NBER WORKING PAPER SERIES

SOME FACTS OF HIGH-TECH PATENTING

Michael Webb
Nick Short
Nicholas Bloom
Josh Lerner

Working Paper 24793
http://www.nber.org/papers/w24793

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2018

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24793.ack

Some Facts of High-Tech Patenting
Michael Webb, Nick Short, Nicholas Bloom, and Josh Lerner
NBER Working Paper No. 24793
July 2018
JEL No. L86,O34

ABSTRACT

Patenting in software, cloud computing, and artificial intelligence has grown rapidly in recent years. Such patents are acquired primarily by large US technology firms such as IBM, Microsoft, Google, and HP, as well as by Japanese multinationals such as Sony, Canon, and Fujitsu. Chinese patenting in the US is small but growing rapidly, and world-leading for drone technology. Patenting in machine learning has seen exponential growth since 2010, although patenting in neural networks saw a strong burst of activity in the 1990s that has only recently been surpassed. In all technological fields, the number of patents per inventor has declined near-monotonically, except for large increases in inventor productivity in software and semiconductors in the late 1990s. In most high-tech fields, Japan is the only country outside the US with significant US patenting activity; however, whereas Japan played an important role in the burst of neural network patenting in the 1990s, it has not been involved in the current acceleration. Comparing the periods 1970-89 and 2000-15, patenting in the current period has been primarily by entrant assignees, with the exception of neural networks.

Michael Webb
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305-6072
mww@stanford.edu

Nick Short
Harvard Kennedy School
nshort@g.harvard.edu

Nicholas Bloom
Stanford University
Department of Economics
579 Serra Mall
Stanford, CA 94305-6072
and NBER
nbloom@stanford.edu

Josh Lerner
Harvard Business School
Rock Center 214
Soldiers Field
Boston, MA 02163
and NBER
jlerner@hbs.edu

Replication files are available at http://www.michaelwebb.co/data/

1 Introduction

Economists have increasingly recognized the impact of legal and economic institutions on the process of growth (La Porta et al., 1998; Acemoglu et al., 2002). One of the most important of these is the patent system, which economists have argued has the potential to provide a critical spur to innovation.[1] Ideally, it has been argued, patents rewards inventors in a manner that is commensurate with the nature of their discoveries (e.g., see the discussion in Scotchmer, 2004). But important questions surround whether the patent system can adequately address fundamentally new areas of technology. Among the barriers to efficacious patent awards that have been highlighted by academics and practitioners are the difficulty of defining patentable subject matter, the relative inexperience of patent examiners with these technologies, and the absence of patented prior art.

Perhaps no area has been more contentious than software patenting.[2] Bessen and Hunt (2007) document the dramatic increase in software patenting between 1976 and 2002. This surge in patenting accelerated after the U.S. Supreme Court's decision in the 1981 case of *Diamond v. Diehr*, which established that software used for industrial purposes could be patentable, and the Court of Appeals for the Federal Circuit's ruling in *State Street Bank v. Signature Financial Group*, which established that a numerical calculation that produces a "useful, concrete and tangible result", such as a price, is patent-eligible. Bessen and Hunt (2007) highlight that the growth in patenting was driven by large manufacturing firms in industries known for strategic behavior. Software patenting seems orthogonal to changes in software R&D, programmer employment, or other indicators. In general, the firms undertaking software patenting actually had lower R&D intensity then their peers.

In this paper, we revisit this territory, focusing on awards in software and related technologies over the past two decades, including cloud computing and artificial intelligence. This region has seen dramatic changes over this period, beginning with a series of widely discussed decisions by the U.S. Supreme Court that have altered the costs and benefits of pursuing these awards. Most notable were the 2009 ruling in *In re Bilski*, which raised substantial questions about the patentability of business method patents, and *CLS Bank International v. Alice Corp.*, a 2014 decision that raised questions as to the extent to which implementing an existing process in software would be patentable.

This period has also seen dramatic changes in firm behavior regarding patents. In particular, firms appear to be increasingly using these awards as part of battles for market dominance. Perhaps no more dramatic illustration is the litigation around smartphone technology: litigation across at least 10 countries enveloped these devices, with at least 50 lawsuits between Apple and Samsung and, until their May 2014 settlement agreement, 20 cases between Apple and Google. (These tabulations do not count the on-going lawsuits between Google and the Rockstar Consortium, whose members include Apple, Black-

[1] There is, of course, a large literature on patents. Representative papers include Griliches (1998), Lanjouw et al. (1998), Jaffe and Trajtenberg (2002), Hall et al. (2005), Galasso et al. (2013), Cohen et al. (2014), Dorn et al. (2016), and Kogan et al. (2017).

[2] Recent literature on software patenting includes Allison and Tiller (2003), Graham and Mowery (2003), Hall and MacGarvie (2010), Mann (2004), and Shalem and Trajtenberg (2009).

Berry, Ericsson, Microsoft, and Sony, and which purchased for $4.5 billion the patent portfolio of bankrupt multinational telecommunications and data networking equipment manufacturer Nortel in 2011). Other illustrations would include litigation around self-driving car software between Uber and Waymo, and the efforts on the part of Microsoft to shield its cloud users from infringement suits through aggressive licensing.

We document a continued dramatic increase in software patents during the course of the twenty-first century, with a 60.2% increase in ultimately successful filings between 2000 and 2013, and a 168.6% increase in applications over the same period. The rate of increase—albeit on a modest base—is far more dramatic for many of the emerging technologies, such as drones, cloud computing, and machine learning. We show that these new technological fields are characterized by rapid bursts of innovation from a relatively small group of inventors, followed by a slowing down in per-inventor productivity as more inventors pursue these opportunities. Invention remains dominated by U.S., Japanese, and Korean inventors, many of them at large firms with a strong patenting history.

While we are unable to answer the question of the social desirability of these awards, we hope to explore these issues in future research. In particular, in subsequent work, we will examine the impact of these awards on market valuations, particularly around the major judicial decisions alluded to above. Nonetheless, given the plausibility that today's legal and economic institutions surrounding intellectual property will have major effects on the future technological landscape of software, artificial intelligence, and cloud computing—as well as technologies that are critically software-dependent, including smartphones and self-driving cars—collecting systematic evidence on patenting around these different technologies appears to be an important task.

The plan of this paper is as follows. We provide an introduction to the data sources and methodology we use in Section 2. Sections 3 through 6 focus on the number of applications and awards, the number and geographic distribution of inventors, and the top assignees respectively. Section 7 seeks to understand the extent to which new patents are coming from current versus new inventors. The final section provides an agenda for future research in these areas.

2 Data and definitions

We use the IFI Claims patents dataset as our source of information on patents.[3] This dataset contains the full text of all published US patent documents through February 2018 obtained from USPTO bulk files. We consider all patent publications with non-missing title, abstract, description, assignee, and inventor fields. In some cases we restrict to granted patents, as described in the text; in other cases we consider all applications. We do not make any other restrictions, such as considering only utility patents or excluding continuation patents.

[3]The dataset is accessed at `https://bigquery.cloud.google.com/table/patents-public-data:patents.publications`.

We describe how we classify patents as corresponding to particular classes of technology below. First, however, we make some observations about the patent data in general. When analyzing patent assignees, we face three data quality issues. The first is name harmonization. For example, patents assigned to IBM may list "IBM" or "International Business Machines", or some misspelling of these, as the assignee. It is important for our purposes to analyze them as belonging to the same, single assignee. The IFI Claims dataset uses the EPO's list of standardized applicant names[4] to harmonize assignee names. This substantially solves the name harmonization problem.

The second issue concerns reassignments. In the course of our analysis, we discovered that in many cases the USPTO data files incorrectly list the inventors of patents as the assignees. This issue appears to begin after the year 2000. A separate USPTO data product, the USPTO Patent Assignment Dataset (Marco et al., 2015), records re-assignments to employers, as well as reassignments for other reasons, such as securitization or acquisition. The dataset attempts to classify reassignments as initial employer assignments or reassignments for other reasons. We plan to use the initial employer assignments if necessary, but not reassignments for other reasons. Currently, we are not accounting for any reassignments.

The third issue concerns subsidiaries. Large parts of a parent company's patent portfolio may be assigned to their subsidiaries. This may be for many reasons: the subsidiary may have joined the parent company through an acquisition; the subsidiary may be used to maintain secrecy about the parent's innovative activities for strategic reasons; there may be tax advantages to assigning patents to subsidiaries in foreign jurisdictions; and patents may be assigned to a subsidiary in order to prepare the subsidiary to be spun out or sold. In one famous example, Ewing and Feldman (2012) identify 1,276 separate shell companies operated by Intellectual Ventures, a company that owns and licenses large numbers of patents. Failing to trace the ultimate owner of a given patent via a corporate tree would mean missing key features of the changing landscape of patent ownership. This is something we are currently working on with Innography, a private provider of patent data that specializes in tracing the ultimate owners of patents. In the current draft, we do not use the Innography data.

2.1 Definitions

We consider patents corresponding to several technological fields as defined in Table 1. Some definitions use keyword inclusion/exclusion criteria for text fields as in Webb et al. (2018). Others use Cooperative Patent Classification (CPC) codes. The fields were selected, first, to represent a range of areas against which to compare our results for software, and second, to highlight the patenting landscape around important emerging technologies. We use two different definitions of software patents: one from Bessen and Hunt (2007), and the other from Webb et al. (2018). These definitions are discussed in Appendix A.1.

To illustrate the kinds of technologies being patented within each of our technology classes, we con-

[4] Available at `https://www.epo.org/searching-for-patents/data/coverage/regular.html`. Last updated 2018. See also Magerman et al. (2006), which describes the harmonization procedure.

sider how the CPC codes to which patents in each technology class were assigned changed between the 1970s/80s period and the 2000s/2010s period. The tables in Appendix A.2 display the 4-digit CPC code prefixes that were most frequently assigned to the patents in each technology class.

