4,733,665, issued in 1988, is a good example of a classic patent. This patent was highly cited almost immediately, and is one of the most highly cited patents in the whole patent system, with more than 1,000 citations. Moreover, the patent continues to receive citations as inventors continue to

[44] Anthony Breitzman and Mary Mogee, “The Many Applications of Patent Analysis,” *Journal of Information Science*, 28 (3) 2002, pp. 187–205.
[45] Ibid.

improve stent designs. In 2006, for example, the Palmaz stent patent received more than 100 citations, even though it was 18 years old at that point.

The authors deliberately removed classic patents from consideration in this project. While they were groundbreaking for their time, most of the patents building upon them are now likely to be making incremental improvements on existing technologies, rather than creating potentially high-risk emerging technologies.

3. Potential Classic

A potential classic is a recent patent that has received many more citations than expected, given its age. If it continues to receive many citations over the course of a number of years, it will become a citation classic.

4. Sleeper Patent

A sleeper patent is one that receives few citations for many years and then suddenly receives a large number of citations. For example, a patent for a material that is light and strong but expensive may garner a few citations until someone works out how to make it more affordably. Once that obstacle is overcome, the patent may receive many citations as inventors reference it for uses in a variety of applications (e.g., airplanes, cars, toys etc.).

A good example of a sleeper patent is Upjohn’s 1969 patent #3,461,461 for minoxidil. This compound was initially developed to treat hypertension, but it was later noticed that one of its side effects was hair growth. Although the patent was issued in 1969, the bulk of its citations came in the 1980s and 1990s, when inventors started developing hair loss treatments based on minoxidil (marketed under the name Rogaine).

5. Hot Patent

Hot patents are essentially the union of potential classics and sleeper patents. This definition is not precise, but underlines the thinking behind the idea of hot patents—i.e., to isolate the subset of patents whose impact on recently issued patents is particularly strong. These could be either older patents that were ignored for many years, but are now receiving many citations; or relatively recent patents that are receiving many citations from current patents.

In other words, a hot patent must not only be highly cited, but must also receive the majority of its citations from recently issued patents. Hot patents differ from citation “classics” like the Palmaz stent patent, a very important technology that continues to be cited. Much of the stent patent’s influence has been on older technologies, so it is not considered a hot patent.

An example of a hot patent is Patent #5,772,905 for nanoimprint lithography, issued in 1998 to the University of Minnesota. This patent has received an impressive total of 127 citations from later patents. More important, 94 of these 127 citations are from patents issued since 2005. The high number of recent citations suggests that this patent is having a strong impact on recently issued semiconductor patents.

The impact of this University of Minnesota patent reflects a recent trend in the semiconductor industry. A current problem in semiconductor development is the lack of a low-cost, mass-producible process to create patterns with ultra fine features in a thin film carried on a semiconductor substrate. To obtain more powerful processors, more circuits need to be etched onto the substrate, which demands ever-finer lines. The University of Minnesota patent promises sub-50 micron features using a low-cost process. Evidence of the promise of this technology is in the large number of citations the patent has received from recent patents issued to organizations such as Micron Technology, Hewlett Packard, MIT, Freescale Semiconductor, and many other high-tech companies and universities.

Hot patents for a given year are identified quantitatively using a formula that employs a sliding scale in the percentage of citations to the patent for that year. For example, the formula for identifying hot patents for the year 2006 is as follows:

$$Ri/Ci > (Yi - 1958))/84 \text{ and } Ri > 9$$

where

- Ri denotes the number of citations patent i has received from patents issued between January 2005 and August 2006
- Ci denotes the total number of citations patent i has received
- Yi denotes the publication year of patent i

This formula relies on a fairly simple concept. A hot patent may be any age, but a large percentage of its citations must come from patents issued in the last 20 months.[46] This percentage varies depending on the age of the patent. For example, a 1977 patent has had 30 years to accumulate citations. If 25 percent of those citations occurred in the last 18 months, it would be reasonable to say that the patent is having a lot of impact on currently issued patents. On the other hand, a 1997 patent has had only 10 years to accumulate citations, so it would have to have a much larger fraction (50 percent) of its citations in the last 18 months to be considered a hot patent.

The second part of the hot patent formula (Ri > 9) means that, for a patent to be considered a hot patent, it must have at least 10 citations from recently issued patents. This restriction removes infrequently cited patents that happen to have been cited only by recently issued patents. Without this restriction, a patent cited only once, for example, could be considered a hot patent if the citation came from a recently issued patent.

6. Next Generation Patent

As outlined above, a patent is considered a hot patent when the majority of its citations are from patents issued in the last 20 months. A patent issued in that 20-month period that references (cites) a hot patent is called a next generation patent. Note that hot patents in general will have many subsequent patents that reference them, but only citing patents issued during the 20-month period are considered next generation patents. For this study, the 2006 hot patents have next generation patents that were issued between January 2005 and August 2006.

