Research Policy 50 (2021) 104217

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Matching patents to compustat firms, 1980–2015: Dynamic reassignment, name changes, and ownership structures

Ashish Arora [a,*], Sharon Belenzon [a], Lia Sheer [b]

[a] Fuqua School of Business Duke University, and NBER
[b] Fuqua School of Business Duke University

ARTICLE INFO

Keywords:
Innovation
NBER patent data
R&D
Market value
Measurement error

ABSTRACT

This paper describes the methodology used to construct a new sample of patents matched to Compustat firms for the period 1980–2015. We compare our data to existing NBER data sets and underscore several methodological improvements, including dynamic matching, company name changes, and ownership structures. We examine how our improved match changes results obtained from the '01 and '06 NBER patent files using comparable samples. Overall, we find that improved patent assignment leads to slightly higher estimates of patent value in market value regressions, as well as to higher estimates of the R&D elasticity in patenting regressions.

1. Introduction

An extensive literature uses patent data to answers questions on the determinants and consequences of inventive activity. Many papers use patent data from the NBER, which matches patents granted by the United States Patent Office to publicly traded American firms (henceforth, Compustat firms). There are two versions of the NBER data. The first, introduced by Adam Jaffe, Bronwyn Hall, and Manual Trajtenberg (Hall et al., 2001), pioneered the use of patent data as indicators of inventive activity at scale. We refer to this dataset as NBER '01, which covers the years 1980–1999. The second iteration of this database was developed by Jim Bessen (Bessen, 2009) for the period 1980–2006 to address some shortcomings in the earlier version (henceforth, NBER '06). In particular, Bessen (2009) included an attempt to improve the dynamic reassignment of patents, wherein as firms changed owners, their patents would get reassigned (in the database) to the new owner.[1]

This paper describes a third iteration of the NBER data (henceforth, ABS), first used in Arora et al. (2021). We extend the data by a decade to 2015.[2] In addition, we reconstruct the complete historical data covered in the NBER data files. We build on Bessen's work and introduce several improvements focusing on better coverage of name changes and ownership structures.[3] We study the implications of our improved matches on patent value and R&D elasticity of patenting.

We combine data from five main sources: (i) company and accounting information from U.S. Compustat 2018, (ii) patents from PatStat; (iii) subsidiary data from historical snapshots of ORBIS files for

* Corresponding author.
E-mail address: ashish.arora@duke.edu (A. Arora).

[1] Some recent papers published in *Research Policy* that rely upon the NBER patent-Compustat firms match include Chen et al. (2020), Figueroa and Serrano (2019), Mezzanotti and Simcoe (2019), Im and Shon (2019), Raiteri (2018), de Leeuw et al. (2019), Figueroa and Serrano (2019), and Silvestri et al. (2018). Other notable papers published in *Research Policy* using this data include Fabrizio (2009), McGahan and Silverman (2006), Bessen (2008) Dushnitsky and Lenox (2005) and Acs et al. (2002). The dataset is used broadly in innovation studies- e.g., Marx et al. (2009), economics- e.g., Hall et al. (2005), management- e.g., Miller et al. (2007), and finance- e.g., Kogan et al. (2017).

[2] The description of the data construction underlying ABS in this paper draws substantially on an unpublished online data supplement to Arora et al., 2021, which also provides more details on how we deal with subsidiaries, name changes, and matching. The case studies in this paper are also drawn from that source. This paper systematically compares the estimates of the market value of patents, and the R&D elasticity of patents between the NBER 01, NBER 06, and ABS. In addition, it provides a simple econometric framework on the possible direction of bias arising from the two main types of measurement error – dynamic matching, and omitted patent matches.

[3] Assignee disambiguation is not an important contribution of our data work. We build on existing name harmonization and string matching approaches. Several patent data projects such as the USPTO's PatentsView, UC Berkeley's Patent Database (Li et al., 2014), and Darden's Global Corporate Patent Dataset (GCDP) (Bena et al., 2017) have advanced assignee disambiguation and assignee-matching techniques. Researchers can use our historical standardized name lists to match with their dataset of interest.

https://doi.org/10.1016/j.respol.2021.104217
Received 19 September 2020; Received in revised form 7 February 2021; Accepted 8 February 2021
Available online 13 March 2021

2002–2015; (iv) mergers and acquisition data from SDC Platinum and (v) company name changes from WRDS's "CRSP Monthly Stock". For ownership and subsidiary data, we supplement NBER '06 with a wide range of M&A data, including SDC, historical snapshots of ORBIS files for 2002–2015, and 10-K SEC filings. We further perform extensive manual checks to uncover firms' structure and ownership changes.