Table 1: Technology class definitions

Technology	Definition
Software (Webb et al.)	Title/abstract include 'software', 'computer', or 'program' AND title/abstract exclude 'chip', 'semiconductor', 'bus', 'circuit', and 'circuitry'
Software (Bessen and Hunt)	Specification includes 'software' OR ('computer' AND 'program') AND title excludes 'chip', 'semiconductor', 'bus', 'circuit', and 'circuitry' AND specification excludes 'antigen', 'antigenic', and 'chromatography'
Smartphones	CPC codes include H04 (electric communication techniques)
Drones	CPC codes include B64C2201 (unmanned aerial vehicles)
Machine learning	Title/abstract include 'machine learning', 'supervised learning', 'SVM', or 'support vector machine'
Neural networks	Title/abstract include 'neural network'
Cloud	Title/abstract include 'cloud comput'
Self-driving cars	Title/abstract include 'autonomous vehicle'
Semiconductors	CPC codes include H01L (semiconductor devices)
Pharmaceuticals	CPC codes include A61K (preparations for medical, dental, or toilet purposes)
Internal combustion engines	CPC codes include F02B (internal-combustion piston engines; combustion engines in general)

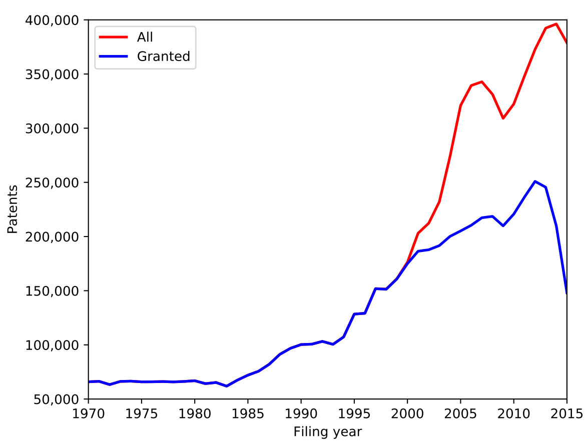
3 Patent applications v. grants

The figures in this section display counts of patent applications and patent grants for each technology by filing year. Note that the USPTO only started publishing patent applications (rather than grants) filed on or after November 29, 2000, and publishes them eighteen months after the effective filing date of the application. The change in 2000 is clear in the figure below.

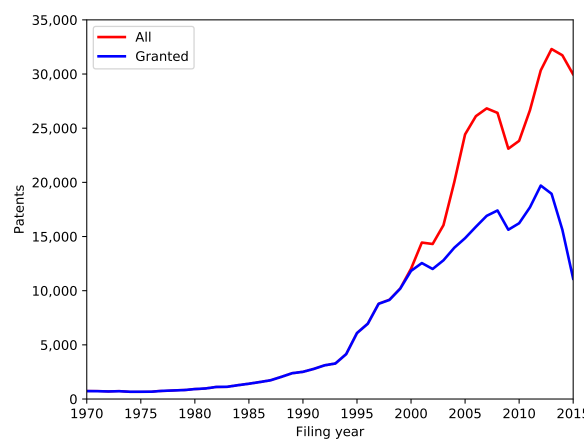
It should be noted that in each case, the number of patent grants falls off at the end of the sample. This reflects the truncation of awards affected by long patent processing times. Many of the awards in 2015 (and even earlier years) had not issued by the beginning of 2018. (For a fuller discussion, see Lerner and Seru, 2017).

The growth of software patents increased sharply after 2000, particularly when patent applications are considered. The number of software patent applications grew by 168.6% between 2000 and 2013. This growth mirrors that of patents more generally: patent applications overall grew by 122.6% over the same period. As we move into more recent technologies, such as cloud, drones, machine learning, and self-driving cars, the growth is far more dramatic. Meanwhile, the number of issued awards in internal combustion engines and pharmaceuticals, included for the sake of comparison, has been nearly flat.

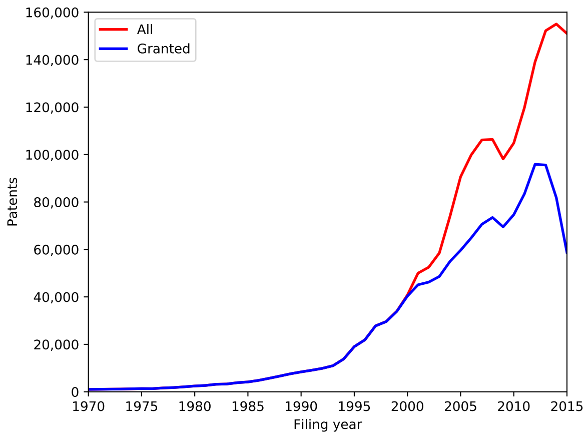
All



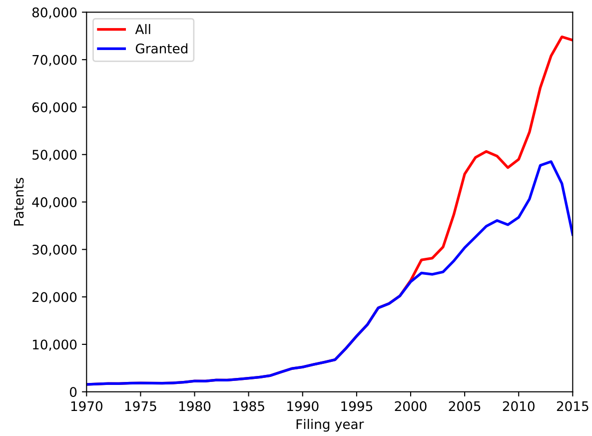
Software (Webb et al.)



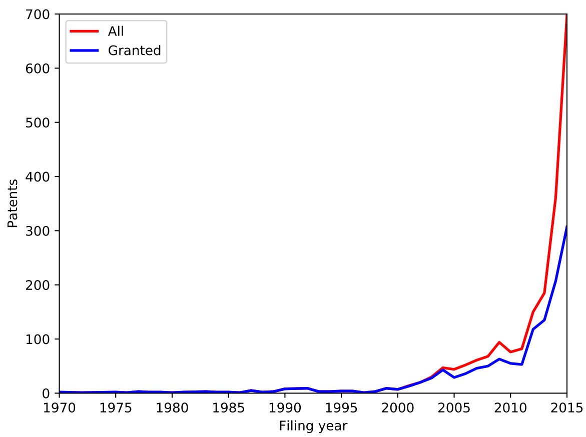
Software (Bessen and Hunt)