---

[46] The choice of a 20-month period is somewhat arbitrary. When the project began in September 2006, the authors' company database covered patents through August 2006, which was chosen as the ending date. The results would be similar if a different period of issuance was chosen instead counting backward from August 2006.

7. Co-citation Clustering

Identifying hot patents and next-generation patents leads to tens of thousands of potentially interesting patents related to emerging high-risk technologies. However, an analyst would have a difficult time making any sense of such an unwieldy set of documents. In general, patents are very specific and it is not uncommon for a single innovation to be covered by dozens of different patents. Therefore, it is reasonable to assume that the patents identified should have some topical overlaps, and should therefore cluster into a few hundred emerging technologies.

Clustering by technology is actually more difficult than it would seem at first glance. Grouping by keyword seems the most promising, except that patents are often not written in standard English. For example, it is not uncommon for something as simple as a bottle cap to be described as a "closure that is removably fastened" or for a zipper to be described as a "separable fastener," a "zip fastener," or a "slide fastener," among other examples of legalese and not simple English. This type of language tends to diminish the usefulness of keyword clustering tools that have been developed largely for search and retrieval within textbooks or newspapers.

In addition, the U.S. Patent and Trademark Office employs a classification scheme, but this does not lend itself easily to the task of clustering patents for the purpose here. USPTO's classification scheme is based on "art unit" and not technology. As a result, a technology may be spread across several art units, but an art unit will not necessarily contain a single technology. For example, a cellular phone patent may fall into 10 different patent office classifications, but several of those classes may also contain patents for radios, antennas, software, and other technologies.

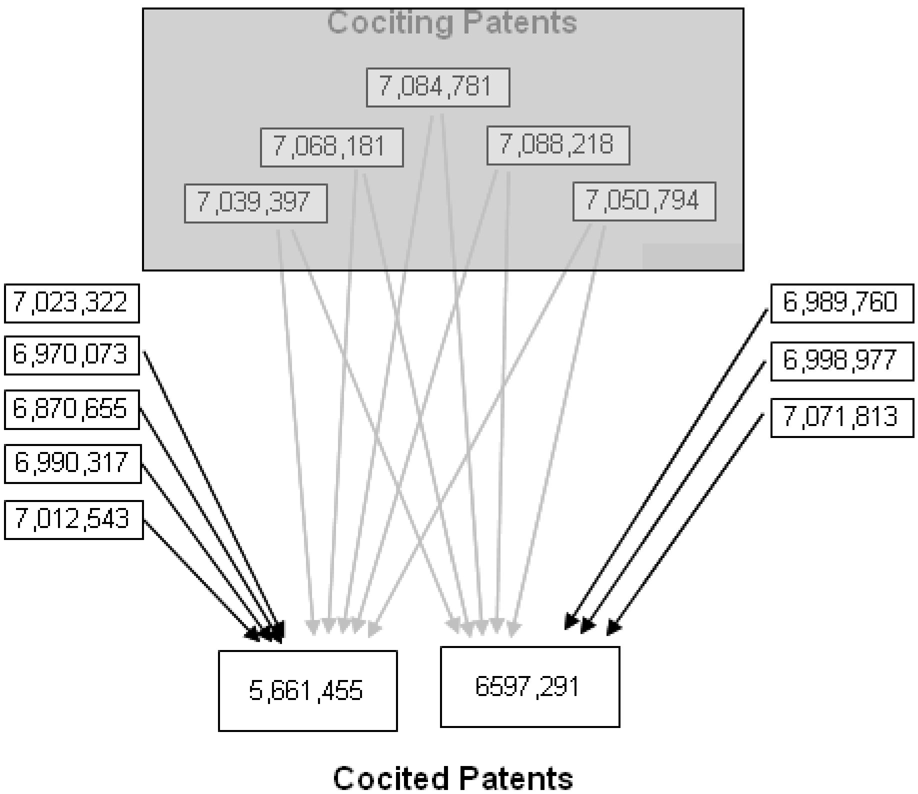
Co-inventor clustering is a method that often works well on small sets of patents issued to the same company, because inventors that collaborate together often work on similar projects. One problem with using inventors to group hot patents (and next generation patents) into clusters of technology is that hot patents are a specialized set that is likely to contain only a fraction of an inventor's career patent output. In addition, hot patents emerge from many different organizations, thus reducing the usefulness of co-inventor clustering.