There are two main areas of improvement. First, we match more patents. About 20% of patents belonging to Compustat firms were omitted or incorrectly matched in the NBER '06 patent database. Second, we achieve better dynamic reassignment. Dynamic reassignment means that, for instance, if a sample firm merges with another firm, the stock and flow of patents of the merged firm are linked to the Compustat record from that point onward, but not before. Similarly, once a subsidiary is divested from its parent Compustat firm, its patents are subtracted from the parent firm's stock from that point onward. Lastly, a significant component of the data upgrade is accounting for changes in names. 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. Finally, we make our data available to all researchers through a public data repository.[4]

We examine how estimates of the "shadow price" of patents in market value regressions and R&D elasticity in patenting regressions change across NBER datasets and our updated data. In general, we find that our data produces higher estimates of patent value from estimating Tobin's Q specifications, as well as higher R&D elasticity estimates from a patent production function.

Section 2 explains how we construct our sample. Section 3 discusses the main challenges we face when matching patent data to Compustat firms. Section 4 presents the patent matching procedure and the process of dynamic reassignment. Section 5 compares ABS data to NBER '06, including detailed case studies to illustrate improvements to the NBER database. Section 6 analyzes differences in estimates of patent value and R&D productivity between existing NBER datasets and our sample, and Section 7 concludes.

2. Sample construction

To construct our sample, we start with all North American Compustat records obtained from WRDS in August 2018 and select companies with active records and positive R&D expenses for at least one year from 1980 through 2015 (inclusive). We define an active record as a year with common shares traded (CSHTR_F). We exclude firms that are not headquartered in the United States based on their current headquarter location in the Compustat 2018 file.[5] We further restrict our sample to companies with at least one patent during our sample period.

Following NBER '06, we aggregate the data to the parent company level, which we refer to as an ultimate owner (hereafter, UO). For example, the company GENZYME CORP (GVKEY 12233) is the ultimate owner of the publicly traded companies GENZYME MOLECULAR ONCOLOGY (GVKEY 117298), GENZYME TISSUE REPAIR (GVKEY 118653), GENZYME SURGICAL PRODUCTS (GVKEY 121742), and GENZYME BIOSURGERY (GVKEY 143176). Yet, GENZYME TRANSGENICS CORP (a.k.a. GTC BIOTHERAPEUTICS, GVKEY 028563) is a standalone company, because it has been spun-off by GENZYME CORP. A major contribution of our data is tracing ownership changes over time and identifying the exact years when GENZYME TRANSGENICS CORP falls under GENZYME CORP and the years when it is a standalone firm. An additional important contribution is accounting for private subsidiaries as well as Compustat, publicly-traded, subsidiaries.

A UO firm enters our sample once it is publicly traded and has at least one patent in stock 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 sample have at least 3 consecutive years of active records in Compustat. In total, we match 1.3 million patents to 4,420 U.S. headquartered Compustat firms and their subsidiaries. These patents account for about 50% of all utility patent grants of U.S. origin.[6] When a patent has several assignees, we match the patent to multiple firms and assign fractional ownership to each assignee (i.e., 1/number of assignees). In case of an ownership change within the sample, patents are dynamically matched to up to five UO firms (see Section 4.2 for detail).

3. Key measurement challenges

Many of the challenges in working with Compustat data arise because Compustat uses GVKEY codes to track companies, but GVKEY codes fail to capture changes over time in ownership structure and firm names, and the same company may have multiple codes over time. For example, Ralston Purina is listed under two different GVKEYs: (i) 1980–1993 under "RALSTON PURINA - CONSOLIDATED" (GVKEY 008935) and (ii) 1993–2000 under "RALSTON PURINA CO" (GVKEY 028701)). Compustat does not link related company identifiers, making it difficult to track companies over time only based on GVKEY. For example, "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" Compustat does not provide information on these changes. We develop a specialized database on current and historical ownership structures and name changes for our sample firms.

3.1. Company names

A major contribution of this paper is identifying name changes of Compustat firms over the sample years 1980–2015. Company names are important because U.S. patent documents list the name, rather than the CUSIP number or GVKEY of the assignee. Patent records contain the owner's name at the time of their publication, whereas companies appear in the Compustat file under their most current name with no records of previous names. The three leading reasons for name changes are: (i) M&A & restructure activity (36%);[7] (ii) change in focus of operation (17%); (iii) brand or subsidiary name adoption (12%).[8] Company names may also change over due to generic name changes (e.g., "MINNESOTA MINING AND MANUFACTURING" changed its name

[4] The data can be freely downloaded from the DISCERN database: 10.5281/zenodo.3594642. The version used for the analysis in this paper is DISCERN version 5. The database is updated on a regular basis, and we welcome contributions from the community in expanding the data - interested researchers are encouraged to contact the authors directly.