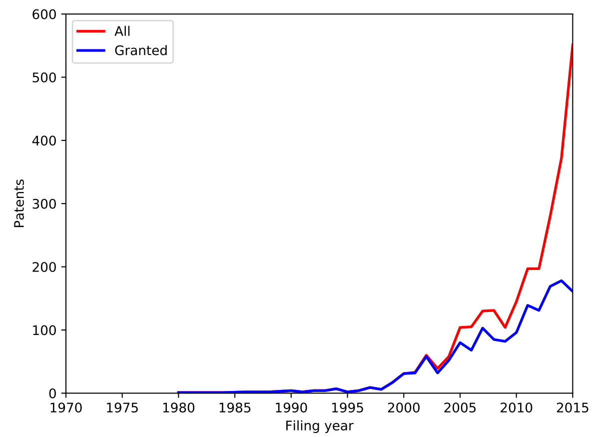
Smartphones



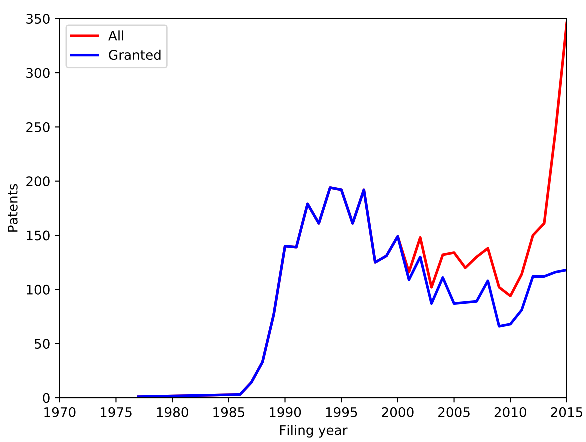
Drones



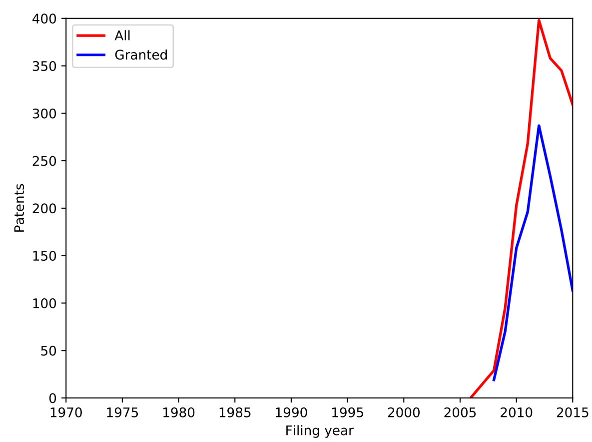
Machine learning



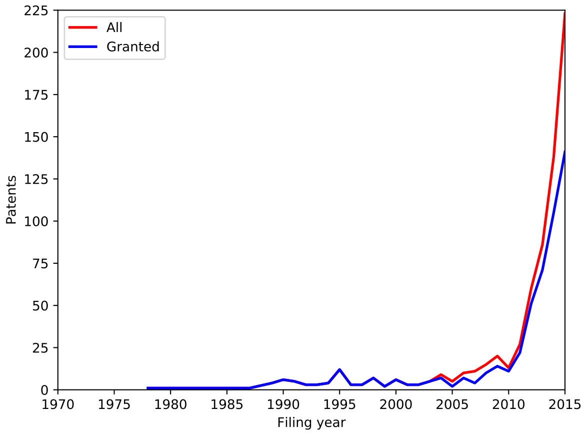
Neural networks



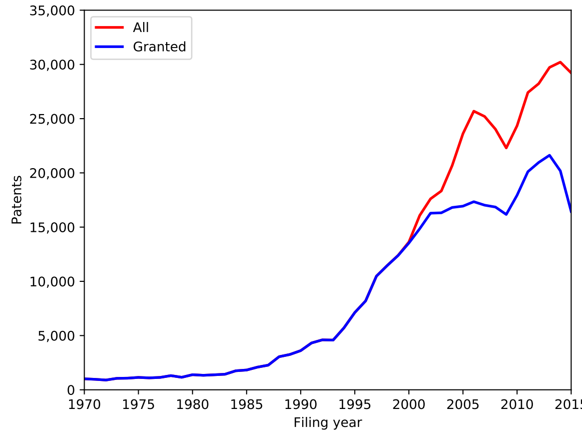
Cloud



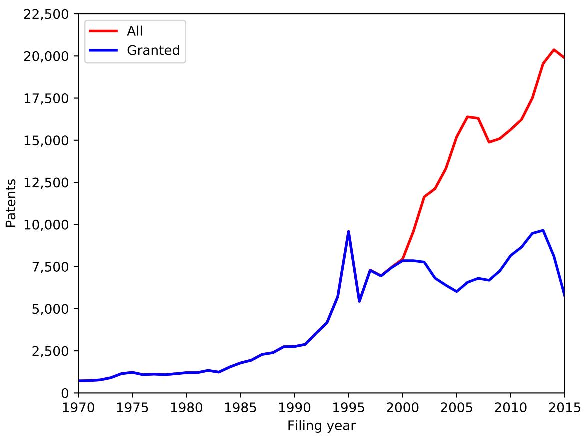
Self-driving cars



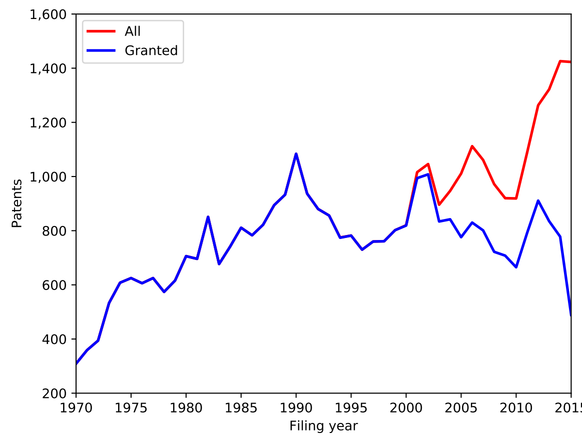
Semiconductors



Pharmaceuticals



Internal combustion engines

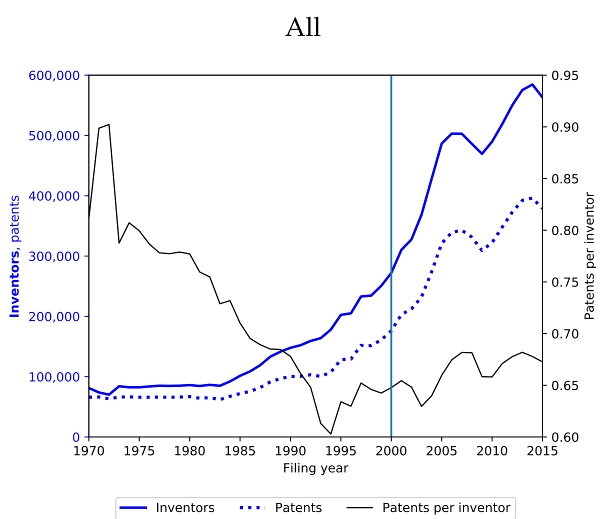


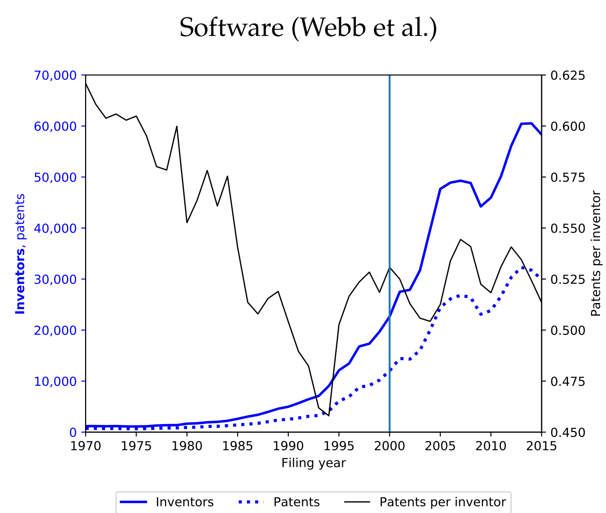
4 Inventors

We calculate the number of patents per inventor for each technology class in each year. Specifically, we divide the number of distinct patent application numbers by the number of distinct harmonized inventor names. Each figure plots, for a given technology class, the patent count, the inventor count, and (on the right-hand-side axis) the patents per inventor. For each technology class, we plot these numbers for all patent applications. The patterns for granted patents are very similar.

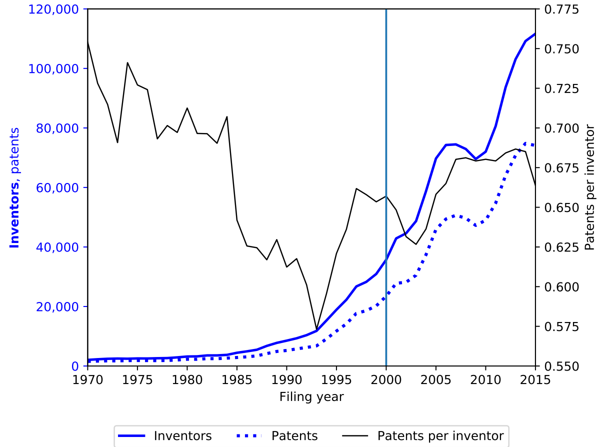
The most striking finding is the rapid fall in patents per inventor once each field opens up. It appears new technology fields are characterized by rapid bursts of innovation from a relatively small group of inventors followed by a fanning out and slowing down in per-inventor productivity. Of course, we cannot control for patent quality – later patents may be more valuable – but typically earlier patents in new fields are more highly cited, so the downward slope in research productivity within fields may even be stronger than depicted, after accounting for patent quality. This is something that is discussed at length in Bloom et al. (2017), reflecting a general trend towards "ideas getting harder to find" over time.

One interesting dynamic seen in many of the technologies is a dip in patents per inventor around 1994. The year 1995 saw one of the most dramatic recent shifts of patent policy, in which the adoption of the Uruguay Round of the Agreement on Trade-Related Aspects of Intellectual Property Rights led to a shift in U.S. awards from seventeen years from the award date to twenty years from the filing date. Patent applicants in the months before June 1995 allowed inventors the longer of the two forms of protection, a potentially valuable consideration. One possibility is that this policy shift triggered filings by a broader range of inventors, leading to a fall in the number of applications per inventor.