Co-citation clustering is a method for clustering patents based on the citing patents that they have in common. For example, if 10 patents cite the same three patents it is reasonable to assume that these 10 patents have something in common. Co-citation clustering originated more than 30 years ago,[47] but works poorly in most patent sets because of the few citations that most patents receive in their lifetimes. The set of hot patents is different, and therefore more receptive to co-citation clustering, because each hot patent has received at least 10 citations, and most have received many more.

Figure 1 illustrates the idea behind co-citation clustering. Two hot patents—5,661,455 and 6,597,291—are cited by 10 and eight next-generation patents respectively. However, five of the next-generation patents are common citing patents to both hot patents, so the two hot patents are clustered together. In this case, both patents are related to technology for electronic transmitters for car accessories, garage door openers, and the like, so it is reasonable to place them together in the same cluster.

[47] Henry Small, "Co-citation in the scientific literature: A new measure of the relationship between documents." *Journal of the American Society for. Information Science*, 24, 1973, 265-269.

Figure 1: Example of Co-citation Clustering



8. Hot Patent Cluster

A set of hot patents is identified as described earlier in section I.A.5. A set of hot patent clusters is formed from the hot patents and all of the patents that cite (reference) them based on co-citation clustering as described in section I.A.7.

9. Next Generation Cluster

A set of hot patent clusters gives rise to a set of next generation clusters. The next generation clusters are, however, not formed by co-citation as were the hot patent clusters. Instead, if P_1, P_2, … P_k are patents in a given hot patent cluster, then all of the next generation patents that reference P_1, … P_k are in the corresponding next-generation cluster.

B. Methodology Details for NIST Emerging Clusters Project

1. Hot Patent and Next Generation Sets

The key concepts section above discusses the method for identifying hot patents and next generation patents for 2006. Using this method, we identified 25,973 hot patents issued between 1975 and 2006, and 76,298 next generation patents issued between January 2005 and August 2006. These next generation patents directly cite the hot patents.

Two additional parallel sets of hot patents and next generation patents were created in a similar manner. The additional sets are used for validation tests. The first additional set, the 1998 set, examines the time period between January 1997 and August 1998, and contains 6,969 hot patents and 30,814 next generation patents. The second additional set, the 2002 set, examines the time period between January 2001 and August 2002 and contains 18,823 hot patents and 74,932 next generation patents.

2. Clustering Specifics

Section I.A.7 above describes the co-citation clustering method, which contains many steps compared to the simple concepts discussed.

Most clustering methods use some variation of hierarchical clustering. The basic idea behind hierarchical clustering is illustrated in the example shown in Figure 2. For each iteration of hierarchical clustering, two 'nearest neighbors' are identified and placed together in the same cluster. In this example, after the first iteration, I and J will be clustered together. After the second iteration, {H, I, J} will be grouped in the same cluster. After the third iteration, there will be two clusters {F,G} and {H, I, J}. After the fourth iteration, there will be three clusters {A, B}, {F, G} and {H, I, J}. If this process continues through nine iterations, only one large cluster remains.

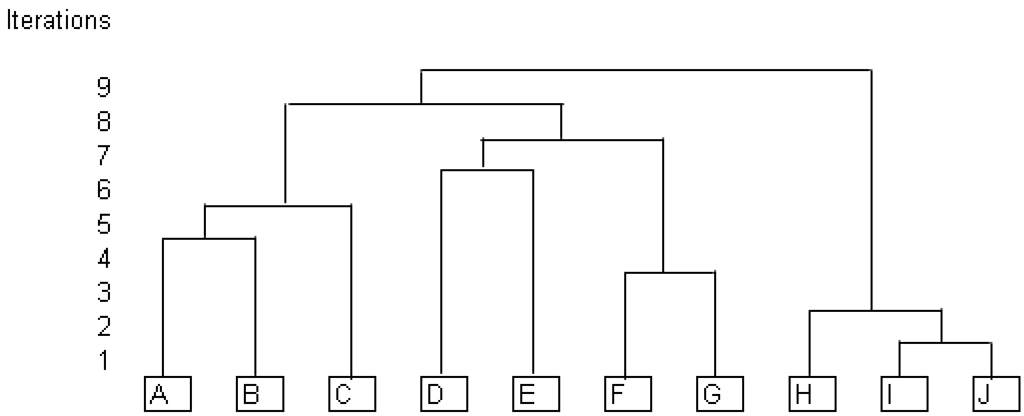
The basic clustering algorithm is as follows:

1. Identify the two most similar documents (patents) and combine them into a cluster.
2. Treat any cluster as a single document.
3. If more than one cluster remains go to 1.

The idea is to cluster the most similar items until the algorithm is stopped before there is a single large cluster. In the example in Figure 2 for instance, it may be best to stop after six iterations, which would create four small clusters {A, B, C}, {D, E}, {F, G}, and {H, I, J}.

