Update to the NBER patent data file: The role of company names and ownership changes in the dynamic reassignments of patents

Ashish Arora[1] Sharon Belenzon[2] Lia Sheer[3]

We introduce a major data extension and improvement to the historical NBER 2006 patent dataset, which should be valuable for all researchers working with patent data. In updating the data to match between Compustat and patents to 2015, we address two major challenges: name changes and ownership changes. These challenges are central to how patents are assigned to firms over time. To be consistent over the sample period, we reconstruct the complete historical data covered in the NBER data files. About 30% of the Compustat firms in our sample change their name at least once. Accounting for name changes improves the accuracy and scope of matches to patents (and other assets), ownership structure and dynamic reassignments of GVKEY codes to companies. Dynamic reassignment means that, for instance, if a sample firm is taken over in a reverse acquisition, the patents of the acquirer are included in the stock of patents linked to the Compustat record from that point onward, but not before.

The following example illustrates the problem posed by name changes and how it affects the existing NBER data file. The data documentation presents details of our data work including additional examples that underscore our numerous improvements to the existing NBER patent data file.

Example: SEALED POWER and GENERAL SIGNAL

This example underscores the consequences of not accounting properly for name and ownership changes.

Up until 1998, SEALED POWER and GENERAL SIGNAL are two distinct entities. Historical Compustat records include the following records for these companies up to 1998:

1) GVKEY 9556, name:
   i) SEALED POWER CORP (1962-1988) – original name
   ii) SPX CORP (1988-1997) -name changes retroactively in Compustat

2) GVKEY 5087, name: GENERAL SIGNAL CORP (1950-1997)

In 1998, SPX Corp acquires General Signal Corp in a reverse merger transaction and General's GVKEY (5087) became the new security of SPX traded retroactively under the new name "SPX CORP". At the same time, the original SPX records are renamed retroactively in Compustat as "SPX CORP-OLD" and stopped being traded. Current (at the time of matching) Compustat records include the following records for these companies for the complete period they are traded:

1) GVKEY 9556, name: SPX CORP-OLD

2) GVKEY 5087, name: SPX CORP

Our approach is to treat these GVKEYs as two separate companies up to 1997 accounting for all relevant names (SEALED POWER CORP, SPX CORP for GVKEY 9556 and GENERAL SIGNAL CORP for GVKEY 5087) in our matches and to dynamically connect the SPX CORP name to General's original GVKEY (5087) only from 1998.

When we examine NBER 2006 patent dataset, we find that the two companies are collapsed under the same company (same PDPCO id) and that for the purpose of Compustat accounting information General's original GVKEY (5087) is used for the complete period while the original SPX GVKEY (9556) is disregarded:

[1] Duke University, Fuqua School of Business and NBER
[2] Duke University, Fuqua School of Business and NBER
[3] Duke University, Fuqua School of Business

Data for SPX Corp in NBER 2006

current name	gvkey	firstyr	lastyr	pdpco	pdpseq	begyr	endyr
SPX CORP	5087	1950	2006	5087	1	1950	2006
SPX CORP-OLD	9556	1962	1997	5087	-1		

Practically, this means that all the patents of SPX CORP are matched to General Signal's financial data up to 1998. To verify we tracked the NBER files and confirmed that indeed SPX patents pre-1998 are matched to General's GVKEY. On the other hand, patents related to "GENERAL SIGNAL CORP" (757 patents without considering related subsidiaries) as well as "SEALED POWER CORP" (36 patents without considering related subsidiaries) are located in the 2006 NBER raw patent match but are not assigned to any Compustat record.

The NBER patent data file does not track ownership of GVKEYs and name changes over time. However, as shown in this example, using the current Compustat name can be misleading. Availability of data on historical name changes enables us to have a better understanding of the firms included in the Compustat sample and their origin. We are able to improve the accuracy of the NBER and linkage to relevant financial data for the period 1980-2015.

Our data is made publicly available and we will be honored by the opportunity of presenting it to the NBER innovation community.

DATA DOCUMANTATION OF THE NBER PATENT DATA UPDATE

This document describes the methodology used to construct our database of publicly listed American firms matched to assignees of patents from the United States Patent and Trademark Office (USPTO) for the period 1980-2015.

We combine data from 5 main sources: (i) company and accounting information from U.S. Compustat 2018, (ii) patents data from PatStat; (iii) subsidiary data from ORBIS; (iv) mergers and acquisition data from SDC platinum and (v) company name changes from WRDS's "CRSP Monthly Stock".

We match: (i) corporate subsidiaries to Compustat ultimate owner (UO) firms; (ii) acquisition data to Compustat companies and their related subsidiaries and (iii) patent data to Compustat companies and their related subsidiaries. We discuss the details of our methodology below.

A. ACCOUNTING DATA PANEL

Our methodology builds and improves on the NBER 2006 patent database (Hall et al., 2001), by extending the time period by a decade (now from 1980 to 2015) and implementing several methodological improvements for the complete sample period.

We start with all North American Compustat records obtained through WRDS in August 2018 and select companies with active records and positive R&D expenses for at least one year during our sample period, 1980-2015[4]. We exclude firms that are not headquartered in the United States based on their current headquarter location. After matching the remaining firms to patent assignees from the USPTO, we further restrict our sample to ultimate owner (UO) Compustat firms with at least one patent during our sample period. A UO firm enters the sample once it is publicly traded and remains in our data until the end of the sample period, unless it is acquired, dissolved, or taken private. All UO firms in our final sample have at least 3 consecutive years of active records in Compustat. Our UO firm level data are supplemented with subsidiary level data. Our final estimation sample consists of an unbalanced panel of 4,090 UO firms and 54,274 firm-year observations. Our unique UO firm identifier in the accounting data panel is labeled as PERMNO_ADJ_LONG. The process of defining an UO firm and its related subsidiaries is explained below.

We face several challenges when working with Compustat data, as following.

1) **Unique company identifier over time.** Compustat uses GVKEY to track companies over time[5]. However, a single company may correspond to multiple GVKEYs within the Compustat database due to changes in ownership and other accounting changes over the sample period (e.g., the pet food company Ralston Purina is listed under 2 different GVKEYs: (i) 1980-1993 under "RALSTON PURINA-CONSOLIDATED" (GVKEY 008935) and (ii) 1993-2000 under "RALSTON PURINA CO" (GVKEY 028701)). The Compustat database does not link related company identifiers making it difficult to track companies over time only based on GVKEY.

