USPTO Economic Working Paper

Office of Chief Economist, U.S. Patent and Trademark Office

U.S. Department of Commerce

The USPTO Historical Patent Data Files

Two centuries of invention

Alan C. Marco
Michael Carley
Steven Jackson
Amanda F. Myers

Working Paper No. 2015-1
http://www.uspto.gov/economics

June 2015

Office of Chief Economist
U.S. Patent and Trademark Office
600 Dulany Street
Alexandria, VA 22314

The USPTO Historical Patent Data Files

Two centuries of innovation

Alan C. Marco
Michael Carley
Steven Jackson
Amanda F. Myers

U.S. Patent and Trademark Office*
Alexandria, Virginia

June 2015

Abstract

Patent classification systems are largely designed for administrative purposes, limiting their value for most research purposes. To address this deficiency, Hall, Jaffe, and Trajtenberg (2001) developed a higher-level classification for the National Bureau of Economic Research (NBER) Patent Citation Data File by aggregating U.S. Patent Classification (USPC) classes into six economically relevant technology categories (and 37 sub-categories) and classified granted patents accordingly. While this classification scheme has proved useful for researchers investigating US patent grants, comparable information on pending or abandoned patent applications has been unavailable for several reasons. We apply the NBER sub-categories to published and publicly-available unpublished patent applications as well as in-force and expired patents to create the *USPTO Historical Patent Data Files*, four research datasets containing time series and micro-level data by NBER sub-category. These new datasets comprise annual information on patent applications, patent grants, and patents-in-force dating back to 1840. Additionally, we provide information on the monthly stocks and flows of utility patent applications and grants from 1981 to 2014. Our hope is that researchers will make use of the data files for primary analysis or as controls for other projects. These data, for the first time, provide for detailed study of patent application disposal and the complex dynamics between new filings, pendency, and abandonment. Historical data enable researchers to put into context recent trends in patenting activity, litigation, and technological change.

Keywords: Intellectual Property Rights, Patents
JEL Classification Numbers: O3

* We thank Charlie DeGrazia for his excellent research assistance with this project. The views expressed are those of the individual authors and do not necessarily reflect official positions of the Office of Chief Economist or the U. S. Patent and Trademark Office.

1 Introduction

The difficulties and limitations associated with allocating patent data organized by complex classification schemes into economically relevant categories are well-recognized among patent researchers (Griliches, 1998). Patent classification systems were created to ensure that applications are routed to the appropriate examiners and to facilitate efficient prior art searching. Accordingly, classifications are primarily based on the technological and functional features of inventions and encompass the entire spectrum of subject matter that can be claimed. Such systems are highly complex and continuously updated to reflect evolving technology. The classification system maintained by the United States Patent and Trademark Office (PTO) is no exception.

The U.S. Patent Classification (USPC) system currently consists of more than 450 unique classes and 150,000 subclasses.[1] USPC classes provide no straightforward link to established product and industry classifications and are too numerous for most research purposes. Additionally, they are not hierarchical in any meaningful sense: classes 101 (printing) and 102 (ammunition and explosives) do not roll up into any common higher level class.[2] To address this, Hall, Jaffe, and Trajtenberg (2001) developed a hierarchical classification based on aggregating USPC classes into 37 (two-digit) sub-categories, which are further aggregated into six main technology categories (see **Table 2** for details). Hall, Jaffe and Trajtenberg applied their methodology to the National Bureau of Economic Research (NBER) Patent Citations Data File. The 1999 version of this data file represented a boon in the use of patent data by economists and other researchers. While any classification system is subject to some element of arbitrariness and subjectivity, the NBER classification has been widely adopted in the research community and provides a useful standard.[3]

Still, until now, NBER classification has only been applied to *patents* granted during a set time period, limiting its value among researchers examining patent applications and longer time horizons. In this paper, we address this limitation by applying NBER sub-categories to applications as well as in-force and expired patents dating back to 1840. We use this methodology to create the *USPTO Historical Patent Data Files*, which the PTO Office of Chief Economist is making publicly available on its homepage

[1] See: http://www.uspto.gov/patents/resources/classification/overview.pdf
[2] The International Patent Classification system and the newly developed Cooperative Patent Classification (CPC) system are hierarchical. However, these classification systems have not been consistently applied to US patents, much less to abandoned applications; in addition, the CPC scheme is revised on a monthly basis. For more information, see http://www.uspto.gov/patents/resources/classification/ and http://www.cooperativepatentclassification.org/
[3] The data file descriptive document (Hall, Jaffe, and Trajtenberg et al, 2001) has been cited over 2,000 times on Google Scholar to date.

(http://www.uspto.gov/economics) as part of ongoing efforts aimed at making data resources widely accessible for researchers.

We use the NBER Patent Citation Data File[4] in combination with PTO data sources to assign NBER sub-categories to all granted, pending, and abandoned applications filed between January 1981 and December 2014. While detailed application data cannot be released for unpublished applications,[5] the `historical_masterfile` file contains micro-level data for published applications, patent grants, and publicly-available unpublished applications. An unpublished application is not publicly available unless referenced by a patent or published application. Since the content of a referencing patent or published application is publicly available, the referenced unpublished application becomes public.[6] However, micro-level data for non-publicly-available unpublished applications are preserved in confidence and are not included in the `historical_masterfile`.[7] From the disaggregated data, we create several useful time-series data files. The `monthly` data file contains monthly time series of the stocks and flows of all publicly-available applications, published and unpublished, and granted patents from 1981 to 2014. The stocks include pending applications and, as we discuss in detail below, patents-in-force; flows include new applications, patent grants, and abandonments. We also create the `monthly_disposal` data file, which provides a more detailed view of the data in the `monthly` file by tracking prosecution outcomes for each monthly application filing cohort over time and across NBER sub-category.

We anticipate that these data files will be especially useful because information on the status of pending or abandoned applications can be difficult to obtain. Applications filed prior to November 29, 2000 were not published.[8] Even applications filed after that date are not published if they are abandoned early (less than 18 months from the earliest filing date) or if the applicant files a request for non-publication.[9] Even for published applications, information on pending or abandoned status is not readily available without searching for particular applications. Bulk downloads of application status are not currently available, and annual office-wide data on the number of incoming applications lack sufficient detail for most research purposes.[10] As a consequence, little is known outside the PTO about abandoned patent applications or the stock of pending

[4] See: https://sites.google.com/site/patentdataproject/Home, accessed 31 July 2014
[5] See 35 USC 122(a)
[6] See 37 CFR 1.14(a)
[7] See 35 USC 122(a)
[8] See 35 USC 122(b)
[9] See 37 CFR 1.213(a) and Hegde (2012)
[10] See http://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm for annual counts of applications since 1840. Additionally, given the fact that application numbers are sequential, some information can be inferred based on the application numbers of granted patents. However, any information based on technological classification is unavailable.

applications. These data, for the first time, provide for detailed analysis of the disposal of patent applications, both published and unpublished. We believe researchers examining the administration of the patent system and applicant filing behavior across technologies will find these data invaluable. The dynamics of new filings, pendency, and allowance are not entirely straightforward.[11] In Section 5, we use plots of cumulative disposal proportions to highlight the complexity involved in studying patent allowance rates over time and across technologies.

Calculating the stock of patents-in-force is not trivial. One must account for irregular statutory changes and other factors that affect patent term. Thus, in order to determine the status of each patent, we combine data on maintenance, parent applications, and patent term adjustments and extensions with different rules applying at time of filing (see Section 4 for details). We suspect that patents-in-force may be particularly valuable control variables for research in such areas as patent litigation and technology change. For instance, in Section 5, we show that patent-related appellate court decisions have remained fairly constant relative to the number of patents-in-force since 1980, suggesting that litigation (pursued through appellate court decision) has kept pace with growth in the stock of patents-in-force.

Monthly time series are limited to the 1981 to 2014 period because internal prosecution data do not exist in machine readable form before 1981. To facilitate researchers examining lengthier historic trends, we create the `annual` data file containing annual time series of applications, grants, and patents-in-force by NBER sub-category from 1840 to 2014. In Section 5, we use these data to put some recent trends in applications and grants into perspective. Notably, we show that the growth rates in patents and new filings experienced in the last 30 years are not unprecedented. We also illustrate how the composition of patented technology in the U.S. has changed over the past two centuries.

The paper is organized as follows. Section 2 provides background on the USPC system. In Section 3, we discuss our data sources and the methodology—depicted in **Figure 1**—for assigning NBER sub-categories to pending and abandoned applications, and in-force and expired patents. Section 4 describes the contents of the `historical_masterfile`, `monthly`, `monthly_disposal`, and `annual` data files as well as the intermediate data files created to generate these four output files. Section 5 highlights some trends observable in the data. Section 6 concludes by suggesting areas of future research employing these data.

[11] See Mitra-Kahn, B., Marco, A., et al (2013) for more information

2 Background

Within the PTO, there are three organizational tiers associated with patent examination: technology centers (TCs), examining art units, and individual examiners.[12] When an application is filed with the Office, it is classified and assigned to a TC and art unit based on its subject matter. Each TC specializes in a particular field (e.g. biotechnology and organic chemistry, computer architecture and software, etc.) and is composed of a number of examining art units. Patent examiners reside within art units, and applications are docketed to examiners by a supervisory patent examiner (SPE) within the art unit. Initial classification always precedes examination, though, as we explain further below, reclassification may occur during prosecution.

The USPC system classifies applications based on subject matter in order to route them through these organizational tiers to the appropriate examiner. Each USPC subject matter division includes a class (which delineates one technology from another) and a subclass (which delineates structural and functional features of the subject matter within the scope of the class).[13] Each art unit, in turn, is responsible for a set of USPC subclasses and contains examiners qualified to examine the technology classified in those subclasses. However, there is no one-to-one correspondence between USPC classes and art units. A single art unit may examine multiple subclasses; or, a single subclass may be examined by multiple art units.

With more than 450 classes and 150,000 subclasses, and ever-evolving technology, there is some degree of ambiguity in classification. Accordingly, the Office provides for reassignment of applications during prosecution. When an application appears on an examiner's docket, he or she may determine that the claimed technology falls under a class or subclass examined by another art unit. The application may then be assigned a new classification and re-docketed to the appropriate art unit. Applications can be reclassified up until the date of grant. In fact, classification itself is not official until the publication of the patent. Before patents are published, classification experts review the primary class,[14] and add several "cross-reference" classes to facilitate prior art searching in subsequent examinations. In contrast, abandoned applications do not receive such scrutiny (even those with pre-grant publication at 18 months). Thus, the quality of

[12] Technically, art units are organized into work groups under the main TC, and TCs themselves are parts of three broad disciplines.
[13] Although USPC subclass data exist, they do not aid in the assignment of NBER categories to applications and patents. We therefore omit these data from the classification algorithm.
[14] Note that the primary class is also referred to as the "original" class. The primary class may be updated due to revisions to the USPC system, thus the original class does not refer to the class-at-issue. For our purposes, we refer to the "initial" class to indicate the class-at-grant and the "current" class to indicate the updated class.

classification is likely to be higher for granted patents than for pending or abandoned applications.

The USPC system frequently undergoes revisions (over 1,000 since 1981[15]) to match the pace of changing technologies. Updates to the system may involve the establishment of new classes as well as revisions to or deletions of old classes. The PTO reclassifies patents according to the classification system in force, which complicates the use of classification data generated at different times. Furthermore, only certain systems reflect updates. To understand which systems are dynamic and which are static, it is useful to note that publications (pre-grant publications or "PGPubs") and published patents are static documents. As an example, class 705 ("Modern business data processing") was created in 1997 from certain subclasses of computer classes 395 and 364. Patent 5,590,269 (granted December 1996) was initially classified in U.S. class 395/209.[16] It was subsequently reclassified to U.S. class 705/7.16.[17] Full-text search databases (patft.uspto.gov or patents.google.com) show the updated ("current") class information for the patent. However, the image of the patent publication— a static document capturing classification at the date of publication—shows U.S. class 395/209. The bulk data released by the PTO in XML form[18] reflects the information contained on the face of patents or PGPubs. Thus, researchers using different data sources may record different classifications for the same patent. The current class information is consistent for granted patents; however, the PTO does not apply classification updates to unpublished abandoned applications (only to pending applications and pre-grant publications). Even with PGPubs that later abandon, undocumented reclassification may occur between publication and abandonment. To account for this, our classification algorithm exploits the scrutiny applied prior to patent publication by matching abandoned applications to granted applications from the same vintage and the same initial classification.

3 Data Sources and Methodology

Calculating stocks and flows of applications, abandonments, and patent grants requires micro-level data on filing date, disposal date, and disposal type (abandonment, grant, or pending) for patent applications filed at the PTO for the relevant period. To further sort these by NBER sub-category requires: (1) a concordance between current USPC class and NBER sub-category, and (2) the current USPC class for all abandonments, grants, and pending applications. The first can be derived from the NBER data file, which

[15] Based on the classification order numbers. *See http://www.uspto.gov/patents/resources/classification/archiverpt.pdf*
[16] See https://docs.google.com/viewer?url=patentimages.storage.googleapis.com/pdfs/US5590269.pdf
[17] See https://www.google.com/patents/US5590269
[18] See http://www.uspto.gov/products/xml-resources.jsp

provides a concordance with the 2008 USPC system. The current USPC class is not available for abandonments, but it can be imputed by creating a concordance between the USPC class at the time of disposal (class-at-disposal) and the current USPC class of granted patents. Below, we describe in more detail each of the source data files, the fields used from each file (summarized in **Table 3**), and the methodology used to create the necessary concordances and compile the historical time series. In the case of internal PTO data, we provide raw data files where possible, though micro-level data for non-publicly-available unpublished applications cannot be released.

Historical data for pending and abandoned applications filed between 1981 and 2014 (inclusive) come from the Patent Application Location and Monitoring (PALM) system. The PALM system, internal to the PTO, is used by patent examiners to monitor the progress of patent applications through prosecution. Any published data (e.g., data used in preparing pre- or post-grant publications) are extracted independently from the patent application documents (the "file wrapper") and undergo a separate quality review; those data are kept in separate systems. However, the PALM data are similar to the meta-data available on the Patent Application Information and Retrieval (Public PAIR) system for published applications.[19]

Because the PALM system is primarily a tracking tool, validated and verified published data are not linked back to the PALM system in any systematic way. Thus, there are some differences between the PALM system and official published records. Where possible we rely on published sources. However, where data on unpublished applications are not available elsewhere, we rely on the PALM data. **Figure 2** presents aggregate counts of applications by filing year and patents by grant year from published sources compared to those derived from PALM. It shows that published and PALM sources converge starting in 1981. Although PALM significantly undercounts applications and patents before this year, we use other internal systems to supplement the published data and assign NBER sub-categories to applications (filed as early as 1981) and patents (granted as early as 1840). The widening gap between PALM application counts and those of the published aggregates reflect, as we explain further in Section 4, an increasing trend in Requests for Continued Examinations (RCEs) and the way in which those are counted.