Hierarchical clustering can be used to cluster via inventors, citations or keywords. The only difference between them lies in how the most similar documents or 'nearest neighbors' are defined—i.e., by common inventors, keywords of citations. In all methods, every pair of patents will have a pairwise distance between itself and every other patent.

Figure 2: Hierarchical Clustering Example



In using the co-citation clustering method, the distance between any two hot patents is measured by how many co-citing patents any two hot patents have in common. Specifically, since there are 25,973 hot patents and 76,298 next generation patents for the first period (the 2006 set), 25,973 vectors are created, each of length 76,298. To fill the vectors, the 76,298 are numbered, citing patents from 1 to 76,298. Each of the 25,973 hot patents' vectors is filled by placing a 1 in each entry where there is a citing patent and a 0 in all other entries. For example, if patent A is cited by the citing patents, labeled 10, 52, 273, and 2400, then the vector for patent A will have 76,294 zero entries and will have a 1 in entry 10, 52, 273, and 2400.

The similarity measure used is a popular one called the Dice coefficient that computes:

Similarity(A,B) = 2*(# Co-citing patents in common)/(# Patents citing A + # Patents citing B).

Using these vectors this is essentially

$$\text{Similarity(A,B)} = \frac{2\sum_{k=1}^{76.298}(A_k \bullet B_k)}{\sum_{k=1}^{76,298}(A_k^2 + B_k^2)}$$

Note that the squares in the denominator can be ignored when the vector elements are 1, but when the vector weights change, these squares become important.

Using the formula above, if patents A and B are cited by 6 and 8 citing patents respectively, and 4 of the citing patents are common to both, the similarity will be 8/(6 + 8) = 8/14 = 4/7 = 0.57. An equivalent way of thinking about this example is that 57 percent of the referencing patents are common to both A and B.

The method described above was amended to account for referencing patterns. The example above gives each citing patent equal weight, which is not necessarily a valid approach. To illustrate this point, consider a patent that references just eight patents versus a patent that references 620 patents. If all citations are weighted equally, than the second patent has more than 75 times the influence of the first patent. Since the relative importance of each of the 620 references is probably less on average than the eight patents in the first example, the influence should reflect that.

To see why this is, consider for a moment keyword clustering instead of co-citation clustering. In keyword clustering, common words are put into a vector instead of co-citing patents. In keyword clustering the word ”article” may appear in 200 patents, but the word ”Quinazoline” may appear only in five patents. In this case it is clear that the word ”Quinazoline” has much more discriminating power than ”article” and the two patents that both mention ”Quinazoline” are much more likely to be related than two patents that both mention ”article.”

To adjust for this phenomenon, a weighting factor has been developed called the Inverse Document Frequency or IDF. IDF is defined as follows:

$$\text{IDF(citing patent)} = \log_2\left(\frac{\textit{source documents}}{\textit{frequency}(\textit{citing documents})}\right)+1$$

The IDF of a patent citing 620 hot patents would be $\log_2(25{,}973/620)+1 = 6.39$, while the IDF of a patent citing only eight hot patents would be $\log_2(25{,}973/8)+1 = 12.66$.

In other words, the citing patents that reference only a few hot patents are given a much higher weight than those that reference many different patents. Once the IDF is added, everything in the description of the vectors and dice-coefficient remains the same, except that entries that were filled with 1 in the discussion above are instead filled with the weights computed via the IDF.

Once the pairwise similarities are computed for all possible combinations, clustering can proceed by identifying the pair with the highest similarity. The algorithm that is used is actually a two-stage algorithm designed to avoid creating overly large clusters. The algorithm is as follows.

1. Compute similarities for all hot patent pairs, yielding (25,973*25,972)/2 = 337,288,378 similarities.
2. Choose pair with highest similarity and cluster together.
3. Choose pair with next highest similarity.

   If similarity is greater or equal to 0.9 then

   a) If both patents are unclustered create a new cluster with patent pair as members.
   b) If one patent is in a cluster and the other is not, add unclustered patent to the other patent’s cluster.
   c) If both patents are clustered, combine the two clusters into a larger cluster.

   If similarity is greater or equal to 0.5, but less than 0.9, then

   a) If both patents are unclustered create a new cluster with patent pair as members.
   b) If one patent is in a cluster and the other is not, add unclustered patent to the other patent’s cluster.

c) If both patents are clustered, combine the two clusters into a larger cluster unless both clusters already contain at least 10 patents. In that case, keep clusters independent.

If similarity is less than 0.5 then stop clustering process.

4. Repeat step 3 until largest similarity is less than 0.5.

Essentially the algorithm works as follows. All of the pairwise similarities are computed and sorted such that the patent pairs with the highest similarities are at the top of the list. The patent pairs are then clustered as in Figure 2 until the similarity scores dip below 0.9. At this point the algorithm continues with one minor change. If two clusters are due to be merged, first check that at least one has fewer than 10 patents.