[5] 18.5% of the related Compustat firms are dropped due to the restriction on U.S. headquartered firms.

[6] We match 58% of all U.S. origin patents that are assigned to a U.S. company or corporation during the period 1980–2015 (assignee type classification is based on PatentsView data). When comparing ABS patents with the unmatched patents, we find that the main difference is in the use of in-text and front page NPL citations (Marx and Fuegi, 2020). Unmatched patents cite on average 1.6 in-text and 2.3 front-page citations more than ABS patents. These patterns are consistent with large firms having a much higher propensity to patent. Specifically, they are likely to file a large number of incremental patents (Cohen and Klepper, 1996), which are less likely to cite science. Unmatched patents are also more likely from science-intensive sectors, such as life-sciences.

[7] For instance,"WESTINGHOUSE ELECTRIC CORP" (GVKEY 011436) acquired "CBS INC" in 1995 and changed its own name to "CBS CORPORATION" in 1997 while maintaining the same GVKEY Compustat firm identifier)

[8] Reverse takeovers are related to this. For example, 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.

in 2002 to “3M”) Name changes imply that we cannot simply match the name of Compustat firms to patent assignees because Compustat includes only the most recent name of the focal corporation. For example, when Google reorganized as Alphabet in the summer of 2015, Compustat updated Google’s name in all historical records to Alphabet. This means that matching patents to Alphabet in 2015 would exclude Google’s patents.

About 30% of the Compustat firms in our sample change their name at least once.[9] To locate historical names, we use the WRDS’s “CRSP Monthly Stock” file, which includes monthly information on names for each security along with 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. For example, while SPHERIX INC is related to two 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. We calculate the starting and end years for each name based on the trading dates in the “CRSP Monthly Stock” file.

Using WRDS “CRSP/Compustat Merged Database - Linking Table”, we link each PERMNO from CRSP to Compustat GVKEY code. The crosswalk between CRSP and Compustat is not straightforward. As shown above, a PERMNO can have multiple related GVKEYs. In such cases, we apply a dynamic match between a PERMNO and Compustat accounting data. However, CRSP also includes cases where under the same GVKEY there are several PERMNO codes.[10] In such cases, we manually checked using 10K-SEC fillings the years that the name was relevant for each GVKEY.[11] We performed extensive checks on the name list, including identifying and distinguishing companies with similar names. Finally, we cleaned and standardized firm names as CRSP tends to abbreviate long words in the company name that it provides (see Appendix A for the name standardization procedure).

3.2. Ownership structure

Compustat does not link parent companies to their publicly traded subsidiaries, nor does it provide information on private subsidiaries. Because patents can be assigned to any legal entity in the corporation (Arora et al., 2014), we need to develop comprehensive data on corporate ownership structure. Moreover, because ownership can change over time, we need to trace these changes at the UO and subsidiary level so that we can assign patents to the relevant UO in each year.

There are several challenges in identifying subsidiaries owned by Compustat firms. First, many of the subsidiaries are private, and manual checks are required to verify which of the several similarly named companies is actually owned by a focal UO. Second, subsidiary ownership changes over time. Companies may spin out their subsidiaries, some of which might subsequently go public or be sold to other firms. A major contribution of our data is developing comprehensive time-varying data on corporate ownership structure.

We rely on two main data sources of information on ownership structure: (i) annual publications of ORBIS by Bureau Van Dyke, which provide annual ”snapshots” of private and public subsidiaries and (ii) SDC Platinum, which provides information on significant ownership changes, such as mergers, acquisitions, and spin-offs.

We use ORBIS’s complete ownership data for each year from 2002 through 2015, because 2002 is the first year ORBIS reported reliable firm coverage for American firms. For earlier years, we rely on NBER files and 10-K SEC filings. To match the names of Compustat firms to ORBIS, we standardize the names of the “Global Ultimate Owner” field in ORBIS, hereafter, GUO, similar to the standardization procedure we used for Compustat firms (see Appendix A). These companies can be UOs themselves or publicly traded subsidiaries of UO firms. Having matched the names of GUOs to all historical Compustat firm names, we retain all the subsidiaries listed in ORBIS of the successfully matched GUOs.

The next step is to match the related subsidiary names to Compustat. We restrict our sample to subsidiaries that are majority-owned by the GUO firm.[12] This yields an ownership link within Compustat between parent firms and their public subsidiaries. We use this information to aggregate the patent matching to the parent company level. In addition, ORBIS provides us with private subsidiaries for each Compustat firms, which we later use in the patent matching procedure.