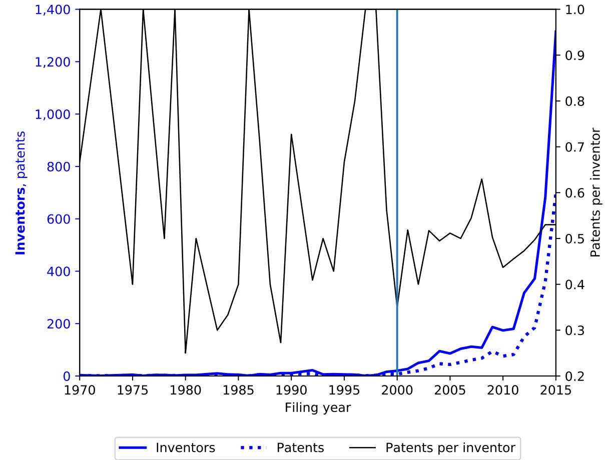




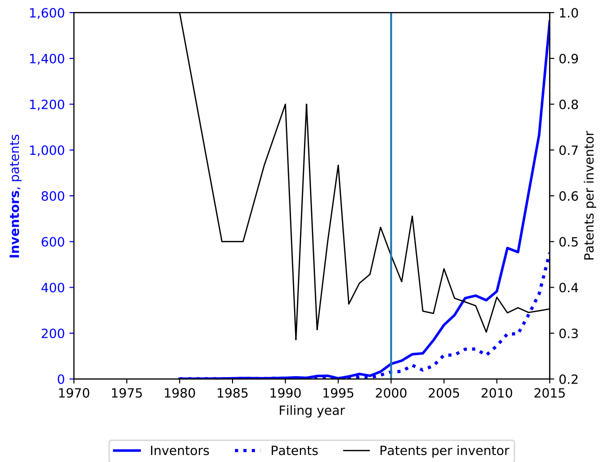
Smartphones



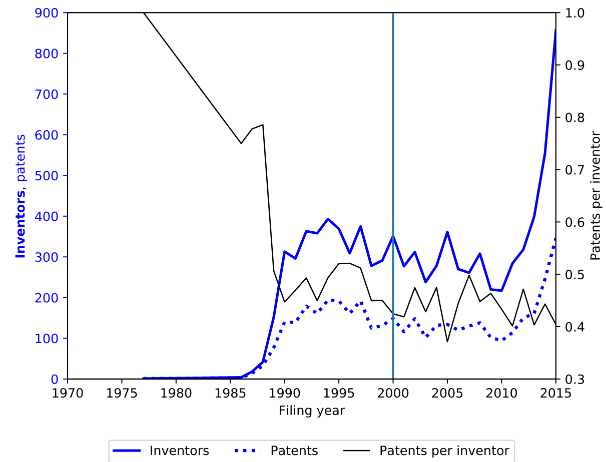
Drones



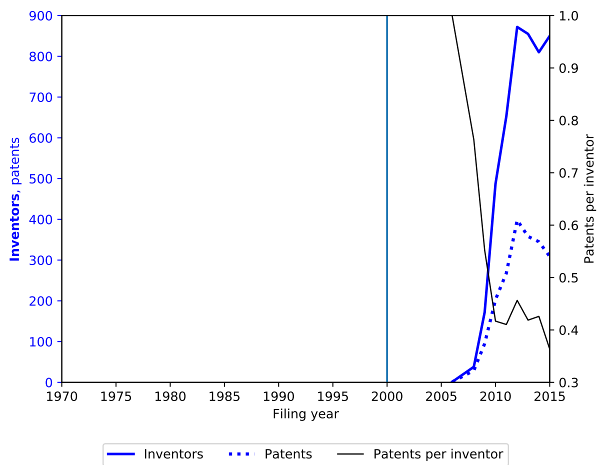
Machine learning



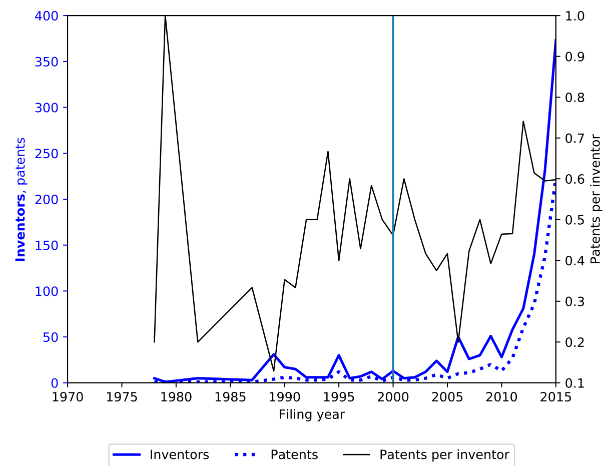
Neural networks



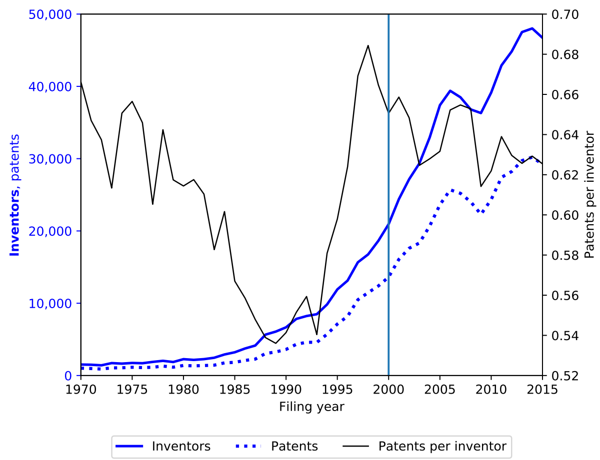
Cloud



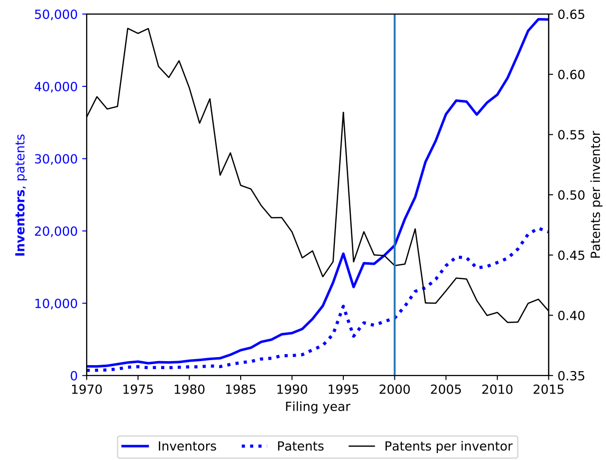
Self-driving cars



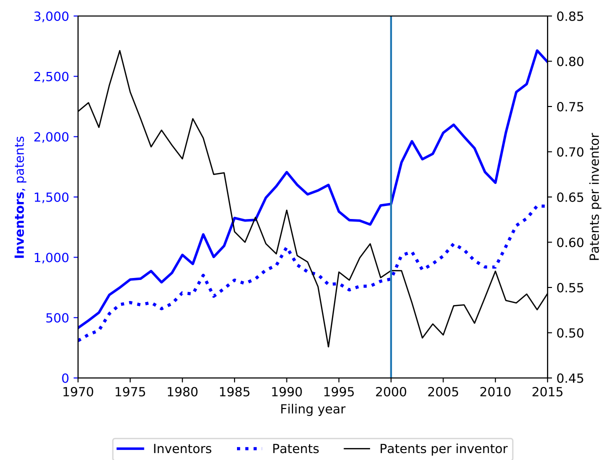
Semiconductors



Pharmaceuticals



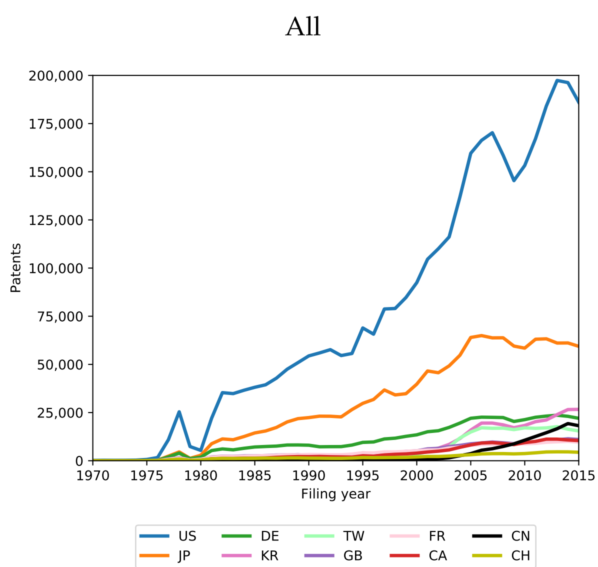
Internal combustion engines

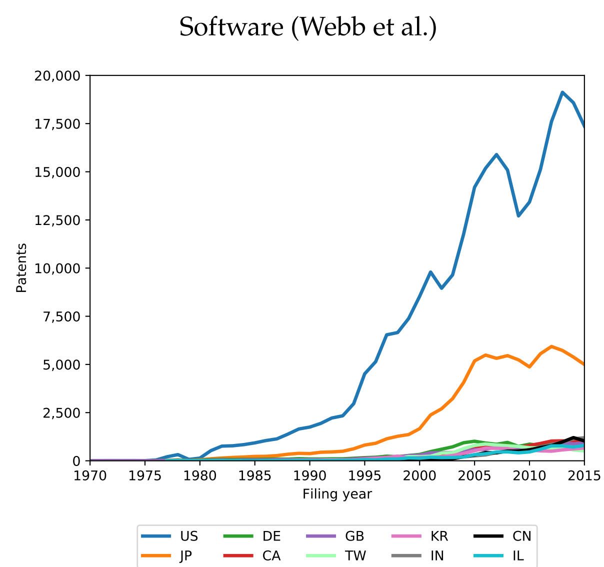


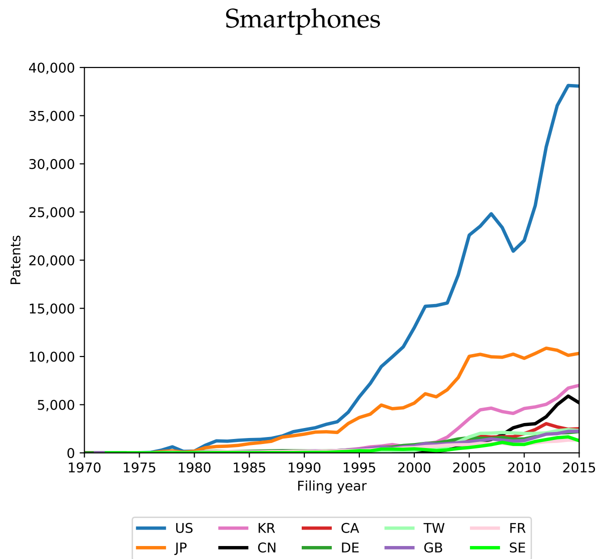
5 Geographic distribution of inventors

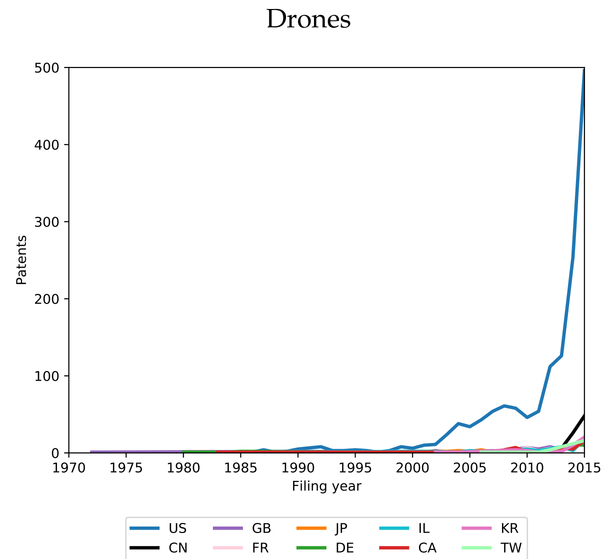
We also considered changes in the geographic distribution of inventors. Below, each figure plots, for a given technology class, the number of patents filed at the USPTO each year by inventors in a given country. For each figure, only the top ten countries by total patent application count over the period 1970-2015 are displayed. (The results for patent grants look very similar.) A few basic results jump out. First, the dominance of the US, which is not surprising given its technological lead and the home bias in analyzing US patent data. Second, the particular dominance of the US in some of the most recent technologies, such as cloud, machine learning, and neural networks (noting the Japanese surge and fall in the 1990s). In comparison, in semi-conductors and combustion engines the US is tracked closely by Japan.

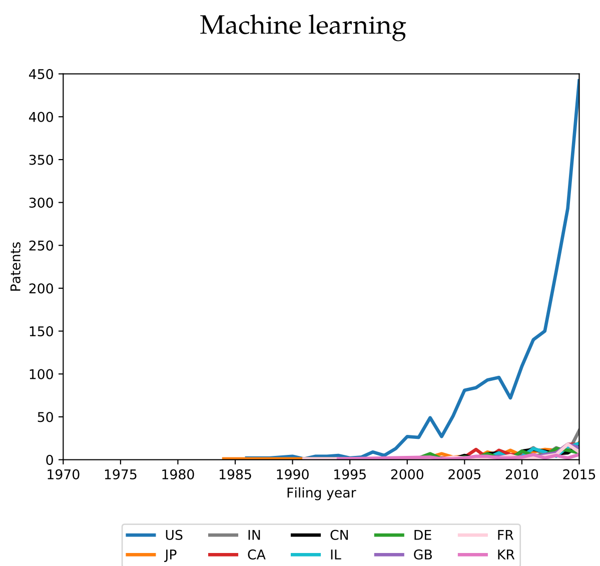
Interestingly, the share of Chinese patentees remains very modest in most technologies, with the exceptions of drones and smartphones, which have seen recent rapid rises.