2) **Name changes.** While scientific publication and patent records contain the owner name at the time of their publication, companies appear in the Compustat file under their most current name with no records of previous names. Company names may change over the course of our sample period due to general name changes[6] and M&As[7], including reverse takeovers[8]. A company with a name change (which might have been accompanied with an ownership change) without a corresponding change in its GVKEY in Compustat may cause us to assign the record incorrectly to its most recent holder for all sample years. Without historical information on the

[4] We define an active record as a year with positive common shares traded (CSHTR_F). We do this to avoid including years with data based on prospectus submitted by the focal company as part of the filing process before the firm became publicly traded.
[5] GVKEY code remains the same, regardless of changes in TICKER, CUSIP, and firm names and thus is preferred on the later as a firm identifier for Compustat records. Compustat database only provides the most recent TICKER, CUSIP and name for each security with no historical info available.
[6] e.g., name abbreviations (for example, "APPLIED MOLECULAR GENETICS INC" changed its name in 1983 to "AMGEN INC"),
[7] e.g., "WESTINGHOUSE ELECTRIC CORP" (GVKEY 011436) purchased "CBS INC" in 1995 and changed its own name to "CBS CORPORATION" in 1997 keeping the same GVKEY Compustat firm identifier.
[8] e.g., in 1993 the private company Dentsply International Inc acquired the public company GENDEX CORPORATION (GVKEY 013700) in a reverse takeover and became publicly traded under the "DENTSPLY INTERNATIONAL INC" name and the original GVKEY.

record's ownership, we cannot correctly link patents and scientific publications to their relevant financial records.

3) **Ownership structure.** A parent company and a majority owned subsidiary may have different identification numbers and records within Compustat. While innovative activities typically take place inside numerous subsidiaries, we aggregate the data to the UO level. Since Compustat database does not link parent companies and majority owned publicly traded subsidiaries, extant manual checks are required.[9] We further link non-publicly traded subsidiaries to their UO firm based on ORBIS data.

4) **Changes in ownership.** Ownership of a firm can change throughout the sample period due to mergers, acquisitions and spinoffs[10]. While firms typically stop being traded independently after an M&A, their existing stock of publications and patents must be reassigned to the new owner. Moreover, in many cases, the acquiring entities continue to file patents and produce scientific publications post-acquisition. Compustat data do not provide information on ownership changes, thus we rely on SDC Platinum's M&A data to dynamically match our patent and publication data.

We implement the following procedures to manage these challenges.

I. NAME CHANGES

One of our key contributions is identifying name changes of Compustat firms over the sample years 1980-2015. To the best of our knowledge, this has not been done consistently for a broad range of companies across many industries over a third of a century. Past research mainly considers the name that appears for each record in the most recent Compustat file (CONM variable) as the relevant name for the complete period the security was traded. The variable CONM, however, is the current name of the Compustat record as of the date the file was downloaded with no historical name information provided by Compustat. As shown above, company name changes may not be accompanied by changes in the original GVKEY firm identifier on Compustat, leading to assigning a record to its most recent holder for the complete sample period. Instead of building on the recent Compustat name, we link our Compustat records to WRDS's "CRSP Monthly Stock" file, which records historical names for each month the security was traded and perform extensive manual checks using SEC filings to validate all related names for our sample period. We find that in our sample 30 percent of Compustat records have more than one related name[11]. Accounting for all historical names significantly improves the accuracy and scope of the matches we perform across various databases as well as the linkage to relevant financial data. We elaborate on our name change methodology below, using several examples.

Example 1: SEALED POWER and GENERAL SIGNAL

The following example illustrates the problem posed by name changes and how it affects the existing NBER data file. This example underscores the consequences of not accounting properly for name and ownership changes.

Up to the year 1998, SEALED POWER and GENERAL SIGNAL are two distinct entities. Historical Compustat records include the following records for these companies up to 1998:

1) GVKEY 9556, name:

   i. SEALED POWER CORP (1962-1988) – original name
   ii. SPX CORP (1988-1997) -name changes retroactively in Compustat

2) GVKEY 5087, name: GENERAL SIGNAL CORP (1950-1997)

[9] e.g., Thermo Electron's publicly traded majority owned spun-out subsidiaries (all of which returned to be privately owned after 1999) need to be accounted under the parent company THERMO ELECTRON CORP (GVKEY 010530) for the complete period.
[10] e.g., "AT&T CORP" (GVKEY 001581) stopped being traded independently in 2005 after it was acquired by "SBC COMMUNICATIONS INC" (GVKEY 009899) which in turn changed its own name to "AT&T INC".
[11] This is comparable to the findings of Wu (2009), who finds that during 1925-2000 over 30% of CRSP-listed firms changed their names at some point after going public. For name changes occurring between 1980-2000 the paper finds that the top 3 reason for name changes are: (i) M&As & restructure activity (36%); (ii) Change in focus of operation (17%); (iii) brand or subsidiary name adoption (12%)

In 1998, SPX Corp acquired General Signal Corp in a reverse merger transaction and General's GVKEY (5087) became the new security of SPX traded retroactively under the new name "SPX CORP". At the same time, the original SPX records are renamed retroactively in Compustat as "SPX CORP-OLD" and stopped being traded. Current Compustat records include the following records for these companies for the complete period they are traded:

1) GVKEY 9556, name: SPX CORP-OLD

2) GVKEY 5087, name: SPX CORP

Our approach is to treat these GVKEYs as two separate companies up to 1997 accounting for all relevant names (SEALED POWER CORP, SPX CORP for GVKEY 9556 and GENERAL SIGNAL CORP for GVKEY 5087) in our matches and to connect the SPX CORP name to General's original GVKEY (5087) only from 1998.

When we examine NBER 2006 patent dataset, we find that the two companies are collapsed under the same company (same PDPCO id) and that for the purpose of Compustat accounting information General's original GVKEY (5087) is used for the complete period while the original SPX GVKEY (9556) is disregarded:

Table A1. Data for SPX Corp in NBER 2006

current name	gvkey	firstyr	lastyr	pdpco	pdpseq	begyr	endyr
SPX CORP	5087	1950	2006	5087	1	1950	2006
SPX CORP-OLD	9556	1962	1997	5087	-1		

Practically, this means that all the patents of SPX CORP are matched to General's financial data up to 1998. To verify we tracked the NBER files and confirmed that indeed SPX patents pre-1998 are matched to General's GVKEY. Moreover, patents related to "GENERAL SIGNAL CORP" (757 patents without considering related subsidiaries) as well as "SEALED POWER CORP" (36 patents without considering related subsidiaries) are located in the 2006 NBER raw patent match but are not assigned to any Compustat record.