Changes in ownership happen for a diverse set of reasons, including mergers, acquisitions, and spinoffs. We rely on SDC Platinum as an additional source of information for changes in ownership at the UO and subsidiary level.[13] We downloaded detailed information on the acquirer and target firm names, acquirer and target firm CUSIPs, type of deal, execution dates, and percentage of shares owned after each transaction. We restrict our focus to deals involving a change in ownership that resulted in majority ownership (more than 50% of shares) for the acquirer, and exclude deals involving asset or business unit acquisitions. We standardize target and acquirer names (see Appendix A) and match them to Compustat firms and their related subsidiaries.

For subsidiaries, execution dates are used to define the years a subsidiary begins or ceases to being owned by the GUO. For UOs, we track up to five ownership changes for each 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 IPO).[14] We assume that if a firm is acquired, all its patents are transferred to the acquirer firm. All subsidiaries are also assumed to move with their parent firm when the parent firm is acquired, unless indicated otherwise. We do not account for reassignments of patents that are not part of the ownership changes that we document.

[9] This is comparable to the findings by (Wu, 2010) that during 1925–2000, over 30% of CRSP-listed firms changed their names at some point after going public. 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 count SILICON GRAPHICS patent stock and patent flow under RACKABLE’s GVKEY starting from 2009.

[10] This is mainly due to significant M&A, including reverse acquisitions that occurred during the years when the firm was not listed.

[11] There is a difference in coverage between CRSP and Compustat for the early sample years. For example, CRSP only includes firms listed in major American exchanges and specifically excludes regional exchanges, while Compustat includes all 10K 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). We manually added missing information from Compustat and checked for historical names.

[12] The 10-K SEC filings used to supplement ORBIS data for the pre-2002 period usually list only majority-owned subsidiaries. Therefore, to be consistent throughout the sample period, we use only majority-owned subsidiaries. Since across the sample period, 75% of the ORBIS subsidiaries are majority-owned, we are capturing most of the subsidiaries. Furthermore, for much of our sample period (2002–2012 files), no more than a fifth of the subsidiaries of our sample firms are minority-owned. Only from 2013, the proportion of minority-owned subsidiaries jump to 40%.

[13] While ORBIS provides time-series information on ownership structure, its main advantage is mapping subsidiaries to parent firms in specific years. SDC, on the other hand, is a specialized database that focuses only on ownership changes.

[14] 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 is eventually acquired and becomes a subsidiary of Abbott in 2001.

4. Assigning patents to firms

Matching patents to their owner for each year involves two main steps. In the first step, we match names of patent assignees to Compustat firms and their subsidiaries. Second, we track changes in the ownership of a patent over time. We explain each step below.

4.1. Matching assignees to company names

We match our sample of Compustat firms and their subsidiaries to names of patent assignees from PatStat, which has approximately 5.3 million patents granted between 1980 and 2015. We 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. We also remove patents assigned to individuals or government entities (for example, an assignee that includes the string "DECEASED" or "U.S. DEPARTMENT"). This procedure leaves us with roughly 5 million patents and 897 thousand unique standardized assignee names, which we match to sample firms as follows.

We begin by matching firm names to assignees using an exact match procedure. For unmatched patents, we implement several fuzzy matching techniques to account for names that are slightly different, but are in fact the same entity. The final step includes manual checks at the assignee name and patent levels to ensure the correctness of the matches. The matching was carried out twice, both for standardized and for original names.[15] 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.

Some 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"), a licensing unit ("MICROSOFT TECHNOLOGY LICENSING LTD"), or a geographic branch or firm location ("BIOSENSE WEBSTER ISRAEL LTD"). For these remaining assignee names not matched during the exact matching process, fuzzy matching was performed using the FuzzyWuzzy library in Python (i.e., Token Set function), and using term frequency-inverse document frequency (TF-IDF). A vector is created for each assignee name using the words contained in it and then compared to the entire list of firm names (that are also vectorized) to find potential matches. When comparing two vectors, the same elements (words) contained in both vectors are marked as "matched", and the similarity between the remaining different elements are calculated using the Levenshtein distance algorithm after sorting the elements alphabetically. The similarity score between the two strings is higher when the elements that match exactly make up a larger portion of the strings and when the remaining (unmatched) part has a smaller Levenshtein distance. To account for multiple scores that indicate a strong match, the top ten potential matches with the highest scores were examined manually to identify the most appropriate match.