We draw the implementation dates for reclassification orders from the Classification Orders Index.[20] We use these dates to construct intermediate datasets containing sequential "classification era" variables that identify the time period each USPC scheme

[19] See http://portal.uspto.gov/pair/PublicPair
[20] See http://www.uspto.gov/patents/resources/classification/archiverpt.pdf

is in force between 1981 and 2014 at the aggregate-, class-, and subclass-level. This field is essential to identifying patent and application counterparts.

Annual counts of patent applications and grants since 1840 are available electronically from the PTO website.[21] However, these data do not provide detailed technology breakdowns. From micro-data, we build an annual time series of patent grants and patents-in-force from 1840-2014. Unlike for patent grants, machine-readable classification information for patent applications is not available prior to 1981.

We develop an algorithm to assign NBER sub-categories to granted utility patents and pending or abandoned utility patent applications using the NBER sub-categories in the NBER data, which contains information for 4.9 million patents granted between 1976 and 2006 (inclusive). Our methodology is outlined below and visually presented in **Figure 1**:

1. Update the USPC classification data of the NBER dataset to December 31, 2014, ("current") values and generate a concordance between current USPC class and NBER sub-category.
2. Assign an NBER sub-category to all patents (based on the concordance generated in step 1).
3. Generate a concordance between USPC class at issue to NBER sub-category for each classification era.
4. Apply the concordance generated in step 3 to assign an NBER sub-category to pending and abandoned applications based on their class-at-disposal or current class.

The NBER classification algorithm matches pending and abandoned applications to granted applications of the same vintage and initial classification and assigns the NBER sub-category of the granted patent to the pending or abandoned counterpart. To refine the concordance, we execute three iterations of the NBER classification algorithm. For each iteration, we use a different construction of the "era classification" variable based on reclassification orders relevant to any USPC classification, USPC classifications, or USPC sub-classifications. We provide more details about the construction of each "classification era" variable below. In the initial iteration, we match class-at-disposal or current class to NBER sub-category with a probability of at least 75%. Using USPC class- and subclass-level "classification era" variables, respectively, the second and third iterations classify previously ambiguous relationships if the relationship is exclusive (a probabilistic match equal to 75%).**Table 1** presents the results of the algorithm. Of the

[21] See http://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm

3.5 million non-granted applications filed between 1981 and 2014, 96.91% were successfully assigned to a single NBER sub-category; 0.52% could not be unambiguously classified, and 2.58% could not be classified due to missing USPC class data.[22] Of the 4.6 million patents granted between 1981 and 2014, 99.95% were successfully assigned to a single NBER sub-category; 0.04% were not classified due to NBER sub-category ambiguity, and 0.01% were missing the necessary USPC class data.

We create monthly (1981-2014) and annual (1840-2014) time series that include an approximate count of patents-in-force. Accurate accounting of patent expiration requires accounting for three policies regarding patent terms: statutory patent term, maintenance payments, and possible extensions or adjustments to the patent term. We account for these features; but, as we discuss below, we do *not* account for the effect that terminal disclaimers[23] may have on patent terms.

Since 1790, the statutory term of a patent has changed in several ways:

- The Patent Act of 1790 allowed for "any term not exceeding fourteen years."[24] Expiration dates for this time period are not available, so we must make assumptions in order to estimate the historical stock of patents-in-force. For this time period, we assume that all patents remained in force for exactly fourteen years from the date of issue.
- The Patent Act of 1836 granted the Commissioner of the PTO the ability to extend a patent's statutory term by seven years.[25] Again, data on which patents were extended does not exist, so we define upper (21 years) and lower (14 years) bounds in the estimation of the stock of patents-in-force.
- In 1861, the term for utility patents was set to exactly 17 years from the date of issuance.[26] This constant term, with no possibility for early expiration, makes it straightforward to calculate the number of patents-in-force through 1980.
- On December 12, 1980, maintenance payments were implemented.[27] All patents issued from applications filed on or after this date are subject to periodic

[22] Applications can take several months to receive an initial USPC classification code. As a result, some (3.69%) of the non-granted applications filed in 2014 were missing the data required to assign an NBER sub-category.
[23] A patentee or applicant may file a terminal disclaimer which disclaims the entire term or any terminal part of the term of a patent to be granted. *See* 37 CFR 1.321(a). A patentee or applicant may file a terminal disclaimer in an application or reexamination proceeding to overcome a non-statutory double patenting rejection. *See* 37 CFR 1.321(c) and MPEP § 804.02. Generally, the disclaimer limits the term of the second patent so that it will expire at the same time as the first. A patentee or applicant may also file a terminal disclaimer in an application or reexamination proceeding to overcome non-statutory double patenting rejection based on a U.S. patent or application that is not commonly owned but was disqualified under 35 USC 103(c) as resulting from activities undertaken within the scope of a joint research agreement. *See* 37 CFR 1.321(d).
[24] See http://www.ipmall.info/hosted_resources/lipa/patents/Patent_Act_of_1790.pdf
[25] See http://ipmall.info/hosted_resources/lipa/patents/Patent_Act_of_1836.pdf
[26] 12 Stat. 246, 249, § 16; 37 C.F.R. §1.362.

maintenance payments and, therefore, the patent status (in-force or expired) depends on whether it was maintained throughout its statutory term.[28] We use maintenance fee payment data collected by the PTO to estimate the stock of patents-in-force.[29]

- The patent statutory term was changed again in 1995 in accordance with the Agreement on Trade Related Aspects of Intellectual Property Rights. Patents issued from applications filed on or after June 8, 1995 are subject to 20-year patent terms beginning on the filing date of the *earliest* application to which priority is claimed.[30] (Patents that were in-force on or applications that were pending as of June 8, 1995 were subject to maximum terms of either 20 years from earliest filing date or 17 years from date of issuance, whichever more greatly benefited the inventor.[31]) U.S. applications may spawn "child" applications through the filing of continuation, continuation-in-part, or divisional applications that claim the priority date of the "parent" application.[32] To determine the earliest filing date requires continuity data for the application. Continuity refers to earlier parent applications that spawned the current application or patent. Failing to incorporate continuity data would result in overestimation of the number of patents-in-force (the filing date listed on the front page of the patent is the filing date for the child application, not the parent).[33]
- The American Inventor's Protection Act of 1999 allowed the patent statutory term to be lengthened beyond 20 years from earliest filing date through patent term adjustment (PTA) and patent term extension (PTE).[34] PTA and PTE are meant to extend the statutory term of the patent when the application was subject to regulatory delay. Omitting PTA and PTE would result in u0nderestimation of the number of patents-in-force because the terms of patents eligible for extension would be truncated.

---

[27] 94 Stat. 3017 § 41; 35 USC 41 or Public Law 96-517

[28] Maintenance fees are due at three times: 3-3.5 years, 7-7.5 years and 11-11.5 years following the patent's date of issue. (A patent may also be maintained if the fee is paid within a six month grace period following each window for an additional charge.) Failure to make a maintenance payment results in the expiration of a patent. See: http://www.uspto.gov/patents/process/maintain.jsp.

[29] Maintenance data can be obtained http://patents.reedtech.com/downloads/PatentMaintFeeEvents/1981-present/MaintFeeEvents.zip

[30] 35 U.S.C. §154(a)(2).

[31] 35 U.S.C. §154(c)(1).

[32] See: http://www.uspto.gov/web/offices/pac/mpep/s2701.html

[33] Continuity data are retrieved from the PTO Patent Technology Monitoring Team (PTMT) Custom Bibliographic Patent Data Extract DVD, available at http://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm#cust_xtract.

[34] See: http://www.uspto.gov/patents/law/aipa/summary.jsp and http://www.uspto.gov/web/offices/pac/mpep/s2750.html. Note that the act was enacted in November 1999, but PTA and PTE were available to pharmaceutical patent applications beforehand on an individual basis.

We use internal PTO administrative data on continuity to identify earliest filing dates for patents filed on or after June 8, 1995. Internal data are also used to identify any PTA/PTE. Grant dates for earlier cohorts, combined with maintenance data, are enough to generate expiration dates, which in turn yield our patents-in-force counts.

Our calculations for patents-in-force are not adjusted for terminal disclaimers[35] because data on terminal disclaimers are not available in machine-readable form. Historically, terminal disclaimers are uncommon and, at present, we do not feel that the accuracy of our counts is significantly compromised by their omission.[36]

Lastly, we do not incorporate any information from administrative post-grant proceedings (e.g., re-examination) or judicial proceedings (e.g., patent litigation) that may influence some or all of the patent claims. Because these data are not publicly available in consistent machine-readable format, we cannot incorporate cancelation or invalidity actions into our patents-in-force calculation.

Published historical data exist only at the annual level for applications (since 1840).[37] Further, much of the classification information is missing for patent records destroyed (or partially destroyed) in the 1836 fire. Thus, our time series for patents-in-force and grants by technology category begin in 1840. The time series for applications cannot be disaggregated by technology category until 1981.

4 Description of Data Files

Using the data generated on NBER sub-category and status of patents across time, we present three intermediate and the four output data files intended for research. Unless otherwise stated, "patents" is used to refer to regular utility patents; this excludes applications and patents for designs, statutory invention registrations, plants, reissues and defensive publications.[38]

The `orders` intermediate data file includes the daily dates (`date`) spanned by each classification era (`era`) between 1981 and 2014. Classification Orders are issued by the USPTO sporadically to account for new and evolving technologies. An era, therefore, is the time between Classification Orders, during which the USPC classification scheme

[35] See footnote 23.
[36] Internal estimates suggest that 11.9% of patents granted in 2012 contain terminal disclaimers. This rate has increased from 3.4% in 1990; thus we expect this to become an increasing problem with respect to calculating the expiration date of patents in the future.
[37] See: http://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.htm
[38] For more information, see: http://www.uspto.gov/web/offices/ac/ido/oeip/taf/patdesc.htm

remains static. An era begins the day of an order and ends the day before the subsequent order. Eras are used in conjunction with class at issue to estimate current USPC class and NBER sub-category.

The **orders_class** and **orders_subclass** intermediate data files include the daily dates (**date**) spanning each classification era (**era**) between 1981 and 2014 at the USPC class-level (**class**) and subclass-level (**subclass**), respectively. Classification Orders provide details on changes to specific USPC classes and subclasses. For each class and subclass, an era begins the day of an order which modifies the class or subclass and ends the day before the subsequent order affecting that class or subclass. For example, Classification Order 1902, issued January 4, 2011, directly modified USPC class 340 and other USPC classes indirectly. As a result, all classes and subclasses affected by Classification Order 1902 will begin a new era. The era will remain constant for USPC classes and subclasses unaffected by Classification Order 1902. Since each Classification Order does not alter the entire USPC system, we use the **orders_class** and **orders_subclass** data files to apply NBER sub-categories where matching class-at-disposal or current class to NBER sub-category yielded an ambiguous result (probability of less than 75%).

The **historical_masterfile** file contains detailed data on each published or publicly-available non-published application, including its application, PGPub and patent numbers (**appl_id**, **pubno** and **patent**, respectively); NBER sub-category (**nber**); current USPC class and subclass (**uspc** and **uspc_sub**, respectively); application, priority, publication and disposal dates (**appl_dt**[39], **prior_dt**[40], **pub_dt**, **disp_dt**, respectively); disposal type (**disp_ty**); total patent term adjustment (**pta**); and estimated expiration date (**exp_dt**).[41] These data are aggregated in various ways in the **annual**, **monthly** and **monthly_disposal** datasets. Note that 2,633 patent numbers are preceded by "X." In 1836, a fire destroyed many patents and applications, some of which were never recovered; the "X" patents indicate those that were granted before the fire. This means that patent 0000001 (granted 7/13/1836) is wholly separate from patent X000001 (granted 7/31/1790).

[39] **appl_dt** values missing from PALM are back-filled using the PTMT Custom Bibliographic Patent Data Extract DVD, available at http://www.uspto.gov/web/offices/ac/ido/oeip/taf/reports.htm#cust_xtract. Since the DVD is only updated through 2013, application dates for more recent filings may be missing.
[40] The **prior_dt** field is pulled from the PTMT Custom Bibliographic Patent Data Extract DVD.
[41] The variable **exp_dt_max** represents the upper limit of the expiration date for the period 1836-1861, during which time the Commissioner of the PTO had the ability to extend a patent's statutory term by seven years. There are no records of which patents did and did not receive extensions, so we provide the upper limit for this period.

The `annual` data file includes annual counts of granted patents and patents-in-force by NBER sub-category, and aggregated filed applications.[42] The `monthly` data file begins in January 1981, the first year of reliable patent application data.[43] It captures many of the same variables as the `annual` file, but also adds abandoned and pending application counts, by NBER category. The `monthly_disposal` data file follows prosecution outcomes over time for each monthly application filing cohort across NBER categories. For all incoming applications to the PTO, there are only two possible terminal outcomes ("disposals"). Applications are either issued as granted patents, or they are abandoned. Applications that have yet to be disposed are pending.[44] The dataset tracks the monthly disposals for each monthly application cohort until all applications are disposed of via issuance or abandonment (unless applications remain pending as of December 31, 2014). Although this dataset tracks entry cohorts,[45] it could be manipulated to capture exit cohorts for a more in-depth analysis of exit pendency or allowance rate by exit cohort. See Appendix 4 for a description of each of the variables in each data file.

The stock of pending applications at the end of the month is determined by the inflow of incoming applications and the outflow of grants and abandonments during that month.

Since our detailed application data begin only in January 1981, the stock of pending applications cannot be accurately measured in the early period. That is, there is left truncation on new applications. **Figure 3** uses the `monthly_disposal` data to plot a survival curve averaged across all monthly filing cohorts in 1981-1983, where "survival" indicates that the application is still pending. For these 36 cohorts, only 1% of applications are still pending as of 73 months after filing. Thus, we report in the `monthly` data file the estimated stock of pending applications beginning 73 months after January 1981 (February 1987) in order to minimize undercounting due to left truncation.