The NBER patent data file does not track ownership of GVKEYs and name changes over time. However, as shown in this example, using the current Compustat name can be misleading. Availability of data on historical name changes enables us to have a better understanding of the firms included in our sample and their origin. We are able to improve the accuracy of their match to the different databases (by using the complete history of firm names) and their linkage to relevant financial data. For this reason, we decided not to use NBER data at all and reconstruct the entire database for the complete sample period (1980-2015).

Compiling historical names

To locate historical names, we use the WRDS's "CRSP Monthly Stock" file, which includes historical monthly information on names for each security alongside its historical CUSIP code and a unique permanent security identification number assigned by CRSP, the PERMNO code, which is kept constant throughout the trading period regardless of changes in name or capital structure.[12] We compute for each name the starting and end year based on their trading dates in the "CRSP Monthly Stock" file.

Using WRDS "CRSP/Compustat Merged Database - Linking Table" we link each PERMNO to Compustat GVKEY code. As shown above a PERMNO can have multiple GVKEYs related to it, however CRSP also has cases where under the same GVKEY we find several PERMNO codes mainly due to significant M&As that occurred during the years or periods the firm was unlisted. Our main firm identifier PERMNO_ADJ builds on the original CRSP PERMNO id with several adjustments. (i) In cases where under the same GVKEY we find several PERMNO codes we replace it with one main PERMNO code[13] – for example, OWENS Corning GVKEY (008214) was split to two PERMNO codes 24811 and 91531 due to it being unlisted between 2003-2005, however we keep PERMNO_ADJ

[12] For example, while SPHERIX INC is related to 2 different GVKEYs (002237 for 1980-2013 and 018738 for 2013-current) it has a unique PERMNO code for the entire period (18148). Similarly, Google Inc PERMNO code is 90319 and it remains the same after the company reorganized as ALPHABET INC in 2015.

[13] In the final accounting data panel, we split firms based on big jumps in sales, patents or publications. For example, we split PERMNO_ADJ 66093 to the period before and after SBC Communications Inc acquired AT&T Corp and became AT&T Inc. PERMNO_ADJ_LONG is the final UO identifier in the accounting data panel after the split.

the same for the complete period (24811). (ii) We manually add a PERMNO_ADJ code for firms in our Compustat sample that did not appear in the "CRSP Monthly Stock" file[14].

We perform extensive manual checks on the name list, including (i) identifying companies with similar names[15]; (ii) adding from SEC fillings additional names that are missing from "CRSP Monthly Stock" file; and (iii) adjusting names manually – the "CRSP Monthly Stock" tends to abbreviate long words in the company name that it provides. We located those cases and manually corrected them to avoid mismatches.[16]

Standardizing firm names

Prior to matching, we standardize firm names to reconcile company names that may be spelled differently across the databases. We compose a standardization code used on both the source and the target names to increase the number of exact matches.

Each company name was first standardized by converting all strings to uppercase characters and cleaning all non-alphabetic characters as well as Compustat related indicators (e.g., -OLD, -NEW, -CL A) and other common words (e.g., THE).

Additionally, an important step in standardizing the company names is standardizing abbreviations. We formed a list that includes over 80 abbreviated words matched to their various original words. For example, LABORATORIES, LABORATORY, LABS, LABO, LABORATORIE, LABORATARI, LABORATARIO, LABORATARIA, LABORATORIET, LABORATORYS and LABORATORIUM were all abbreviated to "LAB". The list was compiled from the most frequently abbreviated words in WOS affiliation field (accordingly, the list is targeted to our sample). This list is presented in Table A2.

Table A2. Most frequent abbreviated words

ADV	AEROSP	AGR	AMER	ANAL	ANALYT	ANIM	APPL	APPLICAT
ASSOC	AUTOMAT	BIOL	BIOMED	BIOPHARM	BIOSCI	BIOSURG	BIOSYS	BIOTECH
BIOTHERAPEUT	CHEM	CLIN	COMMUN	COMP	CORP	CTR	DEV	DIAGNOST
DYNAM	EDUC	ELECTR	ENGN	ENVIRONM	FAVORS	GEN	GENET	GRAPH
GRP	HLDG	HLTHCR	HOSP	INC	IND	INFO	INNOVAT	INST
INSTR	INTERACT	INTL	INVEST	LAB	LTD	MAT	MED	MFG
MICROELECTR	MICROSYS	MOLEC	NATL	NAVIGAT	NEUROSCI	NUTR	ONCOL	ORTHOPAED
PHARM	PHOTON	PHYS	PROD	RES	SCI	SECUR	SEMICOND	SERV
SFTWR	SOLUT	SURG	SYS	TECH	TEL	TELECOM	THERAPEUT	TRANSPORTAT

For each standardized name, we create a cleaner, fully-standardized name by omitting the legal entity endings and other general words (e.g. INC, CORP, LTD, PLC, LAB, PHARMACEUTICAL), where possible, to maximize match rates (e.g., "XEROX CORP" was standardized to "XEROX", "ABBOTT LABORATORIES" to "ABBOTT"). However, in cases where the company name is too short, generic or can match to other strings within the affiliation field, we preserved the original standardized name to avoid mismatches and extensive manual checks on the match results. For example, omitting the legal entity from "QUANTUM CORP" would result in a potential mismatch between "QUANTUM" and "TEXAS STATE UNIV CTR APPL QUANTUM ELECTR DEPT".

[14] There are differences between CRSP and Compustat coverage- for example, CRSP only includes firms listed in USA major exchanges and specifically excludes regional exchanges, while Compustat includes all 10-K filer firms in North America. Moreover, CRSP coverage for major exchanges has expanded gradually over the years (e.g., ARCA was only added from 2006).

[15] For instance, RACKABLE SYSTEMS INC (GVKEY 162907) changed its name to SILICON GRAPHICS INTL CORP after it acquired the public company SILICON GRAPHICS INC (GVKEY 012679) in 2009 – we need to make sure that we count SILICON GRAPHICS related publications and patents under Rackable's GVKEY only from 2009. Similarly, we need to distinguish between the original BIOGEN INC (GVKEY 002226) and the new BIOGEN INC (GVKEY 024468) that was formed only after the merger with IDEC PHARMACEUTICALS CORP in 2003.

[16] It is also worth mentioning that the "CRSP Monthly Stock" file reports acronym firm names with extra space between the initial letters (e.g., E G & G INC and not EG&G INC). This has to be taken into consideration when performing matches to other databases that do not use this format.