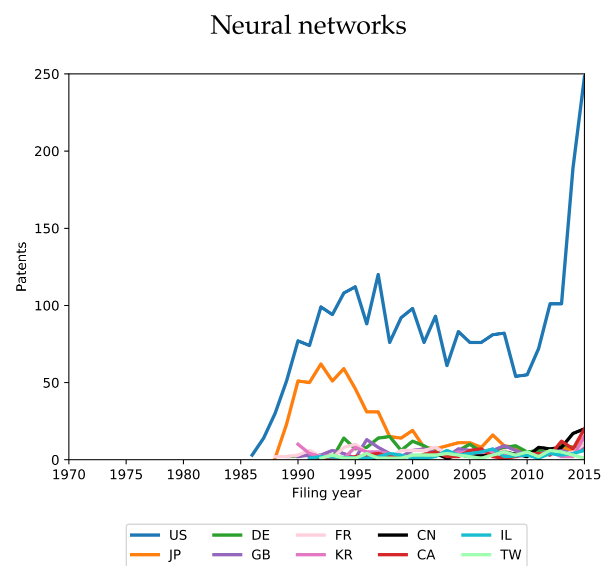


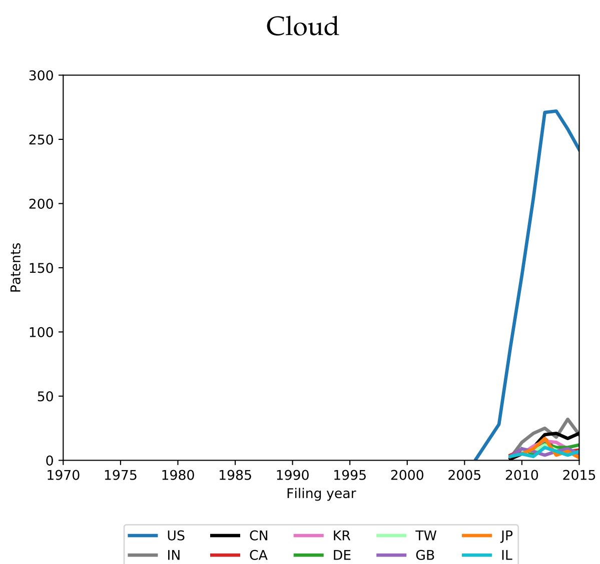


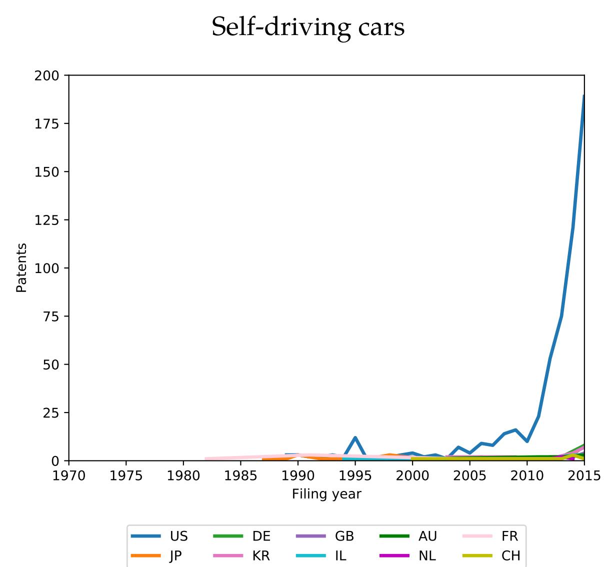

Smartphones
Patents
Filing year
US JP KR CN CA DE TW GB FR SE


Drones
Patents
Filing year
US CN GB FR JP DE IL CA KR TW

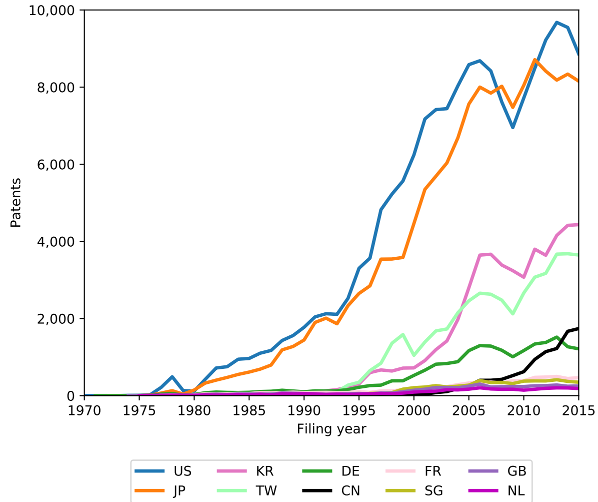

Machine learning
Patents
Filing year
US JP IN CA CN IL DE GB FR KR


Neural networks
Patents
Filing year
US JP DE GB FR KR CN CA IL TW

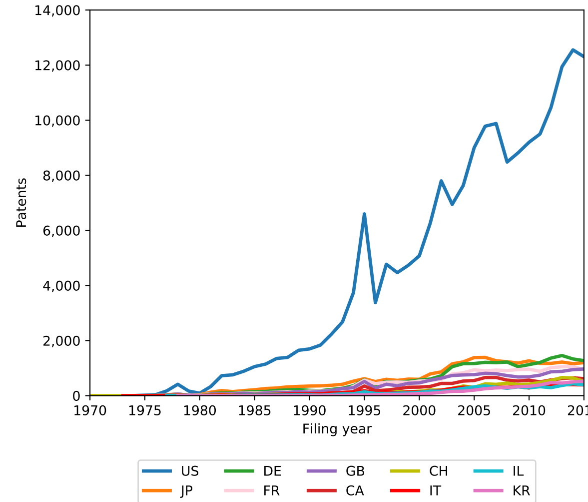

Cloud
Patents
Filing year
US IN CN CA KR DE TW GB JP IL


Self-driving cars
Patents
Filing year
US JP DE KR GB IL AU NL FR CH

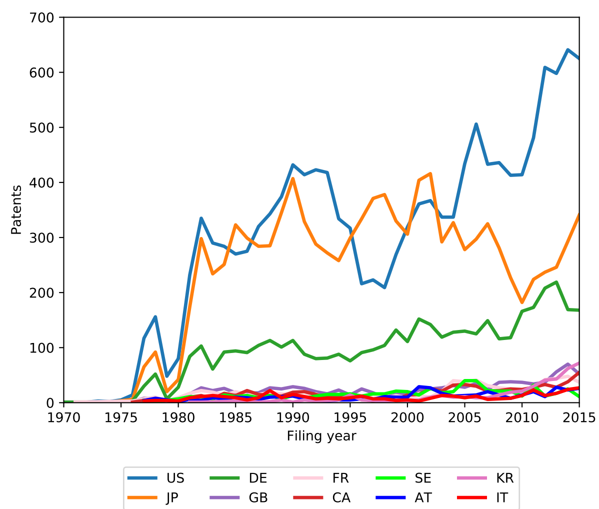
Semiconductors



Pharmaceuticals



Internal combustion engines



6 Top assignees (granted patents)

We also considered changes in the identities of the top assignees in each technology class over time. Each table below displays, for a given technology class, the total number of patents granted to each of the top 10 assignees by total patents granted over the periods 1970-99 (left panel) and 2000-15 (right panel).

What is striking here is how little change there is between patents applied for between 1970 and 1999 and between 2000 and 2015. Consistent with Bessen and Hunt (2007), the applications are dominated by some of the largest U.S., Japanese, and (in the more recent sample) Korean technology firms that have traditionally been active patent filers. As we move into some of the newer technologies, however, we see a more diverse array of patentees, including some newer software-focused firms, Chinese firms, and individual inventors.

Note that these results do not account for corporate ownership structure. For example, when counting any company's patents, we consider only the parent company and do not account for patents owned by subsidiaries. We are working to address this issue.

Table 2: All

1970-99			**2000-15**		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
1	IBM	34,564	1	IBM	83,530
2	Canon KK	24,974	2	Samsung Electronics Co Ltd	60,099
3	Gen Electric	24,747	3	Canon KK	41,470
4	Hitachi Ltd	21,935	4	Sony Corp	30,561
5	Mitsubishi Electric Corp	16,202	5	Toshiba KK	28,850
6	NEC Corp	16,188	6	Intel Corp	26,659
7	Eastman Kodak Co	16,132	7	Microsoft Corp	24,781
8	Philips Corp	15,853	8	Fujitsu Ltd	21,772
9	Toshiba KK	15,780	9	Gen Electric	21,166
10	Motorola Inc	15,489	10	LG Electronics Inc	21,109

Table 3: Software (Webb et al.)