It is important to note that examination decisions by examiners (allowances and rejections) are not technically disposals. For allowances, applicants must pay a fee in order for the patent to be granted. In the data file, we use the day that the patent is

[42] A monthly time-series can be created for patents and patents-in-force using the `historical_masterfile` file.
[43] Without sufficient data on abandonments, discerning between the number of abandoned and pending applications from the raw counts of new applications would be impossible.
[44] Some official documents will distinguish between the stock of unexamined applications (those awaiting their first substantive examination), and the total stock of pending applications. We report the latter of these.
[45] For more discussion and analysis of entry and exit cohorts, see Mitra-Kahn, Marco, et al., . (2013 "Patent backlogs, inventories and pendency: An international framework," IPO & USPTO joint report: http://www.ipo.gov.uk/ipresearch-uspatlog-201306.pdf)

published as the disposal, or issuance, date.[46] Applicants have many different ways to respond to rejections (including appeals to the Patent Trial and Appeal Board or its institutional predecessors, and ultimately to the federal courts), so rejections do not necessarily represent a terminal disposal of applications. Disposal through abandonment can occur in two ways. First, the applicant may expressly abandon an application. Second, abandonment will occur automatically when an applicant fails to respond to an office action (typically within six months). Express abandonment is fairly rare;[47] more often, the applicant will fail to respond to an office action (most frequently a rejection). Thus, the most common paths for terminal disposals are allowance, followed by issuance and rejection, followed by abandonment. When an applicant fails to respond to an office action, the examiner may wait several months before officially filing the abandonment in case the applicant files a petition to revive the application. When the examiner eventually files the abandonment, the abandonment date is back-filled to reflect the initial response deadline.[48] For this reason, there is some degree of right censoring for the six or so most recent months in the time series. The data file goes through the end of 2014, so that we ensure a sufficient period of time to capture all abandonments.

We include only "regular" utility patent applications, excluding provisional applications, re-issue applications, and Patent Cooperation Treaty (PCT) applications. However, we do include utility patent applications that are filed as a national stage entry of a parent PCT application or converted provisional applications. In some official statistics, the total number of filed applications includes RCE filings. RCEs are one version of a "non-serialized continuations," a continued application that retains the serial number of the parent application. Because the parent and child applications are substantially the same, and because the "parent" application is simultaneously abandoned, we exclude RCEs from our counts of new filings and abandonments. However, we do include traditional continuations (including continuations-in-part and divisionals) in our counts because they do not necessitate the abandonment of the parent application (so that the count of pending applications may increase).[49]

In a dynamic context, information about stocks and flows alone is not enough to accurately distinguish between pendency and the rate of abandonment when those may be varying over time. If abandonments generally occur faster than grants, then one will always be comparing abandonments from one filing cohort to patent grants from a

[46] Some confusion exists because grants are sometimes called "allowances," and the grant rate (the number of issuances divided by the number of disposals) is sometimes called the allowance rate.
[47] Internal statistics show that the rate of express abandonments peaked in 2002 at 7.6% (as a proportion of all abandonments) and has steadily fallen to 2.1% in 2012.
[48] For example, if the office action were mailed February 1, the due date for response would be six months later, August 1 (MPEP § 1.134). If the examiner files the abandonment on November 1, the abandonment date will still be August 1.
[49] For more information, see MPEP § 706.07(h)

different filing cohort. For instance, if the number of new applications is growing at a high rate, then the ratio of granted patents to abandonments (the “allowance rate”) would continually trend downwards, even if the application quality and examination quality remained constant.[50] The problem is compounded if the duration of pendency changes over time.

The `monthly_disposal` data file is intended to provide more information about pendency and abandonment for researchers. This data series provides—for the first time—detailed information about disposal and pendency for patent applications. It does not provide information on any particular abandoned or pending application, but instead contains the distribution of disposal dates for each monthly filing cohort from 1981 to 2014. Further, we separate disposals by type (grants and abandonments) and by NBER sub-category. From these anonymized data, researchers can analyze pendency and abandonment at the PTO in a way that was heretofore impossible.

It is important to note that the counts in `monthly_disposal` are cumulative within each `appl_month`, so that they represent stocks, not flows. The number of issued, abandoned and pending applications always sum to the corresponding number of applications for that entry cohort and NBER sub-category. That value is consistent with the number of new applications reported in the `monthly` data file for that month.

5 Discussion

From the *USPTO Historical Patent Data Files*, we highlight certain trends and also demonstrate some possible uses for these data. We start with aggregate trends, illustrating the overall growth in applications and patents, and then we move to technology specific growth rates. **Figure 4** shows the annual count of applications (by filing year) and patents (by grant year) from 1840 to 2014. Many scholars have noted the sharp increase in the number of patent applications in the early 1980s (Kortum and Lerner, 1999). While this is evident in the raw counts, historic time series show that the growth rates experienced in the last 30 years—while high—are not unprecedented. In **Figure 5**, we plot patent and application counts in natural logarithms, enabling us to more easily assess growth rates by comparing slopes. The reference line is equivalent to constant growth over the period, so any portion of the graph that is steeper than the reference line represents a growth rate that is higher than historical averages. The figure shows that the growth rate in patents

[50] See Mitra-Kahn, Marco, et al. (2013) and Carley et al. (2014)

and in applications over the last 30 years is roughly equivalent to the historical average; whereas, the growth rate from 1850 to 1870 dwarfs any modern day growth.

The trend in patents-in-force resembles that of patent grants until the institution of policies creating varying patent terms. **Figure 6** shows the estimated annual number of patents-in-force from 1840 to 2014. For most of this period, the growth in patents-in-force reflects the previous year's patent grants and the stock of expiring patents. Interestingly, between 1980 and the present, while patents grew at a roughly constant rate, the stock of patents-in-force initially experienced no growth. This is because the onset of maintenance fees causes some patents to expire before the end of the statutory patent term. Following implementation of maintenance fees, expiring patents offset the growth in patenting so that the stock of patents-in-force remained fairly constant. By the early 1990s almost all granted patents were subject to maintenance fees, and a steady state growth rate returned.

We expect the stock of patents-in-force to be an important control variable for various time series analysis. Patents-in-force disaggregated by NBER sub-category may provide an indication of how crowded patented technology spaces are over time and relative to each other. Aggregate time series of patents-in-force provide an invaluable weight for various trends. For example, **Figure 7** depicts the annual number of patent-related appellate court decisions per patent-in-force from 1929 to 2006.[51] It clearly shows decisions per patent-in-force trending down for much of the time period and holding largely flat since 1980. Thus, in the last 35 years, the number of appellate court decisions rendered each year has largely kept pace with the stock of patents-in-force.

Likewise, **Figure 8** shows new applications per patent-in-force from 1840 to 2014. Normalizing by the number of patents-in-force reinforces the observation that the growth rate in applications over the last 20 years is high relative to that of the last century, but still lower than that of the mid- to late-19th century.

Figure 9 presents a high level view of the monthly data, showing the annual number of applications filed between 1981 and 2014 by NBER category (see **Table 2** for details on NBER categories and sub-categories). Filings in Computers & Communications clearly stand out, having grown in absolute number at significantly higher rates than all other categories. Chemical applications, in contrast, have been largely flat since 1990. **Figure 9** also highlights differences across technologies in the responsiveness of application

[51] [Reference Henry See Matthew and Turner paper and provide detail on data].(2005). We would prefer to use the number of district court filings. However, these data are not consistently and available in a historical time-series. While appellate court decisions are subject to significant selection problems, they do not suffer from as much under-reporting as district court cases.

volumes to significant policy events or economic cycles. For instance, we see more pronounced peaks in filings prior to the 1995 change in the patent term for Chemical and Drugs & Medical compared to the other categories. Likewise, applications in all NBER categories decreased following the recent financial crisis, though Computers & Communications filings exhibited the sharpest peak-to-trough decline. Post-crisis recovery has also varied. Applications in Computers & Communications and Electrical & Electronics quickly recovered and resumed growth, while Chemical filings rapidly recovered but subsequently held flat. Applications in Drugs & Medical and Other categories recovered at a much slower pace, and the number of Mechanical applications only returned to pre-crisis levels in 2013.

Figure 10 disaggregates the previous figure into NBER sub-categories. At this more granular level, we see that much of the growth in Computers & Communications filings evident in **Figure 9** was driven by surging applications in Computer Hardware and Software and Communications sub-categories, as opposed to the other sub-categories (Computer Peripherals and Information Storage). Likewise, **Figure 10** shows that filings in the Drug sub-category generated most of the volatility around the 1995 patent term change for Drugs & Medical category. This is not surprisingly given that drug providers place a relatively higher premium on lengthy patent terms and are possibly better positioned to accelerate the application process for early-stage drugs or treatments. To provide a clearer view of trends in filings for each NBER sub-category, in **Appendix 1**, we plot applications from 1981 to 2014 by NBER sub-category broken out by quartiles based on 2014 filings.

Figure 11 shows the annual number of issued patents by grant year between 1981 and 2014 by NBER category. As with annual applications, patents in the Computers & Communications and Electrical & Electronics categories vastly outnumber patents in all other categories beginning in the early 2000s. The 1995 policy change also exerted a visible impact on patent counts across all categories, however, the impact occurred several years later, characterized by a sharp increase in patents in the mid-1990s. Interestingly, **Figure 11** shows a relatively sharp increase in patent grants since 2009, as pendency times have decreased. This is evident across all NBER categories even those where incoming applications have held fairly flat over the past decade(s).

Figure 12 further disaggregates annual patents by NBER sub-category to show specific drivers of growth within categories. We provide comparable trend graphs broken out by quartiles based on 2014 grants in **Appendix 2**. The growth in Communications and Computer Hardware and Software patents is quickly evident. The upturn in patenting since 2009 is observed across most NBER sub-categories, suggesting that factors unrelated to particular technologies may be driving the recent growth in issued patents.

Indeed, greater examination capacity at PTO, concerted efforts to reduce backlog and lessen pendency, and an improving economic environment are all likely contributing factors. Release of these data provides a foundation for further inquiry into the responsiveness of different technologies to such factors.

We expect the `monthly_disposal` data file to be particularly valuable for researchers examining application disposal and pendency over time and across technologies. To demonstrate this, in **Figure 13**, we plot the cumulative disposal proportions over months since filing across NBER categories for the cohort of applications filed in January 2002. The pending line corresponds to a typical survival curve, where survival indicates applications that are still pending. The patent issuance and abandonments lines correspond to cumulative incidence functions (CIFs), i.e. the proportion of applications having issued and abandoned, respectively, as of each month from filing. [52] The number of pending applications approaches zero as applications are disposed of via abandonment or issuance, so that, at any point in time since filing, the sum of all three of these curves is 1.0, or 100% of the entering cohort of applications. The CIFs highlight the complexity involved in discussing allowance rates at the PTO (or any patent office). One may care about the terminal allowance rate (at the limit), but allowance rates for each NBER category are not stable across application cohorts or over time. The figure shows that the terminal allowance rate for the January 2002 cohort ranges from 57% (Drugs & Medical) to 81% (Electrical & Electronics). The rate of abandonment exceeds that of issuance for all NBER sub-categories in the months immediately following filing; this trend persists for some NBER categories longer than for others. **Figure 14** is similar to the previous figure, but examines pendency across several cohorts for the Chemical NBER category. It clearly shows how allowance rates may vary over time for applications in the same NBER category. In fact, patents for inventions of this type have allowance rates of 58% in 1981, 55% in 1991, and a much higher 73% in 2001.

Both **Figure 11** and **Figure 12** show a considerable shift occurring since 1980 in the composition of patented technology, i.e. from near uniformity across most NBER categories to the dominance of Computer & Communications and Electrical & Electronics. Using data from the `annual` file, we can put this more recent shift into historic perspective by examining the evolution of patented technology since 1840. In **Figure 15**, **Figure 16** and **Figure 17**, we rank NBER sub-categories (color-coordinated based on category) in terms of the number of patents issued for each decade since 1840.[53]

[52] See Carley et al. (2014) for more information on allowance rates.
[53] Decades were constructed as follows:$1980 = \{1980, 1981, \ldots, 1989\}, 1990 = \{1990, 1991, \ldots, 1999\}, \ldots, 2010 = \{2010,2011,\ldots,2014\}$

The sub-category with the most patents issued in a particular decade is ranked number one, while the sub-category with the fewest patents issued is ranked number 37 (or lower in the case of ties, which only occur early in the time-series when zero patents were issued for some sub-categories). Rankings allow us to track patenting in each sub-category relative to all other sub-categories and discern how the composition of patented technology has changed over the past two centuries. Because **Figure 15** is visually complex, we highlight the trend in rankings for particular subsets of sub-categories in **Figure 16** and **Figure 17** . **Figure 16** features the five sub-categories with the highest ranking (including ties) in the first decade (1840-1849) of the time series. **Figure 17** highlights the five NBER sub-categories with the highest rankings during the 2010-2014 period. We include similar graphs tracking the rankings of the bottom five sub-categories in the first and last decade in Appendix 3.

Figure 16 indicates that, through much of the past two centuries, the most patented technologies were in more traditional areas – Miscellaneous Others (sub-category 69), Miscellaneous Mechanical (59), Materials Processing and Handling (51), Agriculture, Husbandry, Food (61) and Heating (66).[54] These sub-categories remain firmly in the top 10 until the early part of the 20th century, after which each sub-category's ranking falls over time with different rates of decline. Miscellaneous Others remains in the top 10 through 2014, and Miscellaneous Mechanical and Materials Processing and Handling only fall out of the top 10 in the early 2000s. Agriculture, Husbandry, Food, and Heating, however, fall steadily throughout the 20th century to the bottom tier by 2014.

Similarly, **Figure 17** shows considerable variation in the historical rankings of the current top sub-categories. As one would expect, patenting associated with Semiconductor Devices (46), Drugs (31) and Computer Hardware and Software (22) was minimal until the second half of the twentieth but climbed rapidly thereafter.[55] Conversely, over the past century, the number of Communications patents issued has steadily increased relative to other sub-categories, and Miscellaneous Chemical patents remain among the top sub-categories through the entire time series.

Overall, both **Figure 16** and **Figure 17** suggest that shifts in the composition of patented technology occurred slowly for much of the past two centuries, at least for the NBER

[54] The miscellaneous sub-categories are among the most numerous, but they may also be the most varied; this variation may also make recent miscellaneous sub-categories somewhat dissimilar to older ones.