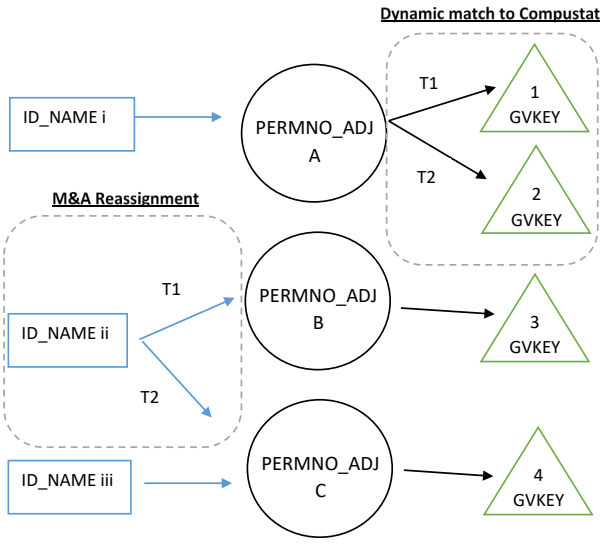
The last step in name standardization is to locate abbreviations that are commonly used by companies instead of their official name. For example, "INTERNATIONAL BUSINESS MACHINES CORP", will also appear under its common abbreviation "IBM" and ''GENERAL ELECTRIC CO" under "GE". We also add the names of prominent R&D laboratories affiliated with companies, such as the T.J. Watson Research Center (IBM) and Bell Labs (initially AT&T and later under Lucent technologies), as authors often omit the name of the company when the address of the laboratory is stated as the publication address.

Constructing the UO name list

All our matching is done at the firm name level. We assign each firm name a unique identifier ID_NAME and indicate the first and last year the name is relevant for a PERMNO_ADJ. We then perform dynamic matching of names to PERMNO_ADJ based on SDC's M&A data. M&A reassignment includes up to five reassignments per name over the sample period (explained in further details below). Our UO historical standardized name list "B2B_UO_name_list.dta" including the dynamic reassignment will become publicly available for researches to match to their database of interest. Main variable of file "B2B_UO_name_list.dta" are described below:

Variable name	
NAME_STD	Historical standardized UO firm names (1980-2015) for firms that were included in our initial Compustat sample[17]
ID_NAME	Name ID unique at name_std-permno_adj1
PERMNO_ADJ$_{0\text{-}5}$	Owner firm id: up to 5 owners + "0" is usually the pre-IPO owner if applicable.
NAME_ACQ$_{0\text{-}5}$	Owner name
FYEAR$_{0\text{-}5}$	First year assigned to owner
NYEAR$_{0\text{-}5}$	Last year assigned to owner

Figure A1. Description of dynamic changes



Note: This figure illustrates the dynamic structure of the data. The dynamic reassignment accounts for: (i) changes in Compustat identification numbers (GVKEY), and (ii) M&A reassignment. Each name (ID_NAME) can be assigned throughout the sample period to more than one firm (PERMNO_ADJ) and each firm can be linked to more than one Compustat record (GVKEY).

[17] The data includes only names of UO parent firms included in our initial Compustat sample. Exceptional are names of top laboratories and names of majority owned publicly traded subsidiaries that appeared in our initial Compustat sample and were collapsed under the UO parent firm. In order to include subsidiary data, the user should match the name list to subsidiary databases such as ORBIS. The standardization code that was used to standardize the names will be available under NAME_STD.do. Standardized names include legal entity and other common words - in cases where users want to match to a cleaner version of the name, they should apply their own script to further clean the names. When matching the name list to other databases users should include extensive manual inspection to matched result.

II. DYNAMIC REASSIGNMENT

We build on the strategy used by NBER patent match (2006) to perform a dynamic reassignment for our subset of UO Compustat firms (see Figure A1). The dynamic reassignment accounts for: (i) changes in Compustat identification numbers (challenge 1 above) - dynamically matching Compustat accounting information for firms that are related to more than one GVKEY record, and (ii) M&A reassignment based on SDC data and construction of a complete name history for the period 1980-2015 (Challenge 4 above). For M&A reassignment, we include up to five ownership reassignments for each firm name that appears in our initial Compustat subsample and was acquired by another firm in our sample. Unless a name is reassigned to another PERMNO_ADJ it stays with the focal firm until the end of sample (or the firm's trading period). We dynamically reassign related patents and scientific publications of the acquired UO firm and its related subsidiaries to acquirer firms accordingly (will be discussed in more detail below).

Each PERMNO_ADJ is then linked to Compustat GVKEYs. For cases where there are changes in Compustat identification numbers over the sample period, we dynamically match PERMNO_ADJ to GVKEYs. In the final accounting data panel, we further split firms based on big jumps in sales, patents or publications. PERMNO_ADJ_LONG is the final UO identifier in the accounting data panel after the split.

Example 2: TIME-WARNER and AMERICAN ONLINE

This example illustrates how properly accounting for name and ownership changes improves the accuracy of patent flow as well as the dynamic reassignment of patents.

Warner Communication and its subsidiaries were independent and publicly traded companies until their merger with Time Inc in 1989 when Time-Warner Inc was formed. In the second half of 2000 Time-Warner was merged with American Online to form AOL Time Warner. In 2003 the company dropped the "AOL" from its name and was renamed Time-Warner Inc. AOL remained a subsidiary until it was spun-out in 2009.

The NBER 2006 patent match reveals:

1) Warner Communication and its subsidiary related patents are correctly matched to WARNER COMMUNICATIONS INC (GVKEY 11284) up to the merger with Time Inc. However, they are not dynamically assigned after 1988 to Time Warner or any other company, implying that the patent stock and patent flow of Time-Warner (and later AOL Time-Warner) from patents related to Warner communication and its subsidiaries (e.g., Warner Bros, WEA Manufacturing (before it was acquired) – above 60 patents up to 2006) are below the true value after the acquisition in1989.

2) TIME-WARNER related patents from 1991 to 2000 (before the merger with American-Online Inc in late 2000) are matched incorrectly to GVKEY 25056, which during those years was solely AMERICAN-ONLINE INC original Compustat financial records. The current name of GVKEY 25056, TIME WARNER INC, which is likely to have misled NBER to link the Time Warner patents to it, was only adopted retroactively in 2003 when the "AOL" was dropped from the official name. Moreover, AMERICAN ONLINE INC and AOL related patents (152 patents up to 2006 based on NBER raw patent match) are not linked to any Compustat record. AOL-TIME WARNER related patents, on the other hand, are matched to a "Pro-Form" Compustat record that is active for only 2 years 1999-2000: AOL TIME WARNER INC-PRO FORM (GVKEY 142022). All of which imply that AOL Time Warner's flow of patents is below the true level throughout the period.

Having a complete history of names enables us to correctly identify each Compustat record and its origin and dynamically match each firm name in our sample to the correct financial records accordingly:

1) AMER ONLINE INC (and later AOL) is matched from 1980 until its spinout in 2009 to GVKEY 25056 and after to AOL INC (GVKEY 183920).