For the top 300 patenting firm names, we performed an additional search to find matching assignee names that were not matched through the fuzzy match process. We search for assignee names with at least five related patents that contain any of the fully standardized firm names after the removal of legal entities. Through this process, we include subsidiaries that have the same organic name as the parent UO firm (For example,"EMERSON" firm name matched with "EMERSON CLIMATE TECH", a division within the firm). The search was conducted through a script that receives the list of assignee names and fully standardized firm names and automatically produces all matching pairs, which were manually checked for legitimate matches.

[15] An additional match was performed after dropping legal entities, to account for firms whose names differ only by the legal entity.

Table 1
Comparison of ABS with NBER for 1980–2006: Patent-GVKEY assignments, U.S. HQ firms.

Comparison 1980–2006	% Patents	Examples
Agreement	80	
Matched to different GVKEY	4	Improved dynamic matching to Compustat records using historical name. Patents of the merged company included under the GVKEY from acquisition, but not before. Example: PHARMACIA: we matched to PHARMACIA & UPJOHN's GVKEY pre-2000 instead to MONSANTO
Only our Sample	14	Newly matched patents due to (i) availability of historical names; (ii) better M&A data; and (iii) Improved matching. e.g. PHILLIPS PETROLEUM CO: 4000+ patents pre-merger with Conoco Inc in 2002; MONSANTO: 2000+ patents pre-merger with Pharmacia.
Only NBER - we matched but irrelevant gvkey-year	1	(i) NBER match (incorrectly) based on 2006 Compustat name: e.g. 1000 patents of RHONE-POULENC patents matched to RORER's GVKEY pre-merger in 1990; (ii) Improved subsidiary coverage: e.g., 450 patents of HUGHES AIRCRAFT are incorrectly linked to GM's GVKEY pre-1985 acquisition.
Only NBER	1	(i) Withdrawn patents: 600 patents (ii) Misc. could not verify connection, typos, and possible mistakes by us

As a final check, we employed a team of RAs to verify that assignees with more than 100 patents were correctly matched by the fuzzy matching algorithm. Existing matches were invalidated when they were not the right match, and new matches were added when more appropriate matches were found.

4.2. Dynamic reassignment

We build on the methodology developed by Bessen (2009) and perform a dynamic reassignment for our sample of UO Compustat firms. Our matching is done at the firm name level. We assign each firm name a unique identifier labeled as ID_NAME and indicate the first, and last year the name is relevant for a PERMNO_ADJ – our UO identifier. We then perform dynamic matching of names to PERMNO_ADJ based on SDC's M&A data. PERMNO_ADJs are dynamically linked to GVKEYs (to link to Compustat data). The same ID_NAME can be linked to multiple PERMNO_ADJ over time, and each PERMNO_ADJ can be linked to multiple GVKEYs within the same year and over time. We include up to five ownership reassignments for each firm name that appears in our initial Compustat sample and acquired by another firm in our sample. Our UO and subsidiary historical standardized name lists, including dynamic reassignments, are publicly available.

5. Comparison with NBER data, 1980-2006

We compare ABS sample to NBER '06 for the period 1980–2006 (Appendix Figure C1 plots the difference in matched granted patents over time). Table 1 presents the comparison results.

About 80% of the patent-GVKEY matches are identical between the NBER and ourselves. We match an additional 18% of the patents mostly due to: (i) improved dynamic linkage of patents to GVKEYs (e.g., Pharmacia), and (ii) linkage of additional patents based on historical name information, wider M&A coverage, and improved matching techniques (e.g., Phillips). In 1% of the cases, we find the same assignment as NBER, but these matches are irrelevant for our sample (e.g., Rhone-Poulenc). Finally, in about 1% of the cases, we are unable to include the NBER matches for a variety of reasons, including possible

Table 2
Difference in means: Mismatched and matched firms (ABS vs. NBER ’06, 1980–2006).

	(1) Diff. in means	(2)	(3) Mismatched	(4)	(5)	(6) Not-mismatched	(7)
	(3) minus (6)	No. Obs.	Mean	Std. Dev.	No. Obs.	Mean	Std. Dev.
Market value	–157.8	21,146	2,427.8	11,917.3	22,302	2,585.5	20,618.5
Tobin’s Q	–0.9**	21,146	4.0	5.7	22,302	4.9	6.3
R&D stock	–108.6**	21,146	248.3	1,054.9	22,302	356.9	2,098.7
R&D expenditure	–19.9**	21,146	58.8	272.8	22,302	78.7	430.7
Assets	307.7**	21,146	1,337.2	6,219.2	22,302	1,029.5	7,606.3
Sales	392.2**	21,146	1,889.6	8,247.3	22,302	1,497.4	8,583.2

Notes: The sample consists of firm-year level observations matched on GVKEY-year pair between NBER ’06, and ABS samples. Mismatched firms are those whose cumulative granted patents in ABS is at least 10% above or below the number reflected in the NBER ’06 sample. ** p<0.01, * p<0.05.