1970-99			2000-15		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
1	IBM	6,365	1	IBM	24,085
2	Microsoft Corp	1,515	2	Microsoft Corp	7,312
3	Sun Microsystems Inc	1,211	3	Sony Corp	7,033
4	Hitachi Ltd	1,133	4	Canon KK	4,767
5	Hewlett Packard Co	1,132	5	Google Inc	4,569
6	Intel Corp	1,074	6	Hewlett Packard Development Co	3,227
7	Fujitsu Ltd	838	7	Fujitsu Ltd	3,145
8	Canon KK	817	8	Intel Corp	2,883
8	Sony Corp	817	9	Hitachi Ltd	2,839
10	Toshiba KK	680	10	NEC Corp	2,392

Table 4: Smartphones

1970-99			2000-15		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
1	Canon KK	6,074	1	Samsung Electronics Co Ltd	20,525
2	Sony Corp	5,512	2	IBM	16,912
3	Motorola Inc	4,850	3	Sony Corp	13,909
4	IBM	4,297	4	Qualcomm Inc	13,371
5	NEC Corp	4,269	5	Canon KK	13,277
6	Matsushita Electric Ind Co Ltd	3,191	6	LG Electronics Inc	12,700
7	Philips Corp	3,189	7	Cisco Tech Inc	9,325
8	Fujitsu Ltd	2,989	8	Microsoft Corp	8,494
9	Lucent Technologies Inc	2,884	9	Intel Corp	8,300
10	Samsung Electronics Co Ltd	2,615	10	Fujitsu Ltd	8,256

Table 5: Drones

1970-99			2000-15		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
1	United Technologies Corp	9	1	Sz Dji Technology Co Ltd	103
2	Freewing Aerial Robotics Corp	7	2	Boeing Co	77
3	Aerovironment Inc	4	3	Amazon Tech Inc	73
3	Messerschmitt Boelkow Blohm	4	4	Aerovironment Inc	54
5	E Systems Inc	3	5	Google Inc	37
6	Sikorsky Aircraft Corp	2	5	Lockheed Corp	37
6	Marconi Gec Ltd	2	7	Honeywell Int Inc	36
6	Nasa	2	8	IBM	26
6	Science Applic Int Corp	2	9	Insitu Inc	24
6	Boeing Co	2	10	US Navy	23

Table 6: Machine learning

1970-99			2000-15		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
1	IBM	8	1	Microsoft Corp	153
2	Lucent Technologies Inc	5	2	IBM	117
3	Thomson Consumer Electronics	4	3	Google Inc	66
3	Microsoft Corp	4	4	Microsoft Technology Licensing LLC	53
5	Barnhill Technologies LLC	2	5	Amazon Tech Inc	37
6	GTE Laboratories Inc	1	6	Yahoo Inc	30
6	Lexis Nexis Group	1	7	Intel Corp	28
6	Akzo Nv	1	8	Health Discovery Corp	23
6	Nat Semiconductor Corp	1	9	Symantec Corp	22
6	Nestor Inc	1	9	Cisco Tech Inc	22

Table 7: Neural networks

1970-99			2000-15		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
1	IBM	77	1	IBM	112
2	Siemens AG	50	2	Google Inc	78
3	Hitachi Ltd	49	3	Microsoft Corp	48
4	Mitsubishi Electric Corp	46	4	Qualcomm Inc	23
5	Motorola Inc	39	4	Microsoft Technology Licensing LLC	23
5	US Army	39	4	Modha Dharmendra S	23
7	Toshiba KK	38	7	Sony Corp	19
8	Matsushita Electric Ind Co Ltd	37	7	Siemens AG	19
9	Fujitsu Ltd	27	9	Nec Lab America Inc	16
10	Sharp KK	26	9	Samsung Electronics Co Ltd	16

Table 8: Cloud

1970-99			2000-15		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
nan	Nan	nan	1	IBM	308
nan	Nan	nan	2	Red Hat Inc	51
nan	Nan	nan	3	Microsoft Technology Licensing LLC	50
nan	Nan	nan	4	Microsoft Corp	43
nan	Nan	nan	5	Oracle Int Corp	35
nan	Nan	nan	6	Google Inc	33
nan	Nan	nan	7	Verizon Patent & Licensing Inc	26
nan	Nan	nan	8	Dawson Christopher J	24
nan	Nan	nan	9	Vmware Inc	23
nan	Nan	nan	10	Amazon Tech Inc	22

Table 9: Self-driving cars

1970-99			2000-15		
Rank	Assignee	Patents	Rank	Assignee	Patents
1	Caterpillar Inc	17	1	Google Inc	147
2	Minolta Co Ltd	5	2	Ford Global Tech LLC	35
3	Nissan Motor	4	3	Waymo LLC	34
4	Trimble Navigation Ltd	3	4	Toyota Motor Eng & Mfg North America Inc	27
4	US Navy	3	5	Uber Tech Inc	15
6	Dyson Ltd	2	6	GM Global Tech Operations Inc	14
6	Commissariat Energie Atomique	2	6	Zhu Jiajun	14
6	Thomson Csf	2	8	Zoox Inc	13
9	Transitions Research Corp	1	8	Ferguson David I	13
9	Sara Avitzour	1	8	IBM	13

Table 10: Semiconductors

1970-99			2000-15		
Rank	Assignee	Patents	Rank	Assignee	Patents
1	IBM	5,335	1	IBM	13,664
2	Micron Technology Inc	3,461	2	Samsung Electronics Co Ltd	12,356
3	NEC Corp	3,438	3	Micron Technology Inc	10,727
4	Texas Instruments Inc	3,379	4	Toshiba KK	8,368
5	Toshiba KK	3,191	5	Semiconductor Energy Lab	8,280
6	Mitsubishi Electric Corp	3,066	6	Taiwan Semiconductor Mfg	7,101
7	Hitachi Ltd	2,836	7	Infineon Technologies AG	4,849
8	Motorola Inc	2,617	8	Intel Corp	4,427
9	Advanced Micro Devices Inc	2,205	9	Samsung Display Co Ltd	4,337
10	Fujitsu Ltd	2,090	10	Sony Corp	3,888

Table 11: Pharmaceuticals

1970-99			2000-15		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
1	Oreal	1,945	1	Oreal	1,616
2	Procter & Gamble	1,275	2	Univ California	1,447
3	Merck & Co Inc	1,131	3	Genentech Inc	1,130
4	Univ California	1,043	4	Novartis AG	1,088
5	Lilly Co Eli	1,020	5	US Health	1,075
6	US Health	789	6	Procter & Gamble	978
7	Colgate Palmolive Co	718	7	Allergan Inc	855
8	Alza Corp	704	8	Univ Texas	793
9	Smithkline Beecham Corp	608	9	Pfizer	588
10	Warner Lambert Co	576	10	Merck Sharp & Dohme	562

Table 12: Internal combustion engines

1970-99			2000-15		
Rank	**Assignee**	**Patents**	**Rank**	**Assignee**	**Patents**
1	Honda Motor Co Ltd	903	1	Ford Global Tech LLC	994
2	Toyota Motor Co Ltd	784	2	Honda Motor Co Ltd	799
3	Yamaha Motor Co Ltd	738	3	Toyota Motor Co Ltd	635
4	Sanshin Kogyo KK	661	4	Caterpillar Inc	393
5	Nissan Motor	565	5	GM Global Tech Operations Inc	342
6	Gen Motors Corp	554	6	Nissan Motor	285
7	Bosch Gmbh Robert	510	7	Bosch Gmbh Robert	230
8	Outboard Marine Corp	382	8	Yamaha Motor Co Ltd	210
9	Mazda Motor	331	9	Borgwarner Inc	193
10	Daimler Benz AG	297	10	Hyundai Motor Co Ltd	187

7 Within/between decomposition

For each technological field, we decompose the growth in patent applications between the periods 1970-1989 and 2000-present into growth in patent applications by current inventors (within) versus new inventors (between).

We write

$$P_t = \sum_i s_{i,t}\omega_{i,t},$$

where $s_{i,t}$ is inventor i's share of **all** patent applications in period t, where coauthoured patents count fractionally so that $\sum_i s_{i,t} = 1$; and $\omega_{i,t}$ is the fraction of inventor i's patents that are patents in the given technological field.

We decompose[5] the change in the share of patenting in a given technological field between two periods as

$$\begin{aligned}
P_t - P_{t-1} &= \sum_i s_{i,t}\omega_{i,t} - \sum_i s_{i,t-1}\omega_{i,t-1} \\
&= \sum_{i\in C} s_{i,t-1}\left(\omega_{i,t} - \omega_{i,t-1}\right) \text{ (within)} \\
&\quad + \sum_{i\in C} \left(s_{i,t} - s_{i,t-1}\right)\left(\omega_{i,t-1} - P_{t-1}\right) \text{ (between)} \\
&\quad + \sum_{i\in C} \left(s_{i,t} - s_{i,t-1}\right)\left(\omega_{i,t} - \omega_{i,t-1}\right) \text{ (cross)} \\
&\quad + \sum_{i\in ENT} s_{i,t}\left(\omega_{i,t} - P_{t-1}\right) \text{ (entry)} \\
&\quad - \sum_{i\in EXIT} s_{i,t-1}\left(\omega_{i,t-1} - P_{t-1}\right) \text{ (exit)},
\end{aligned}$$