2) Warner Communication is matched up to the merger with Time Inc to WARNER COMMUNICATIONS INC (GVKEY 11284) and later dynamically transferred ending up in AOL -Time Warner GVKEY (25056) starting 2001.

3) AOL -Time Warner is matched to AOL -TIME WARNER (GVKEY 25056) starting 2001 after the merger was approved.

4) As a side note- Time Inc is not included as an UO in our sample as it did not have R&D expenses, but it is included as a subsidiary name under the Time-Warner UO company.

Case study 3: PHARMACIA & UPJOHN and MONSANTO

This example demonstrates that having a complete history of names enables us to correctly identify each Compustat record's historical ownership and dynamically match each firm name in our sample to its relevant financial records in each period. For instance, linking each patent to its correct financial record can be a concern for papers that link patents to market value, specifically those distinguishing different types (e.g., high vs. low cited patents), which rely on the specific patent that was matched and not only the quantity.[18]

In 1995 original Pharmacia merged with Upjohn to form Pharmacia & Upjohn. In 2000, original Monsanto merged with Pharmacia & Upjohn to form Pharmacia Corporation (New Pharmacia). Between 2000-2002 the new Pharmacia gradually spun off its agricultural operations to a new created subsidiary, Monsanto Company (New Monsanto). In 2003 the new Pharmacia was acquired by Pfizer and is now a wholly owned subsidiary of Pfizer. Table A3 illustrate how our methodology allows us to correctly compute patent stock and flow for each GVKEY record.

[18] The following are additional examples: (I) Patents of Honeywell before the merger with Allied Signal (3,112 patents) are incorrectly linked to Allied Signal's GVKEY (001300) up to 1999, while the financial records of the original Honeywell Inc are disregarded (GVKEY 5693). (II) Patents of TELEDYNE INC (GVKEY 10405) pre-merger with the publicly traded ALLEGHENY LUDLUM CORP in 1996 (to form ALLEGHENY TELEDYNE INC, which in 1999 was renamed ALLEGHENY TECHNOLOGIES INC after TELEDYNE was spun-off as free-standing public company) are not linked GVKEY 10405 (634 patents up to 1999, of which 597 patents are pre-1996 merger). In addition, ALLEGHENY LUDLUM CORP's (GVKEY 13708) patents (254 patents, of which 240 patents pre-1996 merger) were not dynamically moved to TELEDYNE INC post-merger. This means that in 1996 (post-merger) the patent stock of GVKEY 10405 is missing at least 789 patents (not including related subsidiary patents). (III) For the new Biogen Inc (GVKEY 24468) NBER does not include patents of IDEC pharmaceuticals, who was the owner of the security before Biogen and IDEC merged in 2003 (40 patents)

Table A3. PHARMACIA & UPJOHN and MONSANTO dynamic match

Period	related GVKEY	Relevant Compustat name for period	Most recent Compustat name	Comments	Patent flow per period per our strategy (based on NBER raw patent match, w/o subsidiaries)	Original NBER match
1950-1994	11040	UPJOHN CO	PHARMACIA & UPJOHN INC	Original Upjohn before merger with Pharmacia	**2,091** Upjohn related patents	N/A
1995-1999	11040	PHARMACIA & UPJOHN INC	PHARMACIA & UPJOHN INC	**1995**: Upjohn merged with original Pharmacia to form Pharmacia & Upjohn	**479** Pharmacia &/ Upjohn related patents	N/A
1950-1999	7536	MONSANTO CO	PHARMACIA CORP	Original Monsanto before merger with Pharmacia & Upjohn	**3,228** Monsanto related patents	**2,733** Pharmacia &/ Upjohn related patents (including patents of Pharmacia before it merged with Upjohn). **While Monsanto's 3,228 patents are not linked.**
2000-2002	7536	PHARMACIA CORP ("new Pharmacia")	PHARMACIA CORP	**2000:** original Monsanto merged with Pharmacia & Upjohn to form Pharmacia Corporation (New Pharmacia). All of PHARMACIA, UPJOHN and PHARMACIA & UPJOHN patents are transferred here from 2000. Monsanto's patents are redirected to the new Monsanto spin-off company.	**304** Pharmacia &/ Upjohn related patents	**304** Pharmacia &/ Upjohn related patents
2000-2015	140760	MONSANTO CO ("new Monsanto")	MONSANTO CO	**2000-2002:** Pharmacia Corporation (New Pharmacia)gradually spun-off its agricultural operations to a new publicly traded company, Monsanto Co (New Monsanto). All Monsanto related patents are transferred here from 2000.	**553** Monsanto related patents (2000-2006)	**553** Monsanto related patents (2000-2006). *NBER links Monsanto's patents to GVKEY 140760 from 1997 - while records for 1997-1999 are available on Compustat, they are based on prospective filings when Monsanto was still traded under GVKEY 140760.
2003-2015	8530	PFIZER INC	PFIZER INC	**2003:** Pharmacia Corporation (New Pharmacia) was acquired by Pfizer and is now a wholly owned subsidiary of Pfizer. All of PHARMACIA, UPJOHN and PHARMACIA & UPJOHN patents are transferred here from 2003.	**472** Pharmacia &/ Upjohn related patents(up to 2006)	**472** Pharmacia &/ Upjohn related patents(up to 2006)

III. AGGREGATING DATA TO THE UO FIRM LEVEL

To merge parent Compustat companies and their independent majority-owned publicly traded Compustat subsidiaries (Challenge 3 above), we locate related firms in our initial Compustat subsample based on name similarity as well as by matching the firm names to ORBIS subsidiary data. Where needed, we perform manual checks to confirm majority ownership using SEC 10-K filings. We aggregate the data to the UO parent-company level, accordingly.[19] We further link private subsidiaries to their UO firm based on ORBIS data (will be explained separately below). Accordingly, if a firm's subsidiary publishes scientific articles while the parent company is the assignee registered on the firm's patents, we record both at the UO level and a citation from a patent to a publication would be considered as an internal citation.

B. OWNERSHIP STRUCTURE

Dealing with ownership changes has been a major effort of this revision, especially in regard to reconstructing and improving the NBER patent database. Ownership may change over the years of our sample- these include ownership changes at the UO Compustat firm level as well as at the subsidiary level. Tracking ownership changes is a main challenge we deal with. We rely on two main sources to construct ownership data: (i) SDC Platinum and (ii) ORBIS.