[55] While it may seem counterintuitive to consider anything in existence in the 1800s "computer hardware or software," computer-like inventions did exist. Many so-classified patents from that time cover inventions related to photography (e.g., patent 0007655 issued in 1850, 0030850 issued in 1860, and 0104963 issued in 1870). Similarly, many early "information storage" patents cover inventions related to phonographs (e.g., patents 0200521 issued in 1878, and 0382416 issued in 1888); the earliest patented invention classified as a "semiconductor device" (patent 0755840) was issued in 1904 to cover a detector for electrical disturbances.

sub-categories at the tails of the distribution. Only since the 1950s do rankings show drastic changes in the most (and least) patented sub-categories. The Semiconductor Devices sub-category climbs from the bottom ten to the top three in two decades; whereas it takes seven decades for Apparel and Textiles to go from one the most patented sub-categories to one of the least. Interestingly, the Miscellaneous sub-categories (Chemical, Others, and Mechanical) are persistently among the highest ranked. These sub-categories present an intriguing case for further research as they contain technologies not easily placed into existing sub-categories and, as such, may represent more innovative or radical invention.[56]

Figure 18 shows the change in composition in a different way. The Gini coefficient is traditionally used to quantify income inequality, but its theory can be generalized to examine the distribution of patents across classes as well. Using annual grants, we calculated the Gini coefficient over time for each classification scheme (USPC class and subclass and NBER sub-category). Judging by the USPC class and subclass trends, annual pools of patent grants are becoming increasingly concentrated; the NBER sub-category trend, however, shows a persistently high degree of concentration until the 1930s, reaching an all-time low in the early 1990s, followed by an increasing trend. While these differing trends may be endogenous (based on reclassification orders) we present it here as a novel topic for future research.

6 Conclusion

In this paper, we build upon the contribution of Hall, Jaffe, and Trajtenberg (2001), applying the NBER sub-categories to pending and abandoned applications and in-force and expired patents dating as far back as 1790. Our methodology produces the *USPTO Historical Patent Files*. Based on the `historical_masterfile`, we generate three new time-series datasets for researchers to employ in primary analysis or as controls. The `monthly` and `monthly_disposal` datasets provide detailed information on the stocks and flows of patent applications and patent grants at the PTO by economically relevant technology categories. We believe these data will be a useful resource for economists and other researchers considering the administration of the patent system and applicant filing behavior across technologies over the last several decades. Notably, the complex dynamics of new filings, pendency, and allowance is largely unexamined outside the PTO.

[56] See Strumsky et al (2012) for more research into this topic

Counts of patents-in-force may be particularly valuable control variables for multiple streams of research. We stress the importance of accounting for the stock of patents-in-force when considering trends in patent litigation. Counts of patents-in-force may also capture the level of crowding or congestion within a technology space. How such crowding affects research and development or commercialization may be of particular interest to some scholars.[57]

Lastly, the `annual` data file, and the underlying `historical_masterfile` micro-data, enables researchers to more easily study historic trends in patenting and technology change. These data enable researchers to put recent trends in applications, grants, and even patent litigation into perspective. When aggregated by NBER sub-category, these data inform on the composition and evolution of nearly two centuries of patented technology in the U.S. It shows how rapidly technologies have changed in the last 30 years when compared to the prior 130 years. It provides for study of the responsiveness of applications, grants, and live patents to policy changes and exogenous shocks.

7 References

Carley, Michael and Hegde, Deepak and Marco, Alan C., What is the Probability of Receiving a US Patent? (January 10, 2014). USPTO Economics Working Paper No. 2013-2. Available at SSRN: http://ssrn.com/abstract=2367149 or http://dx.doi.org/10.2139/ssrn.2367149

Graham, S., & Hegde, D. (2012). Do inventors value secrecy in patenting? Evidence from the American Inventor's Protection Act of 1999. http://cardozo-new.yu.edu/sites/default/files/Graham.Stuart.pdf

Griliches, Z. (1998). Patent statistics as economic indicators: a survey. In R&D and productivity: the econometric evidence (pp. 287-343). University of Chicago Press.

Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citation data file: Lessons, insights and methodological tools (No. w8498). National Bureau of Economic Research.

[57] See Lerner (2004) and Lerner (1995)

Henry, Matthew and Turner, John L., The Court of Appeals for the Federal Circuit's Impact on Patent Litigation (February 2005). Available at SSRN: http://ssrn.com/abstract=713622 or http://dx.doi.org/10.2139/ssrn.713622

Kortum, S., & Lerner, J. (1998, June). "Stronger protection or technological revolution: what is behind the recent surge in patenting?". In Carnegie-Rochester Conference Series on Public Policy (Vol. 48, pp. 247-304). North-Holland.

Lerner, Josh, (1995), Patenting in the Shadow of Competitors, Journal of Law and Economics, 38, issue 2, p. 463-95.

Lerner, J., & Tirole, J. (2004). The economics of technology sharing: Open source and beyond (No. w10956). National Bureau of Economic Research.

Mitra-Kahn, B., Marco, A., et al., 2013, "Patent backlogs, inventories, and pendency: An international framework," IPO- & USPTO joint report, http://www.ipo.gov.uk/pro-ipresearch.htm.

Schumpeter, J. A. (1935). The analysis of economic change. The Review of Economics and Statistics, 17(4), 2-10.

Strumsky, D., Lobo, J., & van der Leeuw, S. (2012). Using patent technology codes to study technological change. Economics of Innovation and New technology, 21(3), 267-286.

Tables and Figures

Table 1: NBER classification algorithm results (%, application years 1981-2014).

NBER Category	*Applications*	*Patents*
11-69 (Classified)	96.91	99.95
70 (Not classified)	0.52	0.04
80 (Missing)	2.58	0.01

Our NBER classification algorithm was very successful in assigning a single NBER sub-category to applications and patents. Note that the time required to assign an application its USPC class has not yet passed for many applications filed during 2014. The classification rate for applications filed 1981-2013 is 97.00%

Table 2: NBER categories and sub-categories.

Category	*Category name*	*Sub-category*	*Sub-category name*
1	Chemical	11	Agriculture, food, textiles
		12	Coating
		13	Gas
		14	Organic compounds
		15	Resins
		19	Misc. (chem)
2	Computers & Communications	21	Communications
		22	Computer hardware and software
		23	Computer peripherals
		24	Information storage
		25	Electronic business methods and software
3	Drugs & Medical	31	Drugs
		32	Surgery, medical instruments
		33	Biotechnology
		39	Misc. (drugs&med)
4	Electrical & Electronics	41	Electrical devices
		42	Electrical lighting
		43	Measuring, testing
		44	Nuclear, X-rays
		45	Power systems
		46	Semiconductor devices
		49	Misc. (elec)
5	Mechanical	51	Materials processing & handling
		52	Metal working
		53	Motors, engines, parts
		54	Optics
		55	Transportation
		59	Misc. (mech)
6	Others	61	Agriculture, husbandry, food
		62	Amusement devices
		63	Apparel & textile
		64	Earth working & wells
		65	Furniture, house fixtures
		66	Heating
		67	Pipes & joints
		68	Receptacles
		69	Misc. (others)
7	Not Classified	70	Not classified
8	Missing	80	Missing

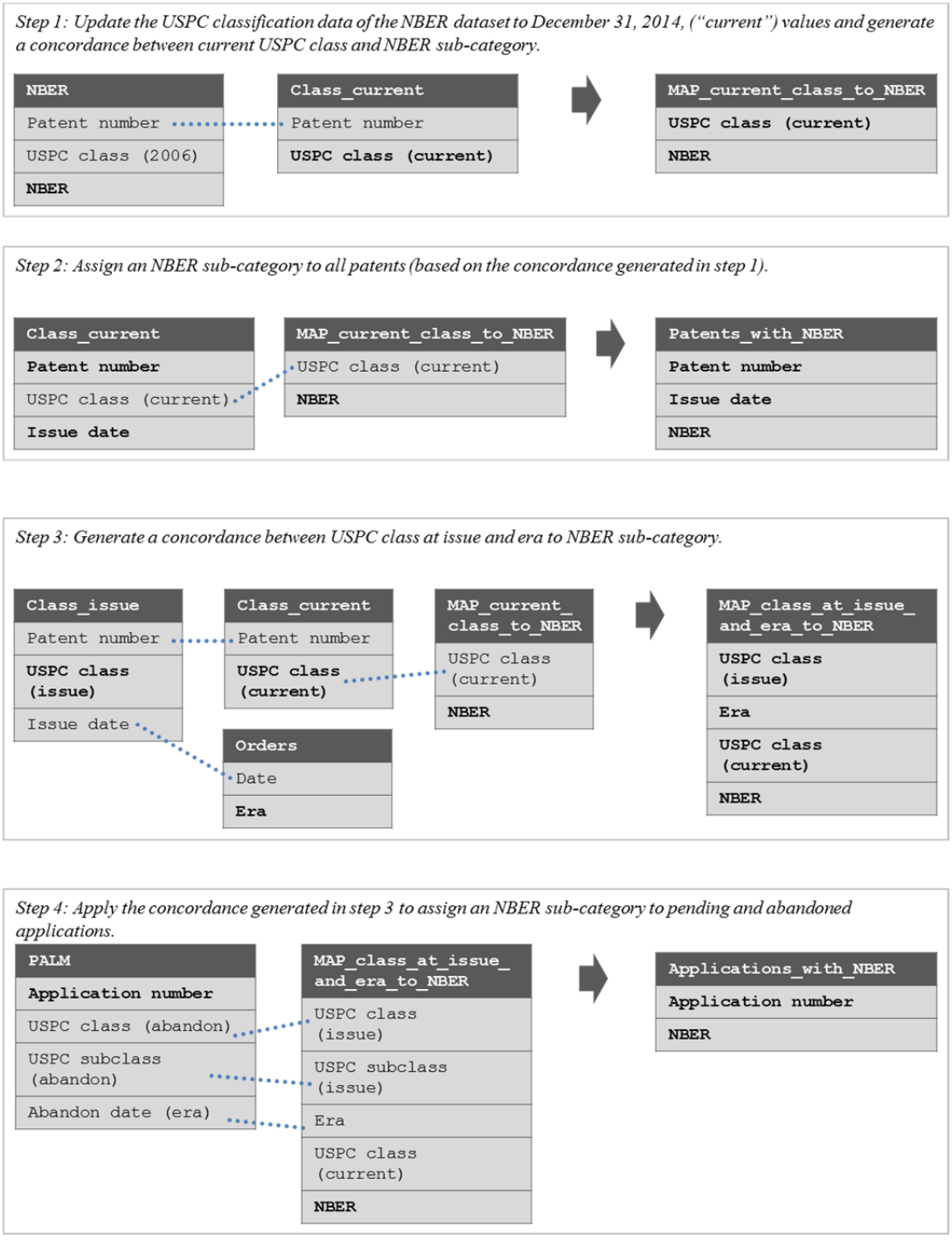
These category and sub-category codes and names are as they first appeared in the NBER paper. We have added sub-categories 70 and 80 (and consequently categories 7 and 8) to account for those applications and patents that our algorithm was unable to label otherwise.

Table 3: Data sources.

Source	*Disposal type*	*Sample*	*Fields*
NBER	Patents	Patents granted 1976-2006	Patent number
			Current USPC class
			NBER sub-category
class_current (internal to USPTO)	Patents	Patents granted 1790-2014	Patent number
			Current USPC class
			NBER sub-category
class_issue (internal to USPTO)	Patents	Patents granted 1790-2014	Patent number
			USPC class at issue
			Issue date
orders	NA	NA	Date
			Era
PALM	Applications	Filings 1981-2014	Application number
			Disposal date
			USPC class at disposal

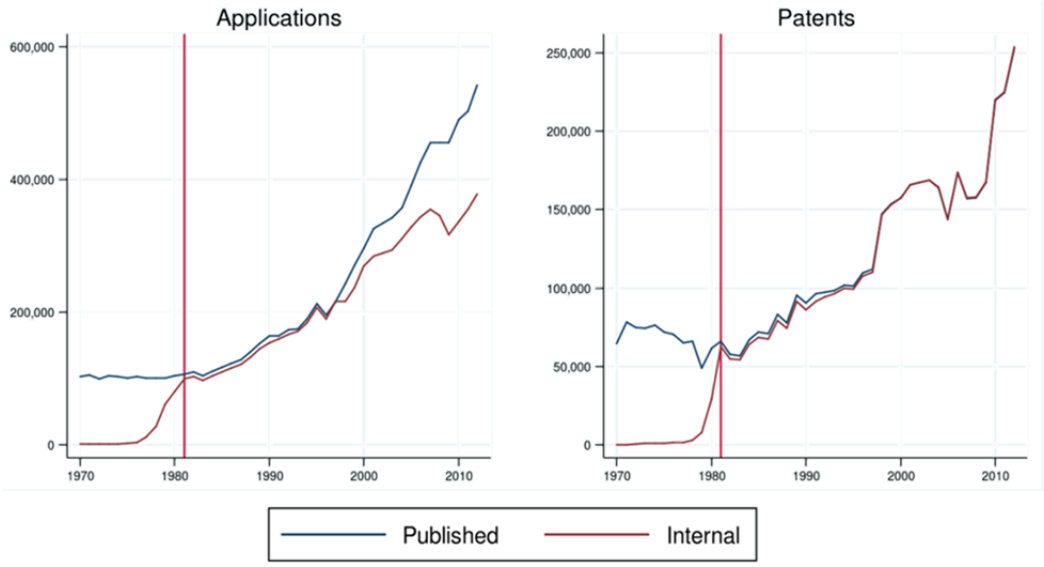
The table above describes the data sources used. See **Figure 1** *for more information on the algorithm used.*

Figure 1: NBER classification algorithm.