Table 3
OLS estimates. absolute percentage difference in annual patents assigned (ABS vs. NBER ’06).

	DV: Absolute difference in annual patents (ABS minus NBER ’06) divided by mean annual patents					
	(1)	(2) Unscaled RHS variables	(3)	(4)	(5) Scaled RHS variables	(6)
VARIABLES	Pooled	Within-firms	Between-firms	Pooled	Within-firms	Between-firms
ln(R&D stock)	0.019*	0.009	0.012			
	(0.009)	(0.019)	(0.009)			
ln(Assets)	0.021**	0.010	0.039**			
	(0.007)	(0.008)	(0.007)			
ln(Tobin’s Q)	–0.001	-0.007	0.017	–0.011	–0.009	-0.001
	(0.006)	(0.006)	(0.010)	(0.007)	(0.006)	(0.011)
Patents / R&D exp.	0.058**	0.063**	0.028**	0.048**	0.062**	0.003
	(0.006)	(0.005)	(0.009)	(0.006)	(0.005)	(0.008)
R&D stock / Assets				–0.003	–0.001	–0.007
				(0.003)	(0.004)	(0.004)
Industry fixed effects	Yes	No	Yes	Yes	No	Yes
Firm fixed effects	No	Yes	No	No	Yes	No
Year fixed effects	Yes	Yes	No	Yes	Yes	No
DV sample average	0.385	0.385	0.361	0.385	0.385	0.361
Number of firms	3,750	3750	3,771	3,750	3,750	3,771
Observations	34,047	34,047	3,771	34,047	34,047	3,771
R-squared	0.08	0.50	0.13	0.07	0.50	0.11

Notes: The sample consists of firm-year level observations matched on GVKEY-year pair between NBER ’06, and ABS samples. The absolute difference in yearly patent count is calculated by subtracting yearly counts reflected in NBER ’06 from values in ABS. The absolute difference is divided by mean of ABS and NBER ’06 patent for the firm in that year. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by GVKEY. ** p<0.01, * p<0.05.

mistakes on our part. In an unreported analysis, we compare the citation patterns of ABS patents with patents matched only by NBER ’06. The difference in yearly weighted forward patent citation per patent (ABS minus NBER ’06) is only 0.066 (0.08 for IPC-year weighted forward patent cites), which is less than 6% of the mean value of forward citations per patent.

Table 2 examines differences in characteristics of firms that are mismatched by NBER ’06 and firms that are matched correctly. For this exercise, the sample consists of firm-year level observations matched on GVKEY-year pair between NBER ’06 and ABS samples for the period 1980–2006. We classify a firm as mismatched if the firm’s cumulative patent stock in ABS is at least 10% above or below the number reflected in the NBER ’06 sample.[16] Overall, Table 2 shows that mismatched firms are bigger as measured by assets and sales, but have a lower Tobin’s Q, and R&D stock and flow.

Appendix Table C1 further presents a list of the top 50 firms with the highest number of average annual difference in matched patents. The number of firms with at least one mismatched patent ranges from 288 in 1980 to 658 in 2006. For these firms, the average absolute difference in matched patent per year is 12.2. Mismatches are higher in chemistry and life science with an annual 13 mismatched patents per firm, as compared to 12 and 10.75 mismatched patents in ICT and semiconductors, respectively. Comparing NBER ’06 to ABS, 1,664 firms are mismatched (approximately 40%). Of the mismatched firms, 129 firms have more patents in NBER ’06 than in ABS. That is, for these firms, ABS revise downwards their number of patents (mostly through dynamic reassignment). Comparing firms with a downward revision of their patents to those with an upward revision of patents, downward revision firms have lower Tobin’s Q (3.4 for downward vs. 4 for upward), yet are bigger in terms of sales (2,271 for downward versus 1,858 for upward) and assets (1,345 for downward versus 1,337 for upward).

Table 3 presents a regression version of the difference in means analysis. The dependent variable is the absolute number of mismatched annual patents between ABS and NBER ’06, normalized by the number of annual patents of the firm. The regression results suggest that larger firms (as measured by assets) are likely to see higher share of mismatched patents. Further, patent-intensive firms (measured as the ratio of patents per R&D expenditures) have a higher share of mismatched patents between ABS and NBER ’06.

5.1. *Case studies*

We present two examples to illustrate the outcome of our patent matching procedure. Appendix B provides additional case studies. The Time-Warner case-study B.1 illustrates how properly accounting for name and ownership changes improve the accuracy of patent flow as

[16] To facilitate comparison between the samples, we consider only patents granted under the current UO.