I. SDC M&A MATCH

Ownership changes of the UO Compustat firms in our sample are tracked through the SDC Platinum database with each firm name dynamically matched to up to five PERMNO_ADJ between years 1980 and 2015. Based on M&A deals available in SDC Platinum from 1980 to 2015, we downloaded detailed information on the acquirer and target firm names, acquirer and target firm CUSIPs, types of deals, execution dates, and percentage of shares owned after each transaction. We exclude deals that we identify as asset or business unit acquisitions.

We restrict the sample to deals involving a change in ownership that resulted in majority ownership (more than 50% of shares) for the acquirer. Execution dates are used to define the years a target firm begins or ends (in case of several acquisitions during the sample period) being owned by an acquirer. We then standardized both target and acquirer names similar to the standardization done for Compustat firm names. We match each deal's target and acquirer firm to our list of Compustat firms using both CUSIP numbers and all standardized historical names. It is important to use historical data as the information is recorded on SDC at the time of acquisition. We retain deals where both acquirer and target firms are matched to a Compustat firm in our sample. We track up to five ownership changes for each target firm name after it enters Compustat and one additional reassignment before it became publicly traded if relevant (i.e., if it was a subsidiary of another Compustat firm in our sample prior to its own IPO)[20]. We Assume that if a firm is acquired all of its patents and publications are transferred to the acquirer firm.

[19] For example, GENZYME CORP (GVKEY 12233) - after verifying ownership on SEC filings: GENZYME MOLECULAR ONCOLOGY (GVKEY 117298), GENZYME TISSUE REPAIR (GVKEY 118653), GENZYME SURGICAL PRODUCTS (GVKEY 121742) and GENZYME BIOSURGERY (GVKEY 143176) are all accounted under their parent company GENZYME CORP (GVKEY 12233). While, GENZYME TRANSGENICS CORP (a.k.a. GTC BIOTHERAPEUTICS, GVKEY 028563) is a standalone alone company in our data as it was not majority owned by GENZYME CORP after it spun-off.

[20] For example, Vysis Inc first enters our sample as a subsidiary of Amoco (1991-1997) and is then spun-off and becomes an UO firm in our sample as an independent publicly traded company in 1998 and eventually acquired and becomes a subsidiary of Abbott in 2001.

Example 4: NABISCO

This example illustrates how we account for ownership changes in our data. During our sample period Nabisco has changed ownership four times. In 1981 Nabisco merged with the publicly traded company Standard Brands to form Nabisco Brands. Then, in 1985 R.J. Reynolds merged with Nabisco Brands to create RJR Nabisco, which eventually became Nabisco Group holding after the tabaco business was spun out in 1999. In 2000, Nabisco was acquired by Phillip Morris, which combined Nabisco with its Kraft brand. Finally, in 2001 Kraft (together with Nabisco) was spun out as a publicly traded company that later on became Mondelez International Inc. In our dataset all Nabisco related patents and publications are dynamically transferred between Compustat records and UO firms based on its ownership throughout the years:

Table A4. Nabisco dynamic match

Years	relevant PERMNO_ADJ	related GVKEY	Original Compustat name for period	Current Compustat name	Comments
1981-1985	14533	7674	NABISCO BRANDS INC	NABISCO BRANDS INC	**1981**: Standard Brands company merged with Nabisco Inc (both publicly traded at the time) to form Nabisco Brands Inc.
1986-1999	14218	9113	R J R NABISCO INC	NABISCO GROUP HOLDINGS CORP	**1985**: R.J. Reynolds Industries merged with Nabisco Brands to form R J R Nabisco Inc (which eventually became Nabisco Group Holding when the tabaco business spun-out of RJR Nabisco in 1999)
2000	13901	8543	PHILIP MORRIS COS INC	ALTRIA GROUP INC	**2000:** Nabisco was acquired by Phillip Morris (today known as Altria Group Inc), which combined Nabisco with its Kraft brand.
2001-2015	89006	142953	KRAFT FOODS INC	MONDELEZ INTERNATIONAL INC	**2001:** Kraft together with its subsidiary Nabisco split from Phillip Morris as a new publicly traded company, which later on becomes Mondelez International Inc.

Examining NBER 2006, we find that for the purpose of Compustat accounting information all Nabisco related patents are linked to GVKEY 9113 from 1950 to 1999. Though the current name related to GVKEY 9113 is "Nabisco Group Holding Corp", based on the historical name information, we know that up to the merger of R.J. Reynolds with Nabisco it belonged solely to R.J. Reynolds. Reynold's patents, on the other hand (Over 419 patents for the period before it spun-out of RJR Nabisco and not including patents of acquired companies such as Heublein Inc), are not assigned by NBER to GVKEY 9113 and they are only being linked to Compustat records after the tabaco business spun-out of RJR Nabisco and became independently traded again under GVKEY 120877 (eventually merging with U.S. operations of British American Tobacco to form Reynolds American Inc). As a result, in 1998 the patent stock in NBER for GVKEY 9113 ("Nabisco Group Holding Corp") is 495 (consisting solely of Nabisco matched patents) whereas it should be 914 if it included R.J. Reynolds related patents. Furthermore, NBER does not dynamically move Nabisco's patent stock or account for its patent flow after 1999 when it was bought by Philip Morris and eventually became part of Kraft (a total of 529 Nabisco related patents up to 2006).

Table A5. Data for Nabisco in NBER 2006

Current compustat record name	gvkey	firstyr	lastyr	pdpco	pdpseq	begyr	endyr
NABISCO GROUP HOLDINGS CORP	**9113**	**1950**	**1999**	**9113**	**1**	**1950**	**1999**
NABISCO INC	7675	1950	1980	9113	-1		
NABISCO BRANDS INC	7674	1950	1984	9113	-1		
NABISCO HLDGS CORP -CL A	31427	1993	1999	9113	-1		

Example 5: CHEMTURA CORPORATION

An example that illustrates how having historical names helps us account for ownership changes in our data and accurately compute the patent stock. Chemtura Corporation traces back to the chemical corporation Crompton & Knowles that was founded in the 19th century. In 1996, Uniroyal Chemical Corporation merged with Crompton & Knowles. In 1999, Crompton & Knowles merged with the publicly traded company Witco to form Crompton Corporation. In 2005, Crompton acquired the publicly traded company Great Lakes Chemical Company, Inc., to form Chemtura Corporation, while Great Lakes Chemical Corporation continued to exist as a subsidiary company of Chemtura.