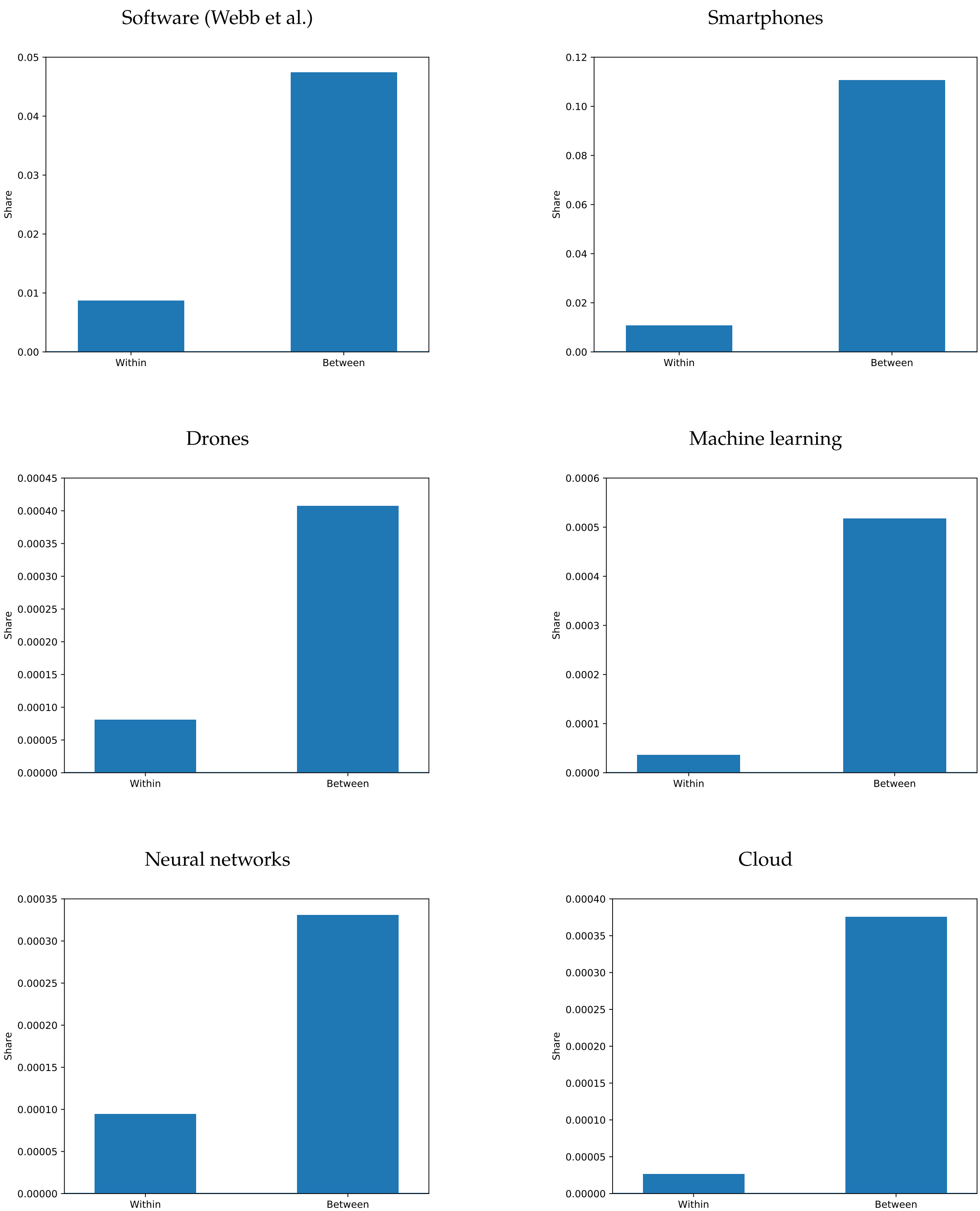
where C is the set of continuing inventors, ENT the set of entrants (who patent in period t but not $t-1$), and $EXIT$ the set of exiters (who patent in period $t-1$ but not t).

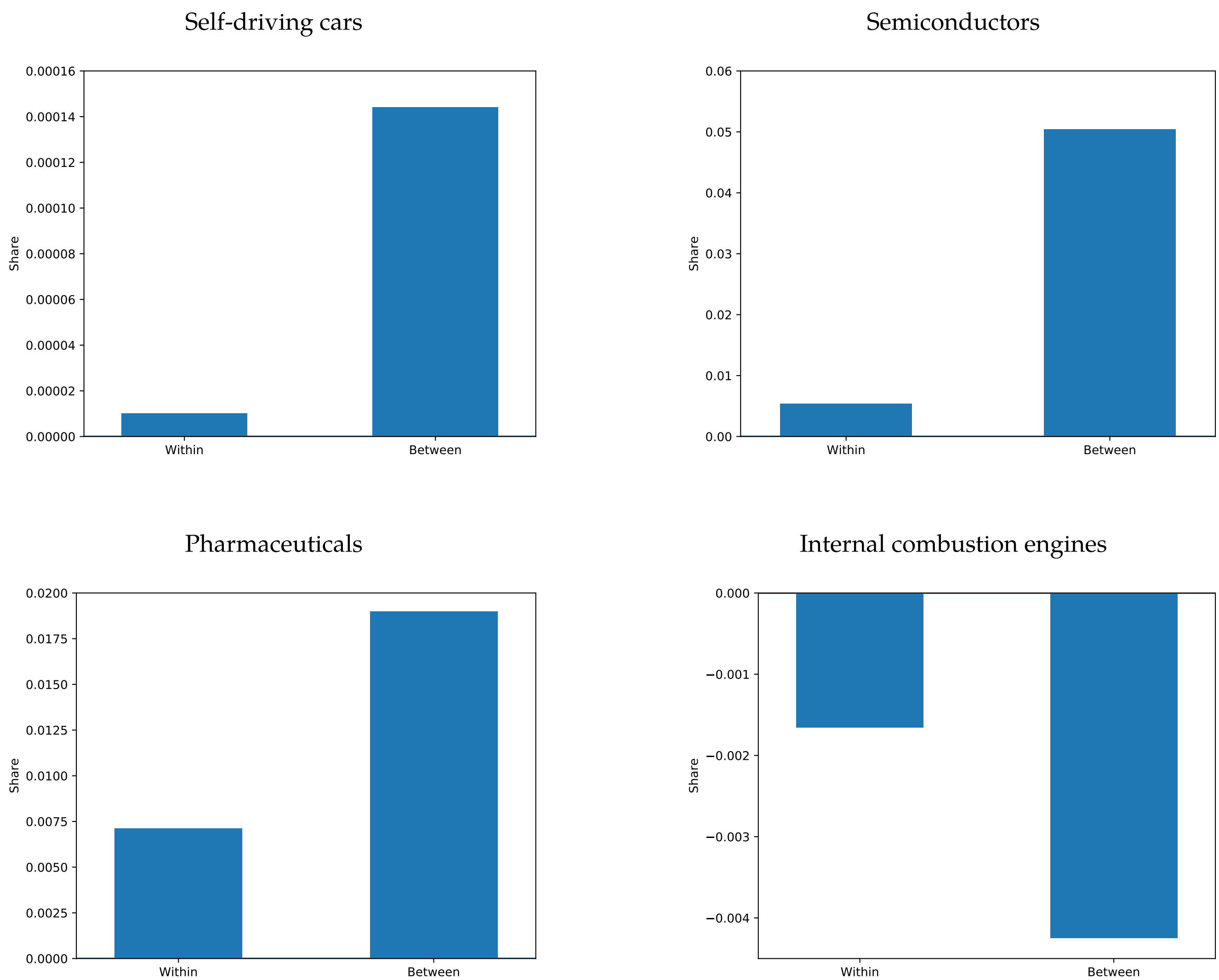
To simplify the presentation, we amalgamate between, cross, entry, and exit into between, i.e.,

$$\text{between} = P_t - P_{t-1} - \text{within}.$$

The results are displayed in the figures below. Most of the new, rapidly growing technologies are almost all between, whereas the older ones — pharma and engines — are heavily within. Thus, in established technological domains, most of the changes in patenting seem to come from dominant incumbent firms, whereas patenting in more cutting-edge areas appears to be dominated by new entrants.

[5]This decomposition is as defined in Haltiwanger (1997). It differs from the Baily, Hulten and Campbell (1992) decomposition in some important respects, as detailed in the appendix of Haltiwanger (1997).


Software (Webb et al.)
Share
0.00
0.01
0.02
0.03
0.04
0.05
Within
Between
Smartphones
Share
0.00
0.02
0.04
0.06
0.08
0.10
0.12
Within
Between
Drones
Share
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0.00030
0.00035
0.00040
0.00045
Within
Between
Machine learning
Share
0.0000
0.0001
0.0002
0.0003
0.0004
0.0005
0.0006
Within
Between
Neural networks
Share
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0.00030
0.00035
Within
Between
Cloud
Share
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0.00030
0.00035
0.00040
Within
Between


Self-driving cars
0.00016
0.00014
0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
Share
Within
Between
Semiconductors
0.06
0.05
0.04
0.03
0.02
0.01
0.00
Share
Within
Between
Pharmaceuticals
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
Share
Within
Between
Internal combustion engines
0.000
−0.001
−0.002
−0.003
−0.004
Share
Within
Between

8 Conclusion

This paper has identified some stylized facts about the growth of patenting in software, cloud computing, artificial intelligence, and related technologies in the twentieth-first century. It highlights the continuing growth of this activity, as well as the continuing dominance of inventors in large U.S., Japanese, and Korean companies. The paper also documents a shared pattern in the evolution of the mixture and productivity of inventors over time.

The analysis raises a number of unanswered issues. Foremost among these is understanding the value of these awards. This question has two dimensions. The first, which we are currently investigating, relates to the private returns to these awards. In particular, we are examining the market reaction to awards before and after some of the key judicial decisions relating to software patents, as well as the different changes in market value around the time of these decisions.

The second, more difficult question was alluded to in the introduction: the impact of these awards on social welfare more generally. To what extent can the seemingly deleterious effects identified by Bessen and Hunt (2007) be corroborated? Do these patterns hold across all aspects of software, or just in certain segments? These issues, while challenging, will reward scrutiny in the years to come.

References

Acemoglu, Daron, Simon Johnson, and James A Robinson, "Reversal of Fortune: Geography and Institutions in the Making of the Modern World Income Distribution," *The Quarterly Journal of Economics*, 2002, *117* (4), 1231–1294.

Allison, John R and Emerson H Tiller, *Internet Business Method Patents*, Washington, DC: The National Academies Press, 2003.

Baily, Martin Neil, Charles Hulten, and David Campbell, "Productivity Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity. Microeconomics*, 1992, *1992*, 187–267.

Bessen, James and Robert M Hunt, "An Empirical Look at Software Patents," *Journal of Economics & Management Strategy*, March 2007, *16* (1), 157–189.

Bloom, Nicholas, Charles I Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," September 2017.

Cohen, Lauren, Umit Gurun, and Scott Duke Kominers, "Patent trolls: Evidence from targeted firms," Technical Report, National Bureau of Economic Research 2014.

Dorn, David, Gordon H Hanson, Gary Pisano, Pian Shu et al., "Foreign competition and domestic innovation: Evidence from US patents," Technical Report, National Bureau of Economic Research 2016.

Ewing, Tom and Robin Feldman, "The giants among us," *Stan. Tech. L. Rev.*, 2012, p. 1.

Galasso, Alberto, Mark Schankerman, and Carlos J Serrano, "Trading and enforcing patent rights," *The RAND Journal of Economics*, 2013, *44* (2), 275–312.

Graham, Stuart JH and David C Mowery, "Intellectual property protection in the US software industry," *Patents in the Knowledge-based Economy*, 2003, *219*, 231.