Table 4
Data for SPX corp in NBER '06.

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		

Notes: PDPCO is NBER's Patent Data Project (PDP) unique company id. FIRSTYR is the first year GVKEY company has data. LASTYR is the last year a GVKEY company has data. PDPSEQ is the GVKEY sequence within PDPCO. If PDPSEQ=-1, the related GVKEY is disregarded. BEGYR is the beginning year for GVKEY within PDPCO. ENDYR is the last year for GVKEY within PDPCO. All patents related to SPX CORP will be accounted under GVKEY 5087 from 1950 to 2006, while the original SPX GVKEY (9556) is disregarded.

well as the dynamic reassignment of patents. The Pharmacia case-study B.2 demonstrates how we use the complete history of names to identify each Compustat record's historical ownership and dynamically match a firm to its relevant financial records in any given year. The Nabisco case-study B.3 illustrates how we account for ownership changes in our data. Lastly, the Chemtura Corporation case study B.4 illustrates how having historical names helps account for ownership changes and accurately compute the patent stock.

5.1.1. SEALED POWER and GENERAL SIGNAL

The following example underscores the mismatching consequences of not accounting properly for name and ownership changes. Until 1998, SEALED POWER and GENERAL SIGNAL were two distinct entities. Historical Compustat records include the following information for these companies until 1998:

1. GVKEY 9556, related names:
   (a) SEALED POWER CORP (1962–1988) - original name
   (b) SPX CORP (1988–1997) -name changed retroactively in Compustat
2. GVKEY 5087, related name: GENERAL SIGNAL CORP (1950–1997)

In 1998, SPX Corp acquired General Signal Corp in a reverse merger, and General's GVKEY (5087) became the new security of SPX traded under the new name "SPX CORP". The original SPX records were renamed retroactively in Compustat as "SPX CORP-OLD". Current Compustat records include the following records for these companies:

1. GVKEY 9556, related name: SPX CORP-OLD
2. GVKEY 5087, related name: SPX CORP

We 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) and connect the SPX CORP name to General's original GVKEY (5087) only from 1998 onward. In the NBER '06 patent dataset (see table 4), the two companies are collapsed under the same company (same PDPCO id), and the accounting data associated with General's original GVKEY (5087) is used for the complete period, while the original SPX GVKEY (9556) is disregarded. Similarly, NBER'06 also assignes the pre-1998 SPX patents to General's GVKEY. 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 not assigned to any compustat record, though they are located in the 2006 NBER raw patent match. This example shows that a considerable number of patents were omitted from the stock of "GENERAL SIGNAL" patents, leading to an undercount of both the flow of patents pre-1998 for this firm, as well as a persistent undercount in its patent-stock.

Table 5 presents mean comparison tests for patents that are matched by both ABS and NBER '06 (338 patents in total, out of which only 286 patents are matched to the correct GVKEY by NBER '06) and 859 patents that are matched only by ABS. We find that patents matched only by ABS are of higher quality as measured by forward patent citations. The difference in yearly weighted forward patent citations per patent (matched by both samples minus only ABS) is minus 0.17 (minus 0.16 for IPC-year weighted forward patent cites).

Table 5
SEALED POWER and GENERAL SIGNAL: Comparison between ABS and NBER '06 patents (1980–2006).

	(1) Diff. in means	(2)	(3) Both	(4)	(5)	(6) Only ABS	(7)
	(3) minus (6)	No. Obs.	Mean	Std. Dev.	No. Obs.	Mean	Std. Dev.
Forward patent cites	−6.039**	338	14.580	17.665	859	20.619	24.464
Yearly weighted forward patent cites	−0.170**	338	0.640	0.731	859	0.811	0.897
IPC-Year weighted forward patent cites	−0.156**	338	0.728	0.712	859	0.884	0.922

Notes: This table presents mean comparison tests for patents that are matched by both ABS and NBER '06 vs. those that are matched only by ABS. We include all matched patents granted between 1980–2006 and related to GVKEYs 5087 and 9556. The unit of analysis is a patent. Forward patent citations are from patents granted up to 2016. ** p<0.01.

Table 6
Example of dynamic name list for Conoco-Phillips.