Based on our strategy we consider all historical names of the current Chemtura Corporation (PERMNO_ADJ 38420) including:

1) CROMPTON & KNOWLES CORP starting 1980
2) CK WITCO CORP starting 1999
3) CROMPTON CORP starting 2000
4) CHEMTURA CORP starting 2005

Most importantly, because we consider the complete set of historical names, we are able to locate all the relevant M&As throughout the years of the publicly traded firms that exist as an independently traded company in our data prior to an acquisition. Accordingly, we dynamically transfer them post acquisition to PERMNO_ADJ 38420:

1) Uniroyal Chemical Corporation (acquired 1996)
2) Witco Corp (acquired 1999)
3) Great Lakes Chemical (acquired 2005)

When we examine NBER 2006 patent dataset we find that the only name that was matched to CHEMTURA CORP (GVKEY 3607) is "CHEMTURA CORP" (PDPASS 13245038). As the Chemtura name was adopted in 2005, only 1 patent was matched for that name. In addition, none of the acquired publicly traded companies were dynamically transferred to CHEMTURA CORP post acquisition. It is likely that a lack of information on historical names led NBER to rely on post-acquisition name (Chemtura) and thus prevented it from accounting for the M&A activities.

By considering all the previous names (without their subsidiaries and the acquired companies) related to GVKEY 3607: (i) Crompton & Knowles Corp; (ii) CK Witco Corp and (iii) Crompton Corp - based on the NBER raw patent match, we locate 220 additional patents up to 2006 that were not linked to any Compustat record that should be assigned to *Crompton & Knowles* (77 patents), *CK Witco* (26 patents)), and *Crompton* (117 patents). In addition, the acquired *Uniroyal Chemical Corp* has a patent stock of 379 patents in 2006 (out of which 185 patents are post acquisition) and the acquired Witco company has a patent stock of 405 in 2006 (out of which 62 patents are from post-acquisition period) and *Great Lake Chemicals* has a patent stock of 183 in 2006 (out of which 3 patents are in 2006, the year after the company was acquired).

Overall, applying our strategy to the raw NBER patent match, we find a patent stock of **1,187** patents in 2006 for GVKEY 3607 as opposed to **1** patent in NBER.

II. ORBIS SUBSIDIARY MATCH

Due to the complexity of measuring large firms' innovative activities, which typically take place inside numerous subsidiaries, we aggregate the data to the ultimate-owner-parent-company level based on majority ownership. There are several challenges in keeping track of subsidiaries owned by UO Compustat firms, which may publish and patent in their own name. First, many of these subsidiaries are private and manual checks are sometimes required to verify which of the several similarly named companies was acquired by the firm. Furthermore, subsidiary ownership may change over the years. Companies may spin out their subsidiaries, some of which might go public or sold to other firms, where they are maintained as stand-alone subsidiaries and continue to patent or publish. Tracking subsidiary ownership is a main challenge we deal with and is explained below.

For firms with at least 50 patents over the sample years at the PERMNO_ADJ UO level, we collect all related subsidiary names using ORBIS and SEC filings as explained below.

We obtained ORBIS files for years 2002 to 2015, which provide us with snapshots of ownership structures for each of the years. One caveat is that the coverage of subsidiaries in the first few years of data files is incomplete.

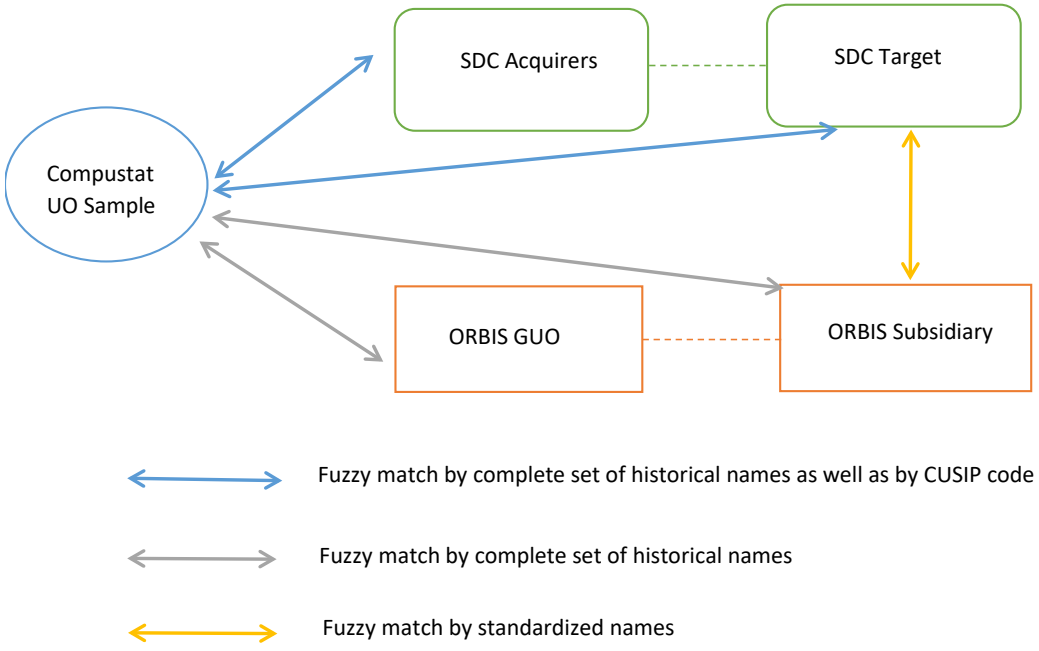
We start by standardizing the names of "Global Ultimate Owner" (GUO) firms and match the names to standardized historical Compustat names of firms with more than 50 patents at the UO level. Once again, it is important to use historical names for this match as the names that appear in each of our ORBIS files are as of the year the file was recorded.

Next, we link the subsidiaries of the successfully matched ORBIS owners to the PERMENO_ADJ of the corresponding parent firms. We restrict our sample to subsidiaries that are majority owned by the parent firm. After standardizing each subsidiary name similar to the standardization done for Compustat names, we obtain the first and last year it appears under a PERMENO_ADJ during 2002-2015. To avoid duplicated matching efforts, we drop subsidiaries that have the same organic name as the parent UO firm as they were already matched at the UO Compustat level. Some subsidiary names appear under more than one PERMNO_ADJ due to acquisitions throughout the years. Because we use yearly snapshots of ownership structure from ORBIS, we are able to account for name changes of subsidiaries over the period.

For firms that exit Compustat before 2002, we manually collect subsidiary names based on their latest available 10-K SEC filing.[21]

Since our sample starts from 1980 and the ORBIS files are only from 2002, we try our best to account for ownership changes of the subsidiaries for the years preceding 2002 using SDC and Compustat databases. We elaborate on our approach below.

Figure A2. Subsidiary matching Description



[21] We do so for top 100 firms based on R&D spending.