Griliches, Zvi, "Patent statistics as economic indicators: a survey," in "R&D and Productivity: The Econometric Evidence," University of Chicago Press, 1998, pp. 287–343.

Hall, Bronwyn H, Adam Jaffe, and Manuel Trajtenberg, "Market value and patent citations," *RAND Journal of economics*, 2005, pp. 16–38.

— **and Megan MacGarvie**, "The private value of software patents," *Research Policy*, 2010, *39* (7), 994–1009.

Haltiwanger, John, "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building From Microeconomic Evidence," *Review of the Federal Reserve Bank of St. Louis*, 1997, *79* (3), 55–78.

Jaffe, Adam B and Manuel Trajtenberg, *Patents, citations, and innovations: A window on the knowledge economy*, MIT press, 2002.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman, "Technological innovation, resource allocation, and growth," *The Quarterly Journal of Economics*, 2017, *132* (2), 665–712.

Lanjouw, Jean O, Ariel Pakes, and Jonathan Putnam, "How to count patents and value intellectual property: The uses of patent renewal and application data," *The Journal of Industrial Economics*, 1998, *46* (4), 405–432.

Lerner, Josh and Amit Seru, "The Use and Abuse of Patent Data: Issues for Corporate Finance and Beyond," Technical Report 24053, National Bureau of Economic Research 2017.

Magerman, Tom, Bart Van Looy, and Xiaoyan Song, "Data Production Methods for Harmonized Patent Statistics: Patentee Name Harmonization," Technical Report, Eurostat 2006.

Mann, Ronald J, "Do patents facilitate financing in the software industry," *Tex L. Rev.*, 2004, *83*, 961.

Marco, Alan C, Amanda F Myers, Stuart Graham, Paul D'Agostino, and Kirsten Apple, "The USPTO Patent Assignment Dataset: Descriptions and Analysis," Technical Report Working Paper No. 2015-2, USPTO 2015.

Porta, Rafael La, Florencio Lopez de Silanes, Andrei Shleifer, and Robert W Vishny, "Law and Finance," *Journal of Political Economy*, 1998, *106* (6), 1113–1155.

Scotchmer, Suzanne, *Innovation and Incentives*, MIT Press, 2004.

Shalem, Roy and Manuel Trajtenberg, "Software patents, inventors, and mobility," Technical Report, SSRN 2009.

Webb, Michael et al., "What Does Artificial Intelligence Do?," February 2018.

A Appendices

A.1 Software definitions

We use two sets of criteria for classifying software patents. The first is taken from Bessen and Hunt (2007). The authors of that paper note: "our concept of software patent involves a logic algorithm for processing data that is implemented via stored instructions; that is, the logic is not 'hard-wired.' These instructions could reside on a disk or other storage medium or they could be stored in 'firmware,' that is, a read-only memory, as is typical of embedded software. But we want to exclude inventions that do not use software as part of the invention. For example, some patents reference off-the-shelf software used to determine key parameters of the invention; such uses do not make the patent a software patent." Bessen and Hunt (2007) manually selected an algorithm, described in Table 1, intended to capture this sense of software, and performed various validation checks.

We also use the criteria for classifying software patents developed in Webb et al. (2018). This algorithm is based on that in Bessen and Hunt (2007), but uses titles and abstracts instead of the specification – also known as the 'technical description' field. This is in order to include patents filed at patent offices other than the USPTO, for which the technical description field is generally lacking in the patent database. Of course, in this paper, we are restricting to patents filed at the USPTO.

As can be seen from the figures in Section 3, the Bessen and Hunt (2007) definition is far more expansive, encompassing almost 5 times as many patents in recent years as the Webb et al. (2018) definition. Note also that using the Bessen and Hunt (2007) defintion, there are almost 5 times as many software patents as semiconductor patents. That said, the two definitions display very similar patterns in all our results.

A.2 CPC assignments

We consider how the CPC codes to which patent applications in each technology class were assigned changed between the 1970s/80s period and the 2000s/2010s period. The tables below display the 4-digit CPC code prefixes that were most frequently assigned to the patent applications in each technology class, ranked by frequency in the second period.

Table 13: All

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
H01L	Semiconductor devices; electric solid state devices not otherwise provided for	129,149	0.027	6,555,914	0.097
G06F	Electrical digital data processing	34,548	0.007	3,693,956	0.055
H04L	Transmission of digital information, e.g. telegraphic communication	14,041	0.003	2,913,666	0.043
H04N	Pictorial communication, e.g. television	43,985	0.009	2,496,628	0.037
A61B	Diagnosis; surgery; identification	44,447	0.009	2,472,649	0.036

Table 14: Software (Webb et al.)

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
G06F	Electrical digital data processing	8,398	0.113	1,029,312	0.233
H04N	Pictorial communication, e.g. television	2,388	0.032	559,174	0.126
H04L	Transmission of digital information, e.g. telegraphic communication	866	0.012	533,809	0.121
G06Q	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for	1,045	0.014	303,479	0.069
A61B	Diagnosis; surgery; identification	1,692	0.023	146,355	0.033

Table 15: Smartphones

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
H04L	Transmission of digital information, e.g. telegraphic communication	14,041	0.102	2,913,666	0.240
H04N	Pictorial communication, e.g. television	43,985	0.319	2,496,628	0.206
H04W	Wireless communications networks	1,870	0.014	1,396,075	0.115
G06F	Electrical digital data processing	2,448	0.018	986,466	0.081
H04M	Telephonic communication	9,257	0.067	526,742	0.043

Table 16: Drones

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
B64C	Aeroplanes; helicopters	197	0.801	34,747	0.464
G05D	Systems for controlling or regulating non-electric variables	2	0.008	5,953	0.079
B64D	Equipment for fitting in or to aircraft; flying suits; parachutes; arrangements or mounting of power plants or propulsion transmissions in aircraft	9	0.037	4,872	0.065
G08G	Traffic control systems	nan	nan	4,029	0.054
H04N	Pictorial communication, e.g. television	1	0.004	1,913	0.026

Table 17: Machine learning

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
G06F	Electrical digital data processing	nan	nan	8,332	0.231
G06N	Computer systems based on specific computational models	4	0.133	5,241	0.145
G06K	Recognition of data; presentation of data; record carriers; handling record carriers	3	0.100	3,954	0.109
H04L	Transmission of digital information, e.g. telegraphic communication	5	0.167	3,536	0.098
G06T	Image data processing or generation, in general	nan	nan	2,599	0.072

Table 18: Neural networks

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
G06N	Computer systems based on specific computational models	139	0.421	6,531	0.212
G06K	Recognition of data; presentation of data; record carriers; handling record carriers	41	0.124	3,546	0.115
G06F	Electrical digital data processing	9	0.027	2,757	0.089
A61B	Diagnosis; surgery; identification	7	0.021	2,397	0.078
G06T	Image data processing or generation, in general	4	0.012	2,085	0.068

Table 19: Cloud

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
H04L	Transmission of digital information, e.g. telegraphic communication	nan	nan	12,386	0.409
G06F	Electrical digital data processing	nan	nan	11,680	0.386
G06Q	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for	nan	nan	1,378	0.046
H04N	Pictorial communication, e.g. television	nan	nan	913	0.030
H04W	Wireless communications networks	nan	nan	695	0.023

Table 20: Self-driving cars

CPC code	CPC code description	1970s/80s		2000s/10s	
		Count	Share	Count	Share
G05D	Systems for controlling or regulating non-electric variables	21	0.344	4,702	0.227
B60W	Conjoint control of vehicle sub-units of different type or different function; control systems specially adapted for hybrid vehicles; road vehicle drive control systems for purposes not related to the control of a particular sub-unit	3	0.049	3,735	0.180
G08G	Traffic control systems	1	0.016	1,600	0.077
G01C	Measuring distances, levels or bearings; surveying; navigation; gyroscopic instruments; photogrammetry or videogrammetry	nan	nan	1,064	0.051
G01S	Radio direction-finding; radio navigation; determining distance or velocity by use of radio waves; locating or presence-detecting by use of the reflection or reradiation of radio waves; analogous arrangements using other waves	11	0.180	1,001	0.048

Table 21: Semiconductors

CPC code	CPC code description	1970s/80s		2000s/10s	
		Count	Share	Count	Share
H01L	Semiconductor devices; electric solid state devices not otherwise provided for	129,149	0.657	6,555,914	0.741
H05K	Printed circuits; casings or constructional details of electric apparatus; manufacture of assemblages of electrical components	7,498	0.038	232,455	0.026
Y10T	Technical subjects covered by former us classification	6,884	0.035	130,387	0.015
G11C	Static stores	2,755	0.014	126,675	0.014
C23C	Coating metallic material; coating material with metallic material; surface treatment of metallic material by diffusion into the surface, by chemical conversion or substitution; coating by vacuum evaporation, by sputtering, by ion implantation or by chemical vapour deposition, in general	1,794	0.009	90,605	0.010

Table 22: Pharmaceuticals

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
A61K	Preparations for medical, dental, or toilet purposes	58,388	0.446	2,137,157	0.500
C07K	Peptides	6,822	0.052	442,875	0.104
C12N	Microorganisms or enzymes; compositions thereof; propagating, preserving or maintaining microorganisms; mutation or genetic engineering; culture media	2,704	0.021	297,071	0.070
C07D	Heterocyclic compounds	6,776	0.052	273,479	0.064
G01N	Investigating or analysing materials by determining their chemical or physical properties	1,308	0.010	136,867	0.032

Table 23: Internal combustion engines

		1970s/80s		2000s/10s	
CPC code	**CPC code description**	**Count**	**Share**	**Count**	**Share**
F02B	Internal-combustion piston engines; combustion engines in general	25,392	0.330	85,086	0.228
F02D	Controlling combustion engines	4,375	0.057	63,136	0.170
F02M	Supplying combustion engines in general, with combustible mixtures or constituents thereof	6,866	0.089	44,232	0.119
Y02T	Climate change mitigation technologies related to transportation	6,293	0.082	36,481	0.098
F01N	Gas-flow silencers or exhaust apparatus for machines or engines in general; gas-flow silencers or exhaust apparatus for internal combustion engines	3,023	0.039	22,992	0.062