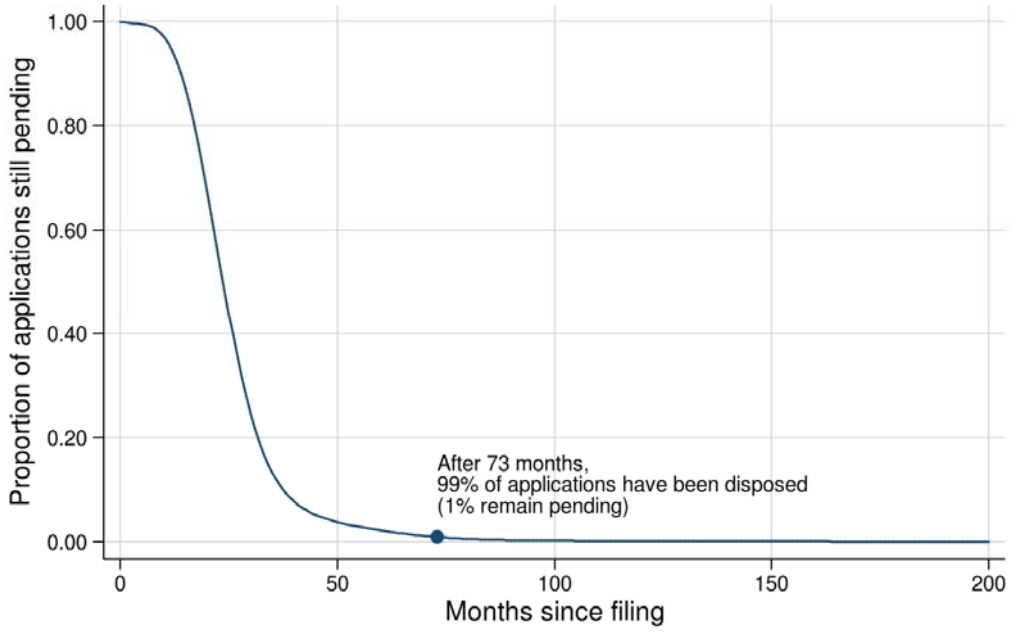
* We thank Charlie DeGrazia for his excellent research assistance with this project. The views expressed are those of the individual authors and do not necessarily reflect official positions of the Office of Chief Economist or the U. S. Patent and Trademark Office.

Figure 2: Comparison of sources.



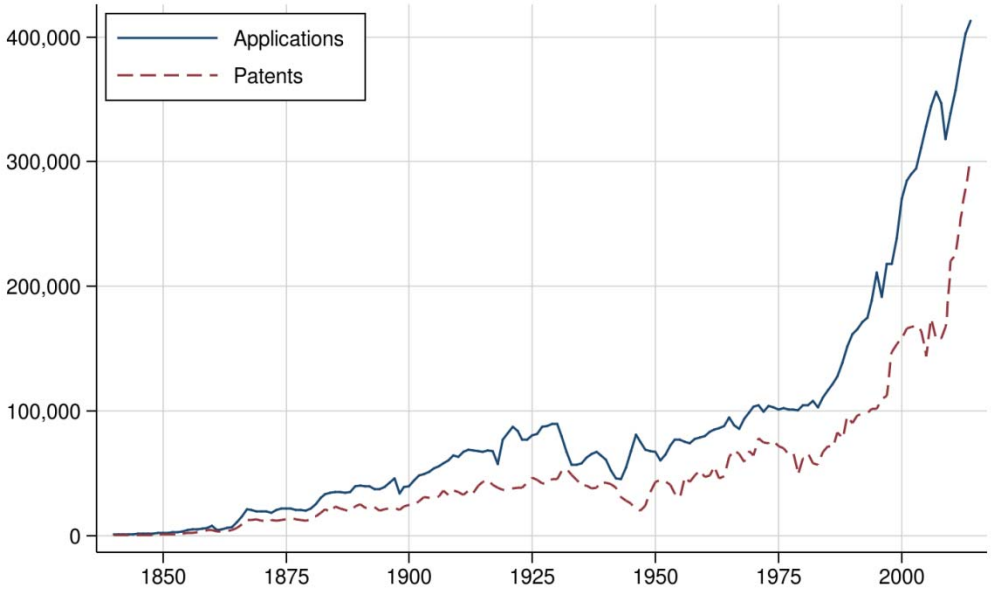
1981 is the first year of reliable internal data and use this as a starting point in several output data files. The growing difference between the number of published and calculated applications is the result of Requests for Continued Examinations (RCEs).

Figure 3: Average survival curve for entry cohorts 1981m1 – 1983m12.



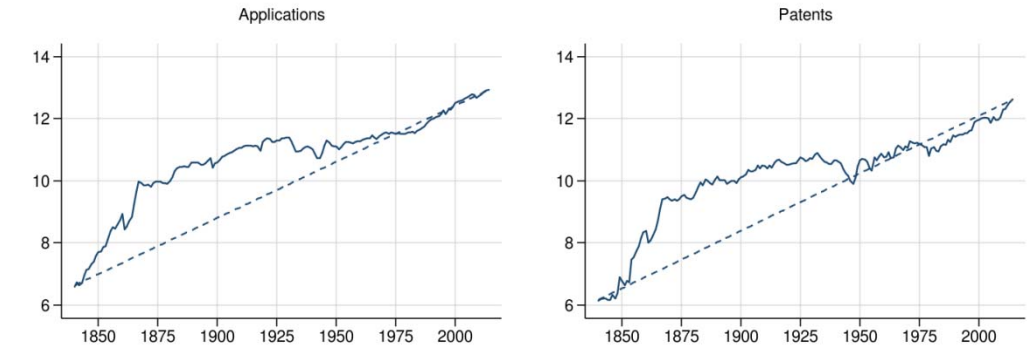
We report in the `monthly_disposal` *data file the estimated stock of pending applications beginning 73 months after January 1981 (February 1987), so that we minimize the expected undercounting.*

Figure 4: Annual count of applications and patents.



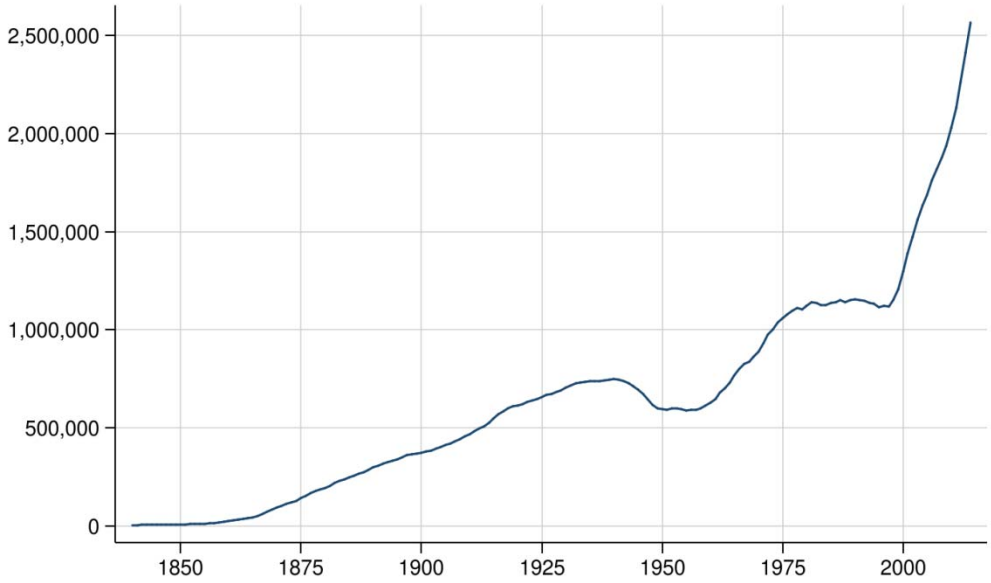
Many scholars have noted the sharp increase in the number of patent applications in the early 1980s. While this is evident in the raw counts, historic data show that the growth rates experienced in the last 30 years are not unprecedented.

Figure 5: Log of annual count of applications and patents.



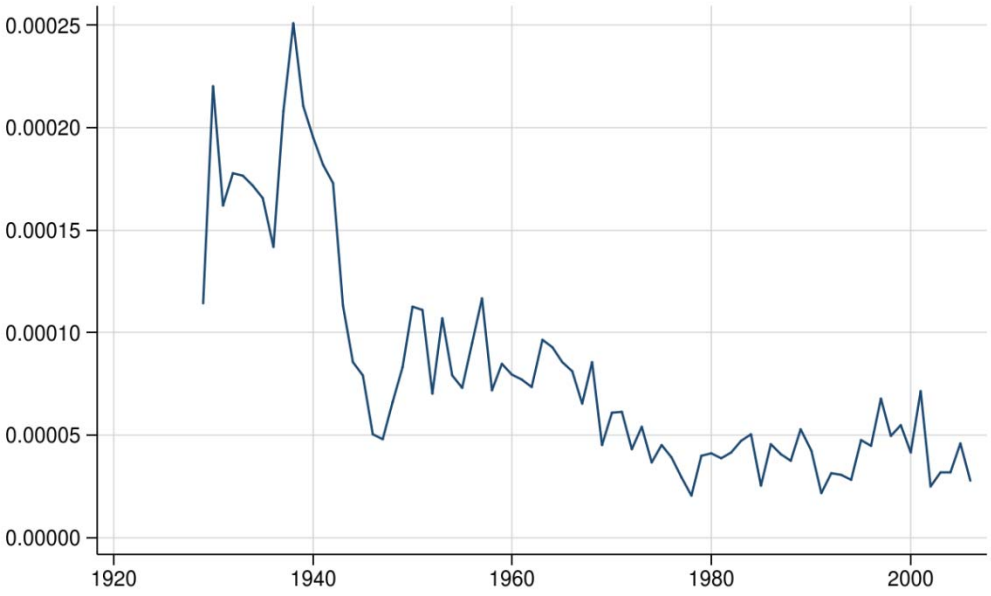
The growth rate in patents and in applications over the last 30 years is roughly equivalent to the historical average. Furthermore, the growth rate from 1850 to 1870 dwarfs any modern day growth.

Figure 6: Annual count of patents-in-force.



Following the implementation of maintenance fees (December 12, 1980), expiring patents offset the growth in patenting so that the stock of patents-in-force remained fairly constant. By the early 1990s almost all granted patents were subject to maintenance fees, and a steady state growth rate returned.

Figure 7: Annual appellate decisions per patent-in-force (1929-2006).

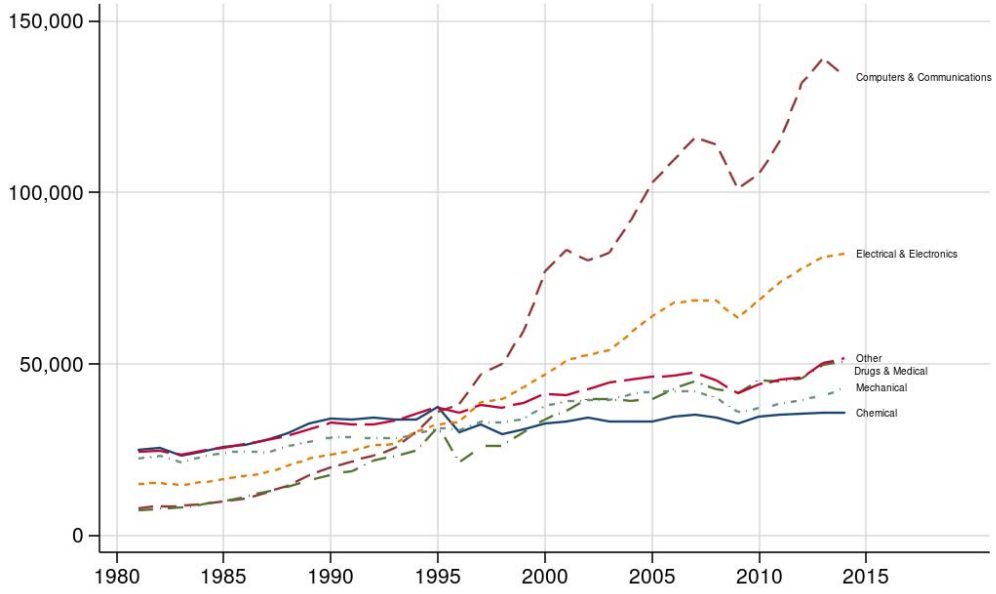


Since about 1980, the number appellate court decisions rendered each year keeps pace with the stock of patents-in-force.

Figure 8: Annual applications per patent-in-force (1840-2014).

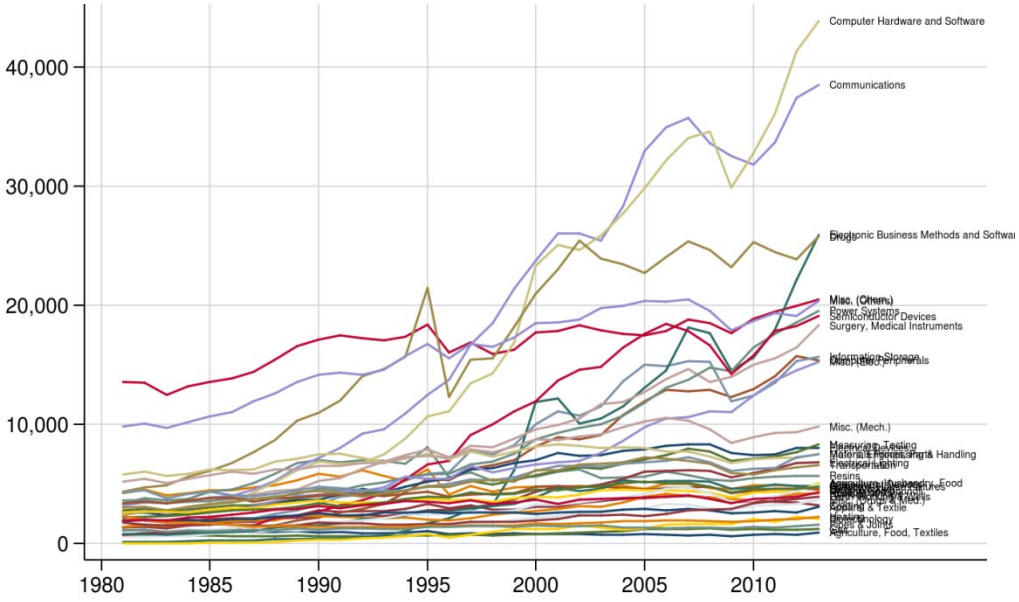
Normalizing by the number of patents-in-force reinforces the observation that the growth rate in applications over the last 20 years is high relative to the last century, but still lower than the mid- to late-19th century.

Figure 9: Annual applications (1981-2014) by NBER category.