1) Fuzzy match between standardized subsidiary names and standardized SDC target name. For the matched result we locate:

a) Cases where the acquirer firm is a UO Compustat firm in our sample, which include:

(i) cases where the acquirer firm has the same PERMNO_ADJ as the parent firm of the subsidiaries. These cases confirm the direct acquisition of the subsidiary by the parent firm and provide us with the start date of the subsidiary (the year of acquisition) under the parent firm.

(ii) cases where the acquirer firm is a UO Compustat firm in our sample that was acquired by the parent firm of the subsidiary (i.e. the PERMNO_ADJ of acquirer and the PERMNO_ADJ of the parent of the subsidiary are related through acquisition). These cases confirm an indirect acquisition of the subsidiary by the parent firm and provide us with the start year of the subsidiary under the parent firm – i.e., year the ORBIS parent firm acquired the Compustat acquirer firm or the year of acquisition of the subsidiary (the latest).

b) cases where the acquirer firm is not a UO Compustat firm in our sample:

(i) if the CUSIP code of the UO parent firm related to the target firm (as indicated in SDC file) is the same as a CUSIP code related to the PERMNO_ADJ of the ORBIS parent of the subsidiary, it indicates that the subsidiary was acquired from the parent firm by the acquirer and provides us with the end date for the subsidiary under the parent firm – the year of acquisition.

(ii) for each acquirer firm's direct CUSIP code we search the complete SDC file for a deal where it was acquired by a firm with a CUSIP code related to the PERMNO_ADJ of the ORBIS parent of the subsidiary. These cases indicate indirect acquisitions, in which the subsidiary was acquired by a non Compustat sample firm that was itself acquired by the subsidiary's ORBIS parent firm. Such cases provide us the start year of the subsidiary under the parent firm –i.e., year the ORBIS parent firm acquired the non-Compustat acquirer firm or the date of acquisition of the subsidiary (the latest).

2) Fuzzy Match of cleaned subsidiary names

As the subsidiary name list includes closely related firm names with different legal entity, we use a clean version of the names that omits legal entity and other common words and we fuzzy match it to both clean Compustat names and the list of clean subsidiary names we found relevant acquisitions for in (1) above.

The fuzzy match to Compustat enables us to link each matched subsidiary name to the dynamic year sequence we constructed for UO Compustat firms. For the fuzzy match to the list of acquired subsidiaries, we adopt the relevant start &/end year we located in (1) above to all related subsidiaries.

3) As an additional check, we manually go over subsidiaries that did not match under 1) or 2) above and appear under more than one parent firm in our ORBIS sample or have more than 100 matched publications or patents. For these cases we check online sources and manually adjust their start and end date. Finally, for subsidiaries we were not able to identify start or end year, we make the assumption that they belong to the UO firm from its start date until end date. However, if the UO firm appeared in Orbis files for more than 3 years before the subsidiary was first linked to it, we adopt the first year the subsidiary is connected to the parent ORBIS firm as the start date of the subsidiary, under the assumption that it was acquired during that year by the parent firm.

All subsidiaries are assumed to move with their parent firm in cases where the parent firm is acquired, unless a subsidiary has a different end date from its parent firm or it is related to a the Compustat dynamic year sequence.

C. MATCHING

I. MATCHING PATENT DATA TO COMPUSTAT COMPANIES AND THEIR SUBSIDIARIES

We procced to matching our firm names to the assignees of the patents granted by USPTO using PatStat, which includes approximately 5.3 million patents for years 1980 through 2015.

We first remove published patent applications (i.e. publication numbers longer than 7 characters), non-utility patents, including Design, Reissue, Plant and T documents, and reexamination certificates. Next, we remove patents assigned to individuals or government entities (for example, an assignee that includes the string "DECEASED" or "U.S. DEPARTMENT"). We are then left with 4.97 million granted utility patents.

In order to compare assignee names to the standardized firm names in our sample, we standardize assignee names similar to the firm name standardization explained above. Assignee name standardization includes converting names to upper case, removing excess spaces, cleaning non-alphanumeric characters and replacing legal entity endings including commonly abbreviated terms (for example, "CORPORATION" is replaced with "CORP"; "LABORATORIES" and "LABS" with "LAB"). At the end of this process, we are left with 897K unique standardized assignee names.

The matching strategy includes several distinct steps. We begin by matching firm names to assignees using exact match. We then perform several fuzzy matching techniques to account for names that are slightly different but are in fact the same entities. Extensive manual checks at the assignee name and patent level were performed to ensure the quality of the matches.

(1) Exact Matching

Exact matching was conducted by comparing standardized assignee names to standardized firm names. Exact matching was carried out twice, once without and once with legal entities. The latter step was performed to account for firms whose names differ only by the legal entity. Extensive manual checks are performed to verify the matches. Special care was taken in cases where firm or assignee names are generic, when several different firms share a common portion of a name, or when firm names contain a common given or family name. To resolve ambiguities, we performed web searches and examined the actual patent documents.

(2) Fuzzy Matching

Fuzzy matching was performed to account for cases where assignee and firm names do not match exactly but are in fact the same firms. Some assignee names are misspelled or include additional letters that prevent an exact match. In other cases, patent assignee names include a specific division title ("ROCKWELL BODY AND CHASSIS SYSTEMS", "ROCKWELL SOFTWARE"), a licensing unit ("MICROSOFT TECHNOLOGY LICENSING LTD", "RCA LICENSING"), or a geographic branch or firm location ("BIOSENSE WEBSTER ISRAEL LTD").

Fuzzy matching was performed by converting assignee and firm names into a term frequency–inverse document frequency (TF-IDF) matrix and calculating a cosine similarity score for each of assignee and firm name pair. To reduce the size of the task, the results were limited to assignees with at least 30 related patents. Given potential matches, final matches were selected manually by examining the top scoring assignee-firm pairs for each assignee.

To overcome the challenge of matching assignee names that include an additional description (such as division or location), we perform a cruder fuzzy matching between each standardized assignee name and the fully standardized firm names to locate cases where a firm name is a substring of an assignee name (for example, "EMERSON" is matched to "EMERSON CLIMATE TECH", a division within the firm). To reduce the size of the task, we perform this matching for the top 300 patenting firm names (based on previous matching procedures) and limit the search to assignee names with at least 5 related patents. Given potential matches, final matches are selected manually.

As an additional robustness check, we employed RAs to manually search the data for firm names with more than 100 patents (based on previous matching procedures) and find matching assignee names that might have been missed during exact and fuzzy matching.

Overall, this process yields 1.29 million patents mapped to the Compustat firms and their subsidiaries in our sample via patent number and NAME_ID. When a patent has several assignees, we match the patent to multiple firms and assign fractional patent ownership to each assignee (i.e., 1/number of assignees).