In this way, large clusters only get to be large by having very high similarities. Note that two patents with a similarity score exceeding 0.9 is roughly equivalent to the patents having more than 90 percent of their referencing patents in common.

For the 2006 hot patent set, the clustering resulted in more than 1,500 clusters of size 2 or more, including 436 of size 5 or more, 241 of size 10 or more, 87 of size 25 or more, 35 of size 50 or more, and five of size 100 or more.

The same algorithm was then applied to the 2002 and 1998 hot patent sets, so that each set could be validated. Note that for each hot patent cluster there is a corresponding next generation cluster that consists of all patents in the next generation that reference a given hot patent cluster.

Appendix B Validation of the Hot Patent Method

An obvious problem with trying to develop a tool that is predictive is that time is needed to determine whether the prediction was accurate. This method, it is hoped, identifies high-risk, emerging technologies. However, it is not clear until time passes whether something that seemed like an interesting idea was truly an emerging technology.

One of the reasons the authors identified two sets of hot patents and next generation patents for earlier times periods (1998 and 2002) was to test the validity of the methods in identifying high-risk technologies. Specifically, they tested whether high-risk technologies exist at a higher rate within the next-generation patents from 1998 and 2002 than in the general population of patents.

I. Validation of Hot Patent Method via ATP Patents

A. The ATP Patent Set

One set of patents known to be based on high-risk research is the set of patents that emerged from projects funded by the National Institute of Standards and Technology (NIST) Advanced Technology Program (ATP). This program sponsors high-risk, emerging technology developments that are deemed too risky to attract private funding. The patents produced by ATP-funded projects have been shown to be high impact and, in some cases, very valuable.[48]

The following description is taken directly from an ATP fact sheet[49]

> The ATP views R&D projects from a broader perspective—*its bottom line is how the project can benefit the nation*. In sharing the relatively high development risks of technologies that potentially make feasible a broad range of new commercial opportunities, the ATP fosters projects with a high payoff for the nation as a whole—in addition to a direct return to the innovators. The ATP has several critical features that set it apart from other government R&D programs:
>
> - ATP projects focus on the technology needs of American industry, not those of government. Research priorities for the ATP are set by industry, based on their understanding of the marketplace and research opportunities. For-profit companies conceive, propose, co-fund, and execute ATP projects and programs in partnerships with academia, independent research organizations and federal labs.
> - The ATP has strict cost-sharing rules. Joint Ventures (two or more companies working together) must pay at least half of the project costs. Large, Fortune-500 companies participating as a single firm must pay at least 60 percent of total project costs. Small and medium-sized companies working on single firm ATP projects must pay a minimum of all indirect costs associated with the project.
> - The ATP does not fund product development. Private industry bears the costs of product development, production, marketing, sales and distribution.

48 NIST-ATP "ATP Economic Studies, Survey Results, Reports and Working Papers," 2006. http://www.atp.nist.gov/eao/eao_pubs.htm. [Last accessed October 2007] and
NIST-ATP "ATP Is Meeting Its Mission: Evidence From ATP Evaluation Studies," 2005. http://www.atp.nist.gov/factsheets/1-a-1.htm. [Last accessed October 2007]
49 NIST-ATP "How ATP Works," 2005. http://www.atp.nist.gov/atp/overview.htm [Last Accessed October 2007].

- The ATP awards are made strictly on the basis of rigorous peer-reviewed competitions. Selection is based on the innovation, the technical risk, potential economic benefits to the nation and the strength of the commercialization plan of the project.
- The ATP's support does not become a perpetual subsidy or entitlement—each project has goals, specific funding allocations, and completion dates established at the outset. Projects are monitored and can be terminated for cause before completion.

> The ATP partners with companies of all sizes, universities and non-profits, encouraging them to take on greater technical challenges with potentially large benefits that extend well beyond the innovators—challenges they could not or would not do alone. For smaller, start-up firms, early support from the ATP can spell the difference between success and failure. To date, more than half of the ATP awards have gone to individual small businesses or to joint ventures led by a small business. Large firms can work with the ATP, especially in joint ventures, to develop critical, high-risk technologies that would be difficult for any one company to justify because, for example, the benefits spread across the industry as a whole.[50]

The ATP program has been widely evaluated. Various case studies provide evidence that the benefits to the nation of the program have far exceeded its cost. Evaluation studies have attributed more than $18 billion in present value social benefits from 40 sponsored projects, which is eight times the total spent by the program ($2.3 billion) to date.[51]

Although the ATP program is controversial, one thing that is clear is that it has led to the acceleration and development of key technologies. The patents that have resulted from the program's cost-shared funding are therefore an excellent test bed for the hot patent method. If the hot patent method truly identifies emerging technologies, then there should be a higher incidence of ATP patents identified by the method than in a random set of patents.