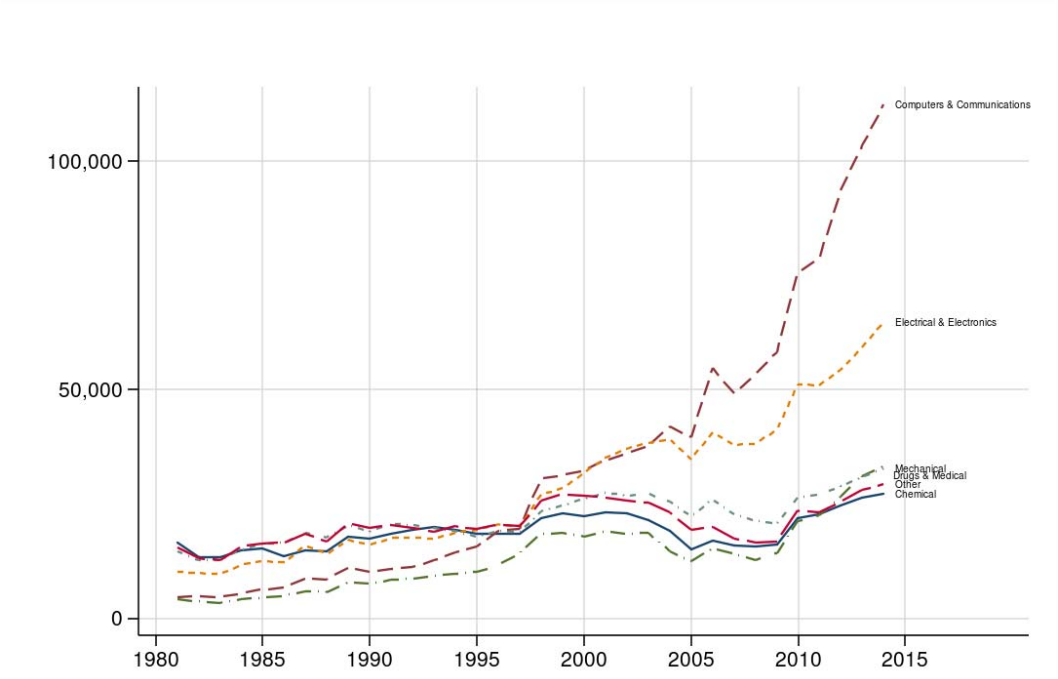
Annual filing rates vary widely by NBER category. Since the mid-1990s, filings in Computers & Communications and Electrical & Electronics have far outpaced other categories. The large swings in applications (especially within the Chemical and Drugs & Medical categories) surrounding the 1995 change in patent term highlights the differences across technologies in the responsiveness of application volumes to significant policy events or economic cycles. Likewise, applications in all NBER categories decreased following the recent financial crisis.

Figure 10: Annual applications by NBER sub-category.



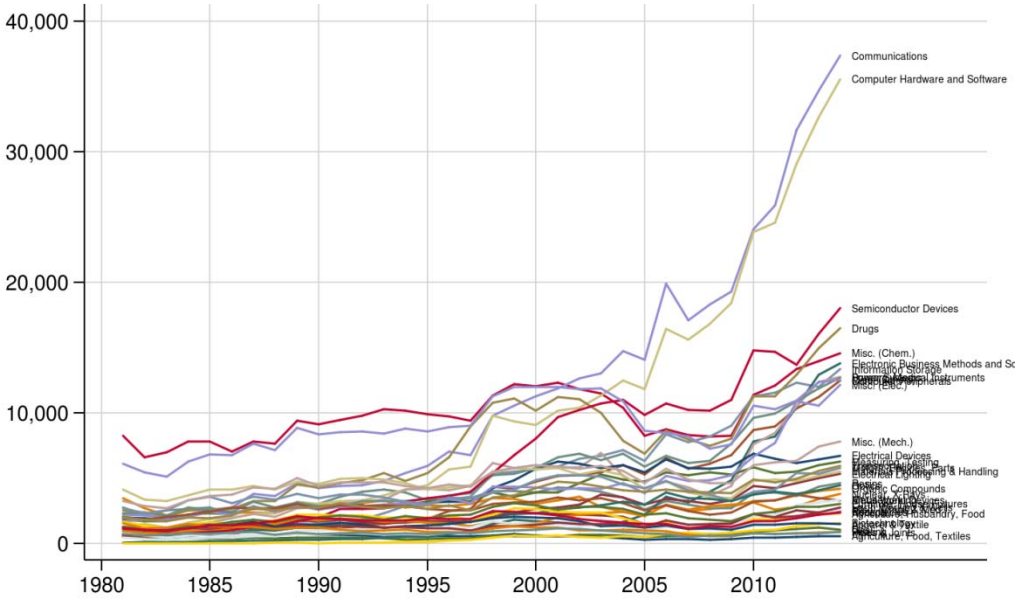
Further disaggregating annual applications by NBER sub-category shows specific drivers of growth within categories. See Appendix 1 for trends broken out by quartiles based on 2014 filings.

Figure 11: Annual patent grants by NBER category.



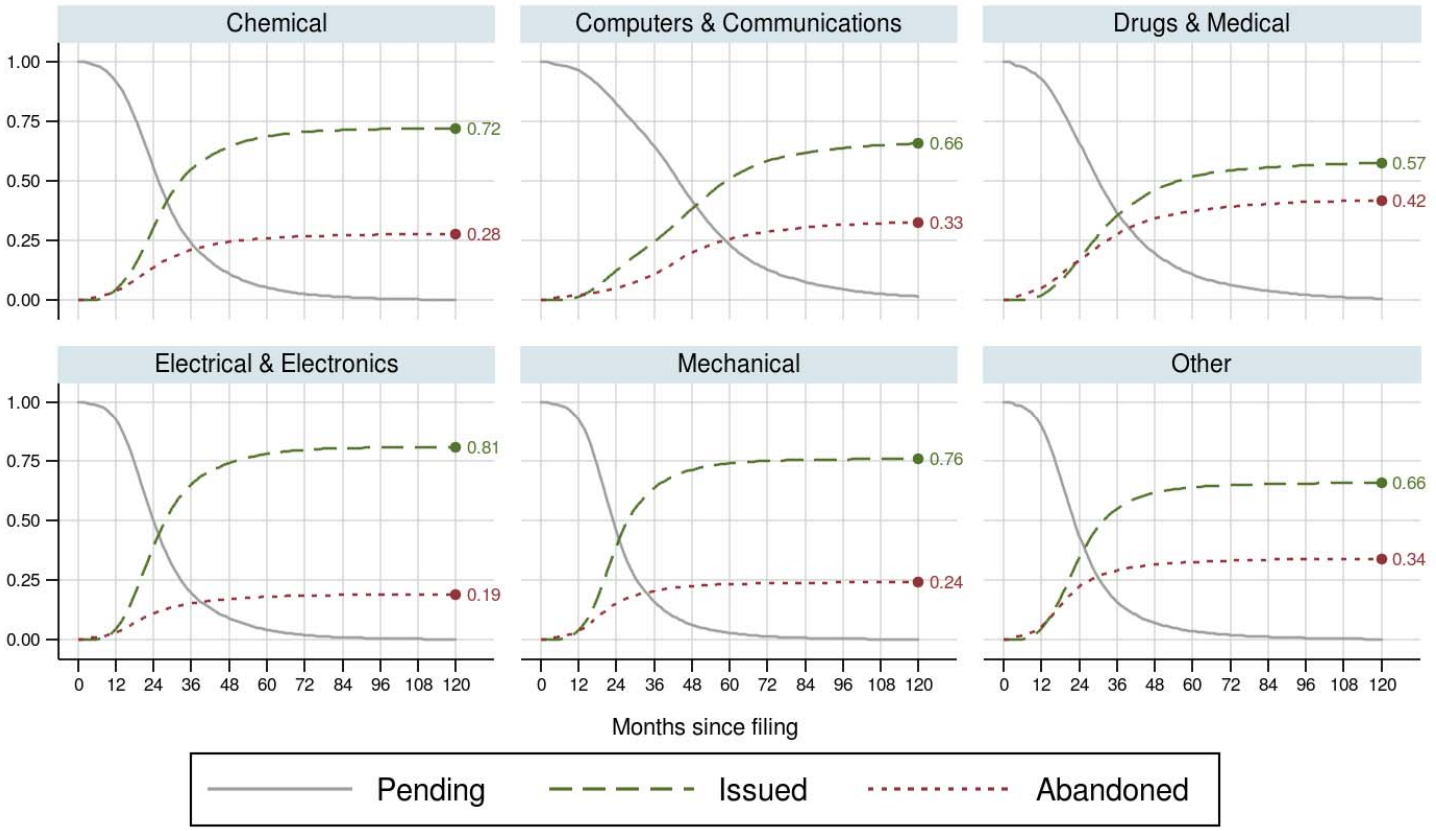
As with annual applications, patents in the Computers & Communications and Electrical & Electronics NBER categories vastly outnumber patents in all other categories beginning in the early 2000s. Interestingly, the number of patents in these other categories has risen sharply and leveled off in recent years.

Figure 12: Annual patent grants by NBER sub-category.



Further disaggregating annual patents by NBER sub-category shows specific drivers of growth within categories. See Appendix 2 for trends broken out by quartiles based on 2014 grants.

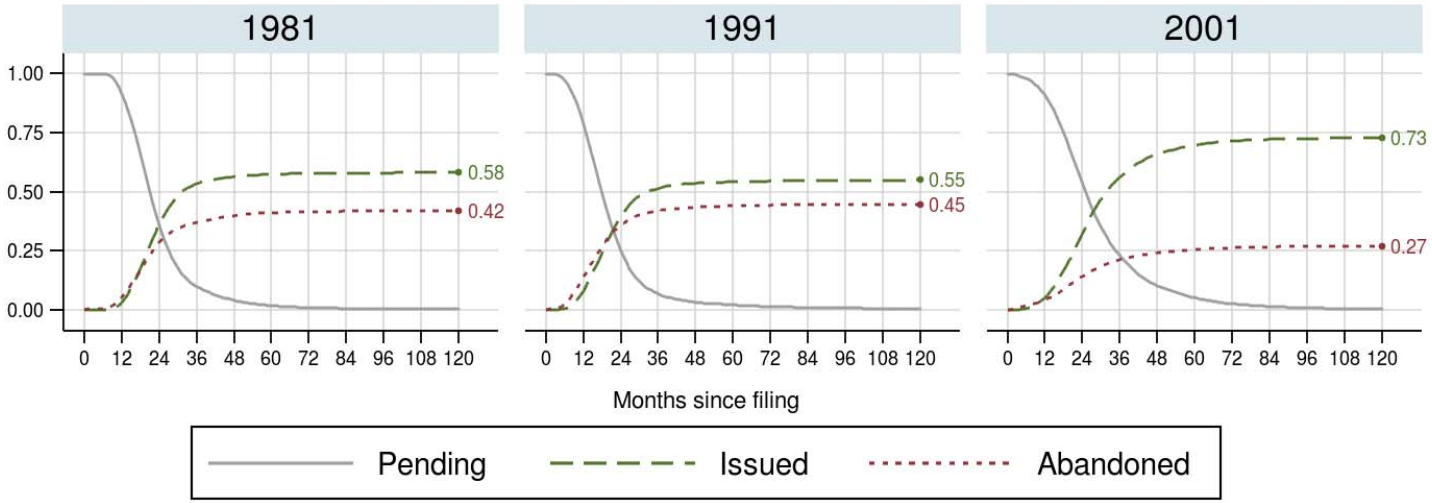
Figure 13: Cumulative disposal proportion by NBER category, January 2002 cohort.



* We thank Charlie DeGrazia for his excellent research assistance with this project. The views expressed are those of the individual authors and do not necessarily reflect official positions of the Office of Chief Economist or the U. S. Patent and Trademark Office.

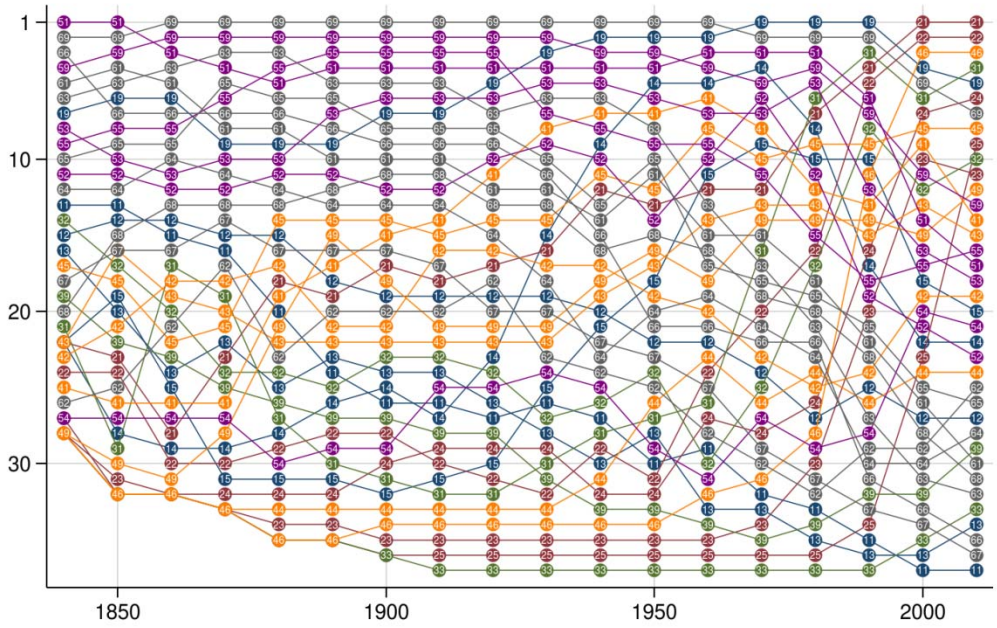
Allowance rates for NBER categories are not stable across application cohorts or relative to each other. In addition, the rate of abandonment exceeds that of issuance for all NBER sub-categories in the months immediately following application; this trend persists for some NBER categories longer than others.

Figure 14: Cumulative disposal proportion for NBER category 1 (Chemical); January 1981, 1991, 2001 cohorts.



Allowance rates may vary over time for a single NBER sub-category, in addition to variation across NBER categories within a single cohort).