B. Results of ATP Validation

In Table B.1, the key results of the validation study using ATP patents are shown. During the time period between January 2001 and August 2002, 274,200 patents were issued by the U.S. Patent and Trademark Office, of which 27 percent are in the next generation set from that period. Also, 107 patents issued during that period were based on research funded by ATP. Now, given a random set of 75,000 patents from the period, about 29 ATP-related patents would be expected within that set ((75000/274200)*107). In fact there are 53 ATP related patents in the 2001-2002 next-generation set, which is almost twice as many as expected (81 percent more than expected, to be precise).

In the earlier time period from January 1997 through August 1998, Table B.1 reveals that there are 102 percent more ATP-related patents in the next-generation set compared with what was expected.

In recent years, ATP's budget has been severely cut and, for three years prior to 2007, competitions were not held. As a result, there are not enough recent patents to perform a similar validation study of the 2005-2006 next-generation patents. However, given the validation studies based on the earlier periods, there is reason to believe that the patents identified by the hot patent method are more likely to contain high-risk, emerging technologies than a random set of patents.

---

[50] Ibid.
[51] Op cit 16.

Table B.1 Key Results of ATP Validation

Patent Sets	Jan 2001 - Aug 2002	Jan 1997 - Aug 1998
Number of U.S. Utility Patents Issued	274,200	205,152
Number of Next Generation Patents	74,932	30,814
Percent of Next Generation Patents	27.30	15.00
Number of ATP Patents	107	66
Number of ATP Patents in Next Generation	53	20
Expected #ATP in Next Generation	29	10
Actual/Expected ATP in Next Generation	1.81	2.02
Number of ATP Projects with Patents	58	43
Number of ATP Projects with Next Generation Patents	29	15
Expected Number of ATP Projects with Next Generation Patents	16	6
Actual/Expected ATP Projects in Next Generation	1.83	2.32

II. Identifying Emerging Clusters

A. Prospective Patent Parameters

As discussed in Appendix A, citation analysis is one method for identifying sets of emerging, potentially high-impact technologies. However, citation analysis and citation metrics are retrospective tools, which means that time must pass before a citation metric can be applied to a patent or patent set. This drawback makes it difficult, based on currently issued patents and their citations, to identify emerging technologies, which are prospective in nature.

Prospective parameters (i.e. metrics that can be measured at the time the patents issue) are thus needed. The authors have identified four parameters—public sector presence, science index, originality index, and reference index—that can be measured at the time the patents issue. Moreover, these parameters have higher scores for clusters that contain ATP patents. This suggests that the parameters, as described, may be useful in identifying emerging, high-risk technologies.

- **Public Sector Presence.** Of clusters containing ATP patents, 56.1 percent contain at least one patent assigned to a university or government agency (only 13.9 percent of all non-ATP clusters contain a public sector patent). Projects that meet ATP's selection criteria pursue—through the involvement of companies, universities, and government laboratories—high-risk, enabling research with potential for delivering broad economic impact. Hence, clusters that contain public sector patents are more likely to describe technologies that are early-stage, and therefore higher risk.

- **Science Index.** ATP clusters contain 166 percent more science references than peer patents of the same age and technology class. Non-ATP clusters contain 61 percent more science references than expected (this suggests next generation patents have high science linkage, but ATP clusters have unusually high science linkage). It follows that emerging, leading-edge technologies tend to reference scientific articles as prior art, rather than just older patents. This makes sense intuitively, as one would expect that incremental improvements in a technology would reference only older patents, whereas emerging, high-risk technologies are more likely to build upon the latest scientific advances.

- **Originality Index.** ATP clusters have an originality index 9 percent above the expected value. Non-ATP clusters have an originality index 3 percent above the expected value. The originality metric measures the extent to which a patent combines disparate ideas. Patents that combine several technologies to create a new technology are deemed more original, and "enabling" for generating broader downstream commercial applications than patents that build upon a single technology area.

- **Reference Index.** ATP clusters have 2.5 times as many prior art references as expected. Non-ATP clusters have 1.6 times as many as expected. This metric is also somewhat related to the originality index. Again, patents that build upon many ideas will tend to have more prior art references than simple, incremental improvement patents.

B. Deriving a Scoring Equation

Monte Carlo methods are numerical simulation methods where sequences of random numbers are used to perform the simulation. The methods are used in a variety of applications in chemistry, physics, economics, mathematics and other areas. The methods go back hundreds of years, but have gained in popularity in the last 50 years because they allow one to model complex problems by running an experiment repeatedly on a computer.

A typical application of a Monte Carlo method is where one has multiple parameters that affect an outcome, but it is not clear how each parameter should be weighted to optimize the outcome. In this case, there are several measurable parameters related to patent clusters, but it is not clear which ones indicate a cluster is likely to be a high-risk emerging cluster.

For this purpose, a Monte Carlo simulation is essentially an optimization problem. The goal is to find coefficients A, B, C, and D for the following equation:

Score = A*(Public Sector Avg) + B*(Science Index) + C*(Reference Index) + D*(Originality),

so that the score for ATP-related clusters is high and the score for non-ATP clusters is low.

Since the ATP patent set is relatively small compared to the universe of patents, it is used to derive a scoring equation that can then be applied to the general universe of patent clusters. If the scoring equation works, this would mean that the high scoring clusters are more likely to contain emerging, high-risk technologies than low-scoring clusters.

To create the scoring equation, more than 100,000 simulations were run with random values of A, B, C, and D for the 2002 next generation clusters. The coefficients that were selected had the desired effect that ATP-related clusters scored much higher than non-ATP clusters. The final scoring equation is:

Score = 87*(Public Sector Avg) + 43*(Science Index) + 4*(Reference Index) + 1*(Originality).

C. Validating the Scoring Equation

The scoring equation derived in the previous section was developed using the 2002 next-generation clusters that contain ATP-related patents. This scoring equation can then be used to identify clusters with the same characteristics as the ATP-related clusters. These clusters are considered to be more likely to contain emerging technologies.

Table B.2 shows the results of the scoring equation applied to all clusters in the 2002 set. Note that the ATP clusters have an average score that is more than twice as high as the non-ATP clusters. The difference in median scores is even more dramatic.

Table B.2 Scoring Differences between ATP and Non-ATP Next Generation Clusters (2002 Set)

	ATP Related Clusters	Non-ATP Related Clusters	ATP / Non-ATP
Avg. Score	121	55	2.2
Median Score	86	33	2.6
Percent with Score > 86	50%	18%	2.8
Percent with Score > 50	80%	35%	2.3

Looking at things slightly differently, half of the ATP-related clusters score above 86, but only 18 percent of the non-ATP clusters score above 86. Similarly, 80 percent of ATP-related clusters score above 50, but only 35 percent of non-ATP clusters score above 50.

Appendix C Fluid Handling Firms

- designs, manufacturers and markets compressor and vacuum products and fluid transfer products.
- filtration, separation and purification technologies, made by the company using its filter media, and other fluid clarification and separations equipment for the removal of solid, liquid and gaseous contaminants from a range of liquids and gases.
- flagship flushometer products (hydraulic plumbing components that feature a unique diaphragm-type valve), toilets and urinals do their jobs with greater water efficiency. The company also makes faucets, sinks, showerheads, bed pan washers, and hand dryers, several of which are sensor-activated. Its HEALTHMINDER products include hygiene-friendly dispensers for soap, hand sanitizers, and air fresheners.
- fluid and metering technologies, health and science technologies, dispensing equipment and fire, safety and other diversified products.
- fluid handling solutions
- fluid system components, which include plug, pinch, and radial diaphragm valves, sanitary fittings, and welding systems. Its products are used in the oil and gas, power, chemical, food and beverage, and semiconductor industries, as well as by bioprocess and pharmaceuticals research companies.
- leading manufacturer of industrial spray nozzles.
- lubrication systems, pumps, and other equipment and tools—heavy-duty pumps, grease fittings and accessories, automated lubrication systems, and a battery-powered grease gun.
- manufacturer of hydraulic and electrohydraulic controls and control valves for makers of off-highway equipment as well as passenger cars and light trucks.
- pneumatic valves, controls systems, and safety products for the fluid power industry
- precision-engineered flow control equipment, such as pumps, valves and seals, for critical service applications.
- ultrahigh-pressure (UHP) water pumps and systems that are used to cut and clean materials as an alternative to traditional cutting methods, such as lasers, saws or plasma.
- worldwide leader in magnetically levitated Bearingless Motor technology, specializing in supplying medical blood pumps to the medical community and ultra-pure fluid handling devices for both semiconductor and industrial applications. The patented technology permits the motor and magnetic bearing to be combined into a single unit with products that achieve maximum reliability, long life, and the ability to pump precious fluids in the harshest of environments.