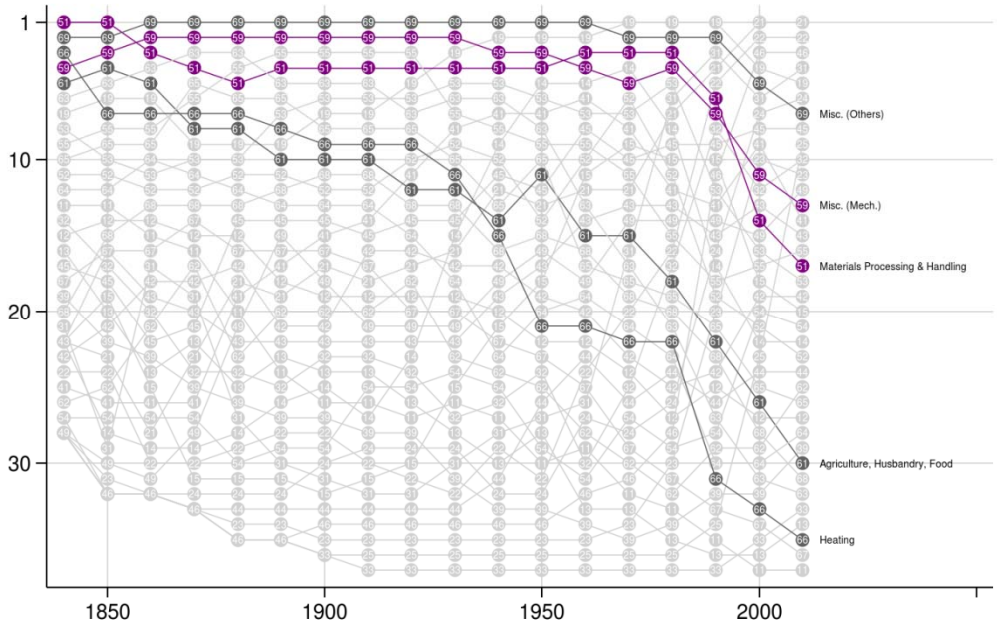
Figure 15. NBER sub-category rank by number of patents issued in the associated decade.



The growth of patents in each NBER sub-category evolves dynamically over time. See the following figures for targeted views of selected sub-categories.

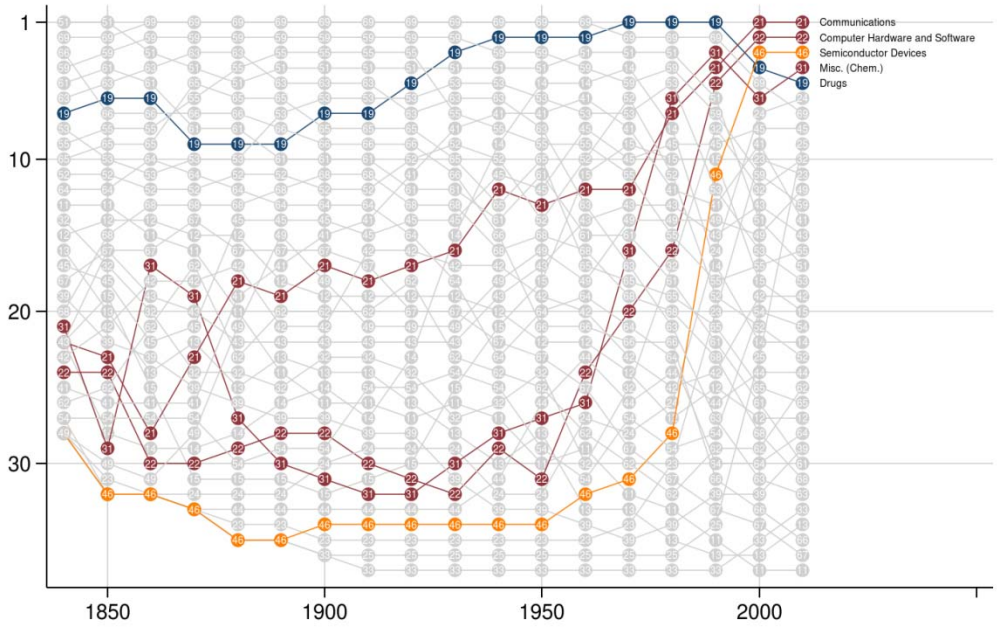
* We thank Charlie DeGrazia for his excellent research assistance with this project. The views expressed are those of the individual authors and do not necessarily reflect official positions of the Office of Chief Economist or the U. S. Patent and Trademark Office.

Figure 16: NBER sub-category rank by number of patents issued in the associated decade (with top five of 1840-1849 highlighted).



NBER sub-categories Miscellaneous Others (sub-category 69), Miscellaneous Mechanical (59), Materials Processing and Handling (51), Agriculture, Husbandry, Food (61) and Heating (66) remain firmly in the top 10 until the early part of the 20th century, after which each sub-category's ranking falls over time with different rates of decline. Miscellaneous Others remains in the top 10 through 2014, and Miscellaneous Mechanical and Materials Processing and Handling only fall out of the top ten in the early 2000s. Agriculture, Husbandry, Food, and Heating, however, fall steadily throughout the 20th century to the bottom tier by 2014.

Figure 17: NBER sub-category rank by number of patents issued in the associated decade (with top five of 2010-2014 highlighted).



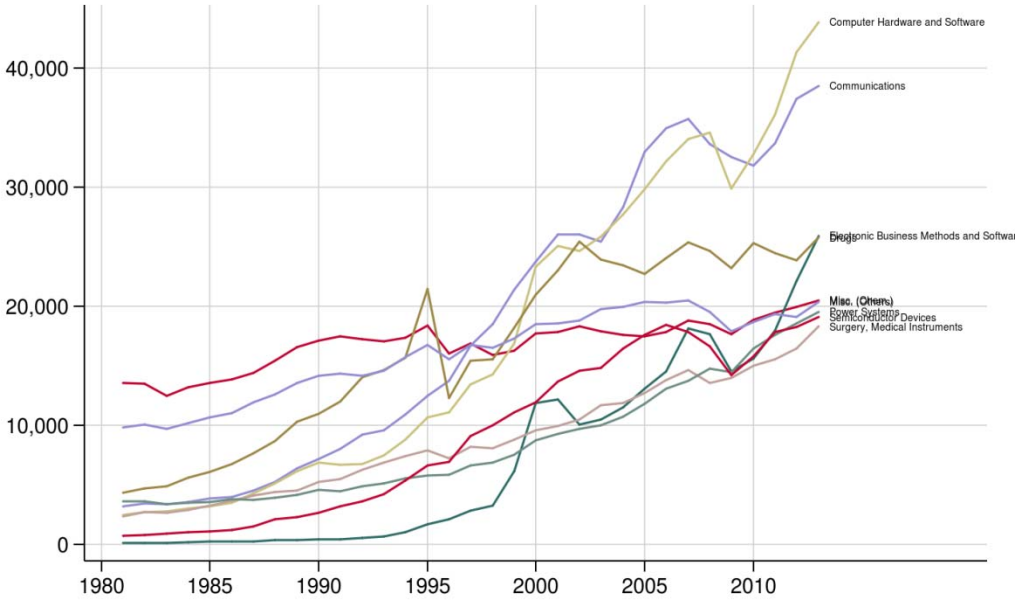
Patenting associated with Semiconductor Devices (46), Drugs (31) and Computer Hardware and Software (22) was minimal until the second half of the twentieth but climbed rapidly thereafter. Conversely, over the past century, the number of Communications patents issued has steadily increased relative to other sub-categories, and Miscellaneous Chemical patents remain among the top sub-categories through the entire time series.

Figure 18. Gini coefficient over grant year, by classification scheme.

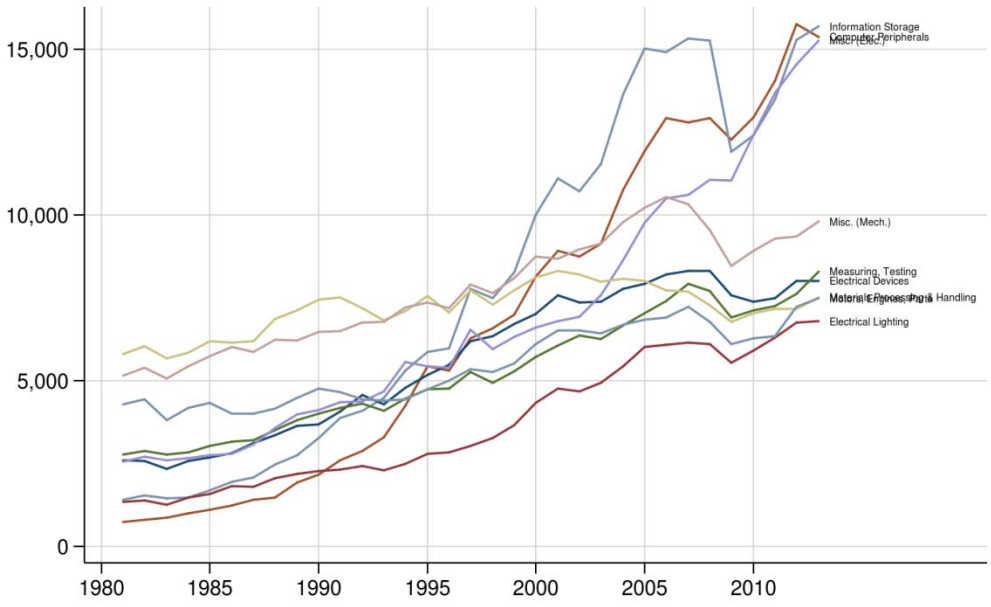
The Gini coefficient, in the context of patents, shows the concentration of grants across patent classes. All three classification schemes show increasing concentration in the most recent decades, but divergent trends over time, notably between the 1950s and 1990s. Although current concentration levels are high, they are not unprecedented.

Appendix 1: Applications from 1981 to 2014 by NBER sub-category broken out by quartiles based on 2014 filings

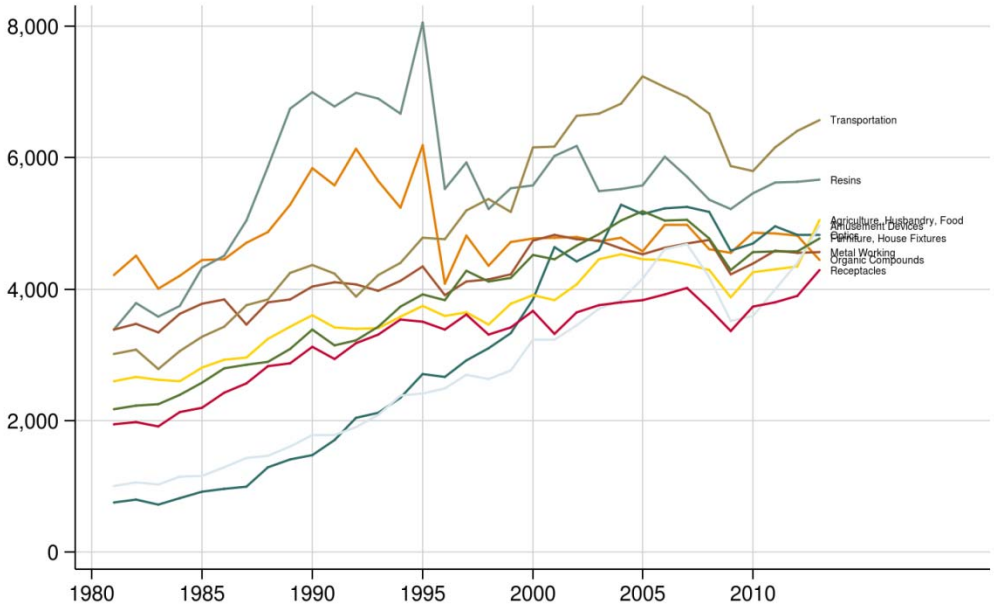
Quartile 1 (NBER sub-categories with the most filings in 2014):



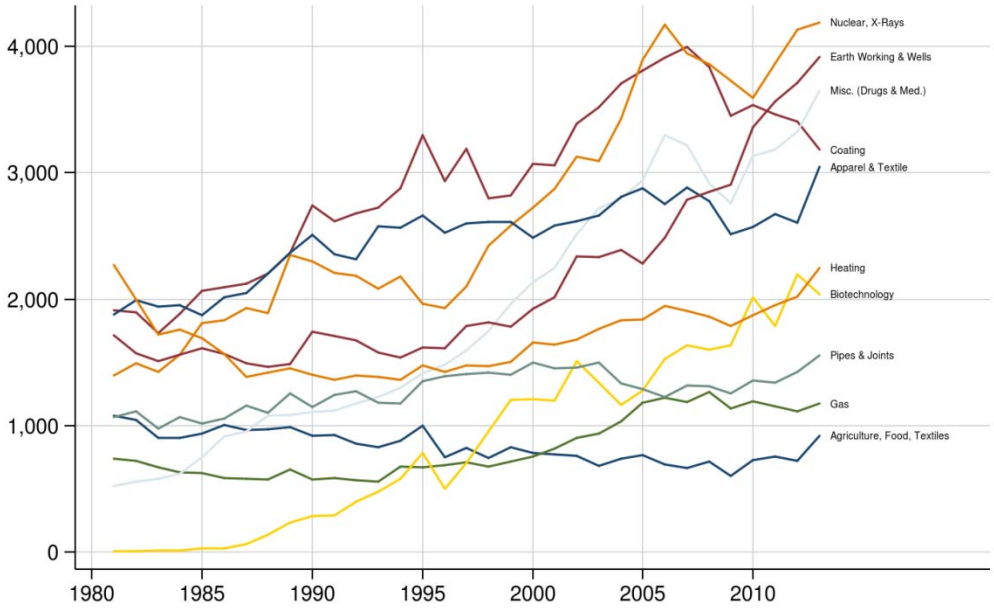
Quartile 2 (NBER sub-categories with the second most filings in 2014):



Quartile 3 (NBER sub-categories with the third most filings in 2014):

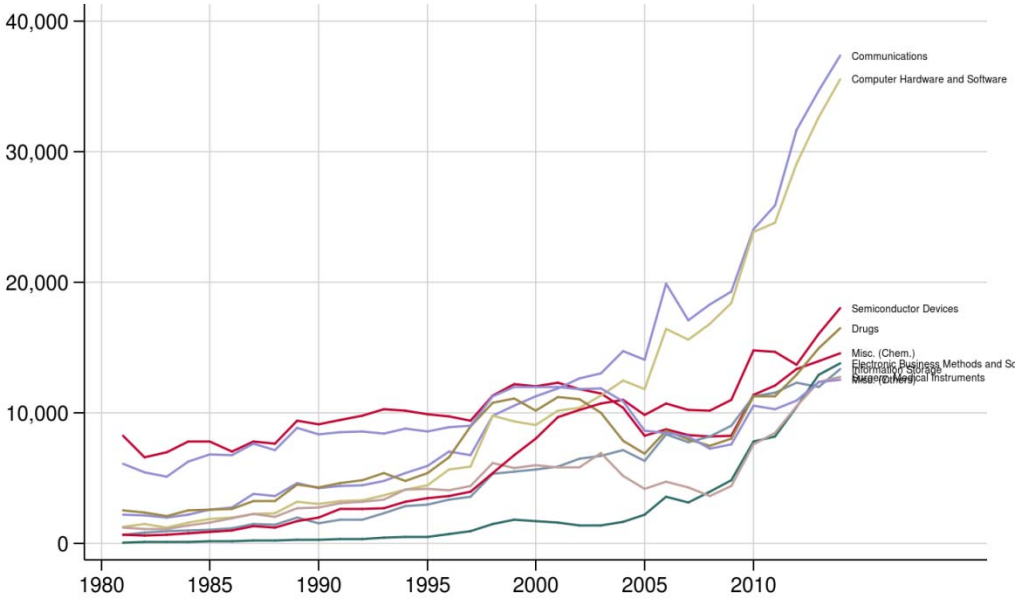


Quartile 4 (NBER sub-categories with the fourth most filings in 2014):