Appendix D Unclassified Firms

Firm	Size	Description
Advanced Lighting Technologies	L	designs, manufactures and sells metal halide materials, components and systems,
Amerasia International Technology	S	one of the leading forces in development and patented applications of advanced material and adhesive solution for electronic interconnection and packaging.
Antaya Technologies	S	glass power and signal connection systems and automated and semi-automated controlled soldering equipment.
Basic Resources	S	flux supply of primary and secondary recyclers in the aluminum industry
Burstein Technologies	S	coupling a new and exciting technology with the instant information processing capability of the Internet to revolutionize chemical and biological testing and the availability and flow of related information and products. Our platform will enable individuals to use specially designed CD-R discs and slightly modified CD/DVD reader either as a standalone instrument or attached to an ordinary personal computer to deliver a wide range of tests including clinical laboratory diagnostics, biological warfare agent detection, forensic DNA tests, and food and water contamination tests.
Clinical Data	S	worldwide provider of pharmacogenomics and molecular services, as well as genetic testing and clinical diagnostics to improve patient care and clinical outcomes.
DIGITAL CONTROL	S	directional drilling tracking equipment
Healthways	L	Health and Care Support solutions to help people maintain or improve their health, and as a result, reduce overall healthcare costs. To e
Immersion	S	provides haptic technologies that allow people to use their sense of touch in operating digital devices. The Company develops and manufactures or licenses hardware and software technologies and products for original equipment manufacturers (OEMs).
Incept LLC	S	a technology incubator in
Information Resources Management LLC.	S	provides area businesses with premium information technology solutions including; Consulting services, Networking Services, Project Management services, IT Assessments, Documentation, and Training.Headquartered in Columbus, we provide small businesses and Nonprofit organizations with a pool of talented technology professionals to assist with information resources management needs
Intuitive Surgical	L	
MSP	S	Advanced instruments, products and services to the semiconductor, pharmaceutical, and air pollution industries. We are known around the world for advanced aerosol particle control, offering ingenious solutions to complex aerosol problems involving the generation, deposition, sampling and/or measurement of nano particles from 3nm to 100 µm in size. Well-conceived displays, features and interfaces are combined with our advanced technology
Mine Safety Appliances	L	Sophisticated safety products typically integrate any combination of electronics, mechanical systems and advanced materials to protect users against hazardous or

Mirus Bio	S	Mirus Bio provides gene transfer technologies to the pharmaceutical industry. The company's reagent kits assist researchers with in vitro transfection, in vivo delivery, DNA and RNA labeling, and RNA interference. Besides its kits, Mirus Bio is also developing its own therapies based on its intravenous technology, called Pathway IV, that facilitates the delivery of plasmid DNA to targeted muscles. One of those therapies is MyoDys, a gene therapy designed to treat muscular dystrophy by delaying loss of muscle function. Mirus Bio is also developing treatments for peripheral ischemia and anemia. The company was founded in 1995 by Jon Wolff, James Hagstrom, and Vladimir Budker
Morning Pride Manufacturing	S	Firefighters Protective Clothing and Equipment
Nalco Holding	L	Integrated water treatment and process improvement services, chemicals and equipment programs for industrial and institutional applications
Navigation Technologies	L	premium-quality digital map data.
PTS	S	component level rebuilds of a wide variety of consumer electronic products.
Production Resource Group LLC	L	provides lighting, audio, scenery, video, and labor services for entertainment companies in markets such as corporate and special events, tradeshow, theater, concert touring, television and film production, retail, and themed attraction.
Reflexite	L	high visibility, reflective products, as well as, lighting optics, display optics, and instrumentation .At Reflexite our business is the Management of Light
Silicon Genesis	S	SiGen is a provider of technology, processes and equipment to semiconductor, solar, display and optoelectronics markets. To that end the Company achieves its revenue goals through a combination of licenses, royalties, services and equipment sales.
Spraycool	S	develops products and technologies used in cooling and packaging applications for high-performance electronic systems
Symyx Technologies	S	scientific research and development (R&D) integration partner to companies in the life sciences, chemical, energy, consumer products and electronics industries.
Synaptics	S	developer and supplier of custom-designed user interface solutions that enable people to interact with a variety of mobile computing, communications, entertainment and other electronic devices. The Company targets the personal computer (PC) market and the market for digital lifestyle products, including portable digital music and video players, mobile phones, and other select electronic device markets with its customized interface solutions
Woodstream	S	poison free pest control and care products for pets and wildlife; hardware and pet supplies Steam traps, Liquid strainers, Liquid traps, Steam strainers, Network packet data synchronization device, Hardware or telephony adapters, Transceivers and media converters
TiVo	S	TiVo is a provider of technology and services for digital video recorders (DVR).
Website Pros	L	Websites and Web services focused on helping small and medium-sized businesses
Xybernaut	S	research, development, manufacture, marketing and sale of mobile/ wearable computing and communication systems and software and service solutions designed to enhance productivity and improve product management, asset management and the accuracy, timeliness and utilization of critical enterprise data.