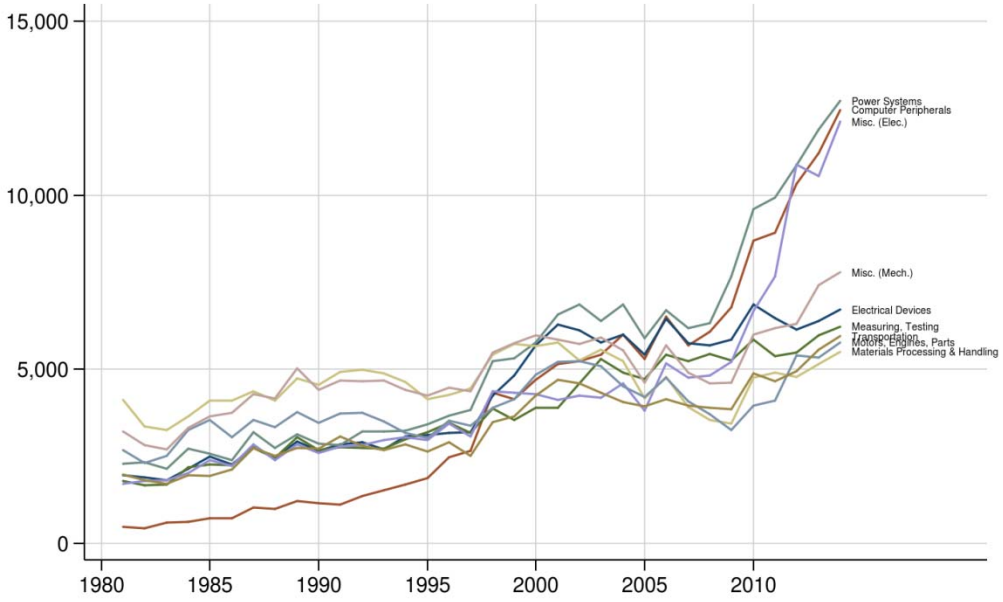


Appendix 2: Patent grants from 1981 to 2014 by NBER sub-category broken out by quartiles based on 2014 filings

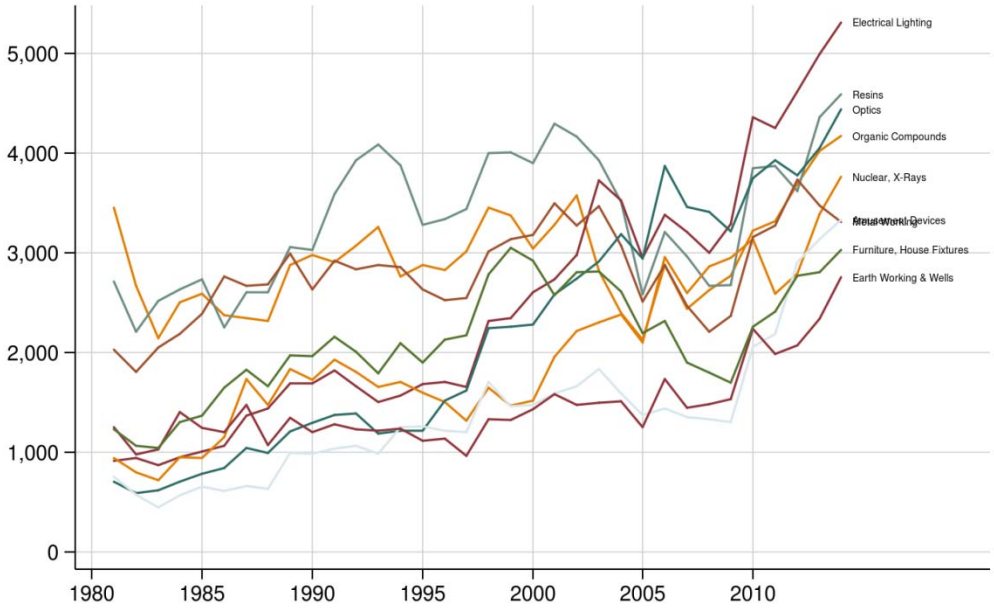
Quartile 1 (NBER sub-categories with the most patent grants in 2014):



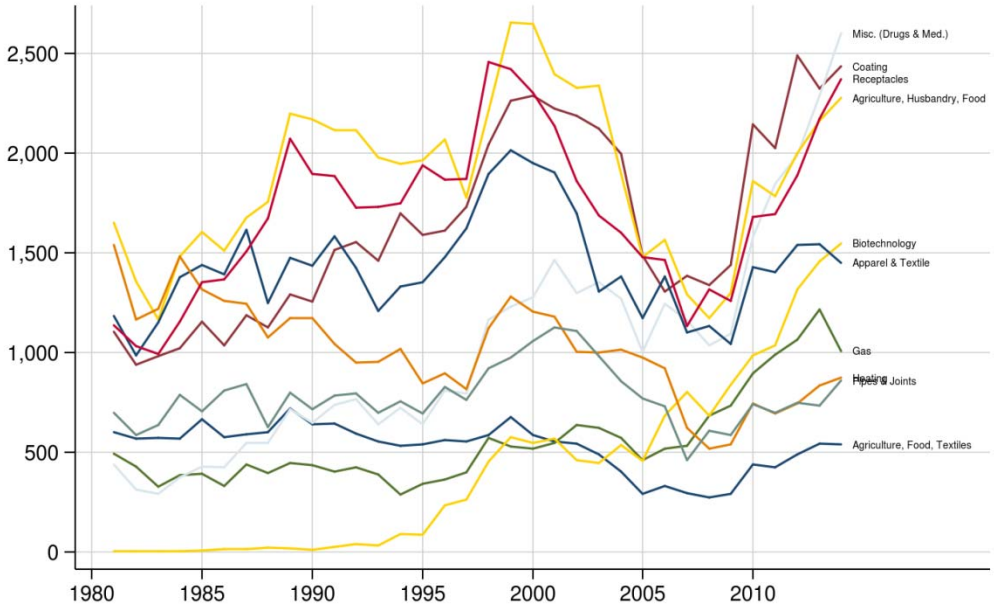
Quartile 2 (NBER sub-categories with the second most patent grants in 2014):



Quartile 3 (NBER sub-categories with the third most patent grants in 2014):

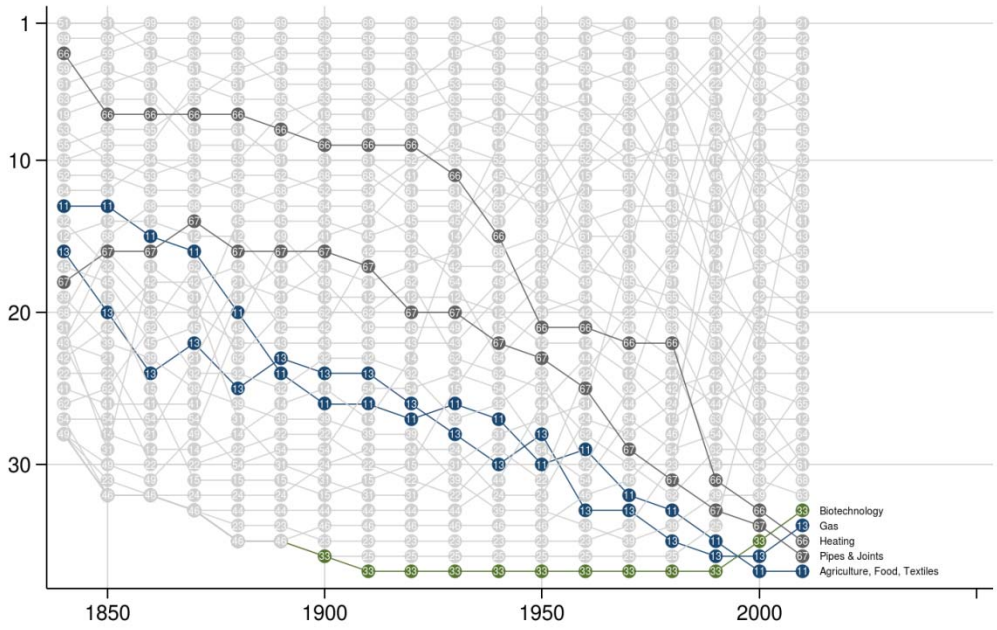


Quartile 4 (NBER sub-categories with the fourth most patent grants in 2014):

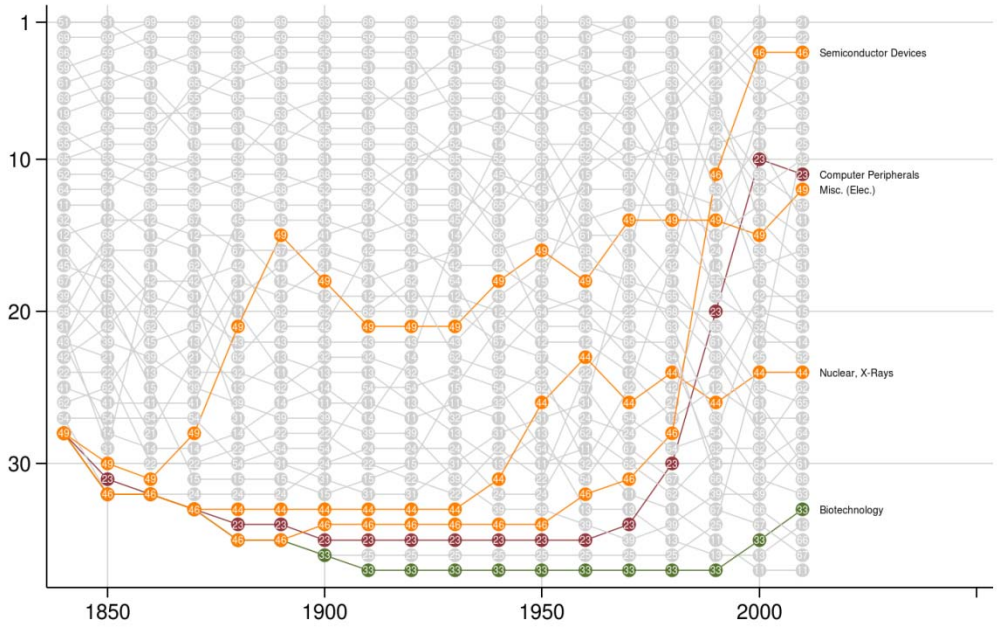


Appendix 3: Additional NBER sub-category rank plots

NBER sub-category rank by number of patents issued in the associated decade (with bottom five of 2010-2014 highlighted):



NBER sub-category rank by number of patents issued in the associated decade (with bottom five of 1840-1849 highlighted):



Appendix 4: Codebook

The table below outlines the contents of each data file (where a variable's inclusion is indicated by an ×):

Variable	annual	monthly	monthly_ disposal	historical_ masterfile
appl_dt				×
appl_id				×
appl_month			×	
calendar_month			×	
disp_dt				×
disp_ty				×
exp_dt				×
exp_dt_max				×
month		×		
months_since_filing			×	
nber				×
patent				×
prior_dt				×
pta				×
pub_dt				×
pubno				×
[total|nber#[#]]_abn		×	×	
[total|nber#[#]]_app	×	×		
[total|nber#[#]]_inforce	×	×		
[total|nber#[#]]_inforce_max	×			
[total|nber#[#]]_iss	×	×	×	
[total|nber#[#]]_pen		×	×	
uspc[_sub]				×
year	×			

The table below describes all of the variables included in this data release:

Name	Variable	Values or Explanation
appl_dt	Application date	Note: Stata counts and presents daily values as the number of days since January 1, 1960; users of other programs may see the value 7671 for January 1, 1981, 7702 for February 1, 1981, and so on.
appl_id	Application serial number	See note for appl_dt
appl_month	Application month	Note: Stata counts and presents monthly values as the number of months since January 1960; users of other programs may see the value 252 for January 1981, 253 for February 1981, and so on. Cohort identifier for **monthly_disposal** data file.
calendar_month	Calendar month	See note for appl_month
disp_dt	Disposal date	See note for appl_dt
disp_ty	Disposal type	ABN = Abandoned ISS = Issued PEN = Pending
exp_dt	Expiration date	See note for appl_dt
exp_dt_max	Upper limit of expiration date	Intentionally missing for patents not granted during the 1836-1860 period.
month	Month	See note for appl_month
months_since_filing	Months since filing	
nber	NBER sub-category	See **Table 2** for details
patent	Patent number	Values pre-fixed by “X” were recovered from the 1836 fire; values pre-fixed by “-“ are placeholders for non-granted applications (and equal to “-“ plus the application ID number)
prior_dt	Priority filing date	See note for appl_dt
pta	Patent term adjustment	Measured in days
pub_dt	PGPub date	See note for appl_dt
pubno	PGPub number	
[total|nber#[#]]_abn	Applications abandoned during month/year: total or NBER [sub-]category #[#]	
[total|nber#[#]]_app	Applications filed during month/year: total or NBER [sub-]category #[#]	Excludes provisional applications, PCT applications, RCEs and other non-serialized applications
[total|nber#[#]]_inforce	Estimated patents-in-force within year: total or NBER [sub-]category #[#]	Measured at the end of the time interval (month or year)
[total|nber#[#]]_inforce_max	Upper limit of estimated patents-in-force within year: total or NBER [sub-]category #[#]	Intentionally missing for patents not granted during the 1836-1860 period

[total|nber#[#]]_iss	Applications issued during month/year: total or NBER [sub-]category #[#]	Excludes reissues
[total|nber#[#]]_pen	Applications pending at end of month/year: total or NBER [sub-]category #[#]	Measured at the end of the time interval (month or year); intentionally missing in the monthly dataset until February 1987.
uspc[_sub]	Current USPC class [subclass]	
year	Year	