ID Name	Name std	Fyear 0	Nyear 0	Permno Adj 0	Name ACQ 0	Fyear 1	Nyear 1	Permno Adj 1	Name ACQ 1	Fyear2	Nyear 2	Permno Adj 2	Name ACQ 2
2384	CONOCO INC	1981	1997	11703	DU PONT E I DE NEMOURS & CO	1998	2001	86368	CONOCO INC	2002	2015	13928	PHILLIPS PETR CO
7325	PHILLIPS PETR CO					1998	2002	13928	PHILLIPS PETR CO	2003	2015	13928	CONOCO PHILLIPS
2385	CONOCO PHILLIPS					2002	2015	13928	CONOCO PHILLIPS				
7324	PHILLIPS 66	1980	2011	13928	CONOCO PHILLIPS	2012	2015	13356	PHILLIPS 66				

Notes: This table presents the dynamic reassignment name list related to Conoco-Phillips. ID_NAME is the unique standardized name id. NAME_STD is the standardized firm name. PERMNO_ADJ(0–5) is the UO firm id. A name can be matched dynamically up to 5 times (1–5) and to an additional pre-IPO owner if applicable (0). NAME_ACQ(0–5) is the related UO name. FYEAR(0–5) is the first-year for ID_NAME within PERMNO_ADJ. NYEAR(0–5) is the last-year for ID_NAME within PERMNO_ADJ.

Table 7
Conoco-Phillips: Comparison between ABS and NBER '06 Patents (1980–2006).

	(1) Diff. in means	(2)	(3) Both	(4)	(5)	(6) Only ABS	(7)
	(3) minus (6)	No. Obs.	Mean	Std. Dev.	No. Obs.	Mean	Std. Dev.
Forward patent cites	-2.744	205	14.961	14.702	4,054	17.704	29.002
Yearly weighted forward patent cites	0.138	205	0.844	0.856	4,054	0.706	1.109
IPC-Year weighted forward patent cites	0.099	205	0.972	1.020	4,054	0.873	1.370

Notes: This table presents mean comparison tests for patents that are matched by both ABS and NBER '06 vs. those that are matched only by ABS. We include all matched patents granted between 1980–2006 and related to GVKEYs 11430 and 8549. The unit of analysis is a patent. Forward patent citations are from patents granted up to 2016. ** p<0.01.

5.1.2. CONOCO and PHILLIPS PETROLEUM

This example illustrates the importance of historical names and ownership changes. In 1981 Conoco was acquired by Dupont, which later spun it off as a publicly traded company, which was eventually acquired in 2002 by the publicly traded company Phillips Petroleum. The merged entity was renamed ConocoPhillips. We use the CRSP monthly stock file to locate all historical names of related securities. Table 6 shows the name list for Conoco-Phillips.

Three important changes resulted from this name list. First, we located more than 4,000 patents issued to Phillips Petroleum prior to the merger with Conono, which were not previously matched. We assigned these patents to Phillips Petroleum and included in its patent stock until 2002, and thereafter included in ConocoPhillips patent stock. Second, we dynamically reassign the patent stock. For example, U.S. patent Num. "5404954" was granted to Conoco Inc in 1995. At that time, Conoco was a subsidiary of Dupont. We included in Dupont's patent flow for 1995, and in Dupont's patent-stock for 1996–1997. However, from 1998, when Conoco is spun-off as an independent publicly traded company, this patent was subtracted from Dupont and added to Conoco's patent stock, and in 2002 transferred to ConocoPhillips patent stock. Lastly, while the majority of Phillips Petroleum patents are pre-Conoco merger, patents were granted under the Phillips Petroleum name up to 6 years after the merger with Conoco. This emphasizes the importance of dynamic name matching.

Table 7 presents mean comparison tests for patents that are matched by both ABS and NBER '06 versus those that are matched only by ABS (more than 4,000 patents) for all related GVKEYs during 1980–2006. We don't find a significant difference in terms of forward citations between the two groups.

6. Econometric analysis

The standard measurement error model assumes that measurement error is not correlated with either the dependent variable or any independent variable. We generalize the standard model to examine the implications of omitting patents, and of failing to dynamically reassign patents, for estimates of the market value of patent stock, and R&D elasticity of patents.

6.1. Sources of measurement error and likely direction of bias: A simple econometric framework

Using boldface to represent vectors, let $\mathbf{Y}$ represent Tobin's Q (i.e., market value over assets) so that Y_{it} represents the Tobin's Q for firm i and time t. Similarly, let $\mathbf{X}$ represent patent stock over assets. Ignoring control variables, the typical regression is of the form

$$Y_{it} = \alpha_0 + \alpha_1 X_{it} + \epsilon_{it} \tag{1}$$

Suppose X_{it} is measured with error. The error-laden variable is denoted by $X^*_{it} = X_{it} + m_{it}$, where m_{it} is measurement error. The actual specification estimated is

$$Y_{it} = \alpha_0 + \alpha_1 X^*_{it} + \epsilon_{it} \tag{2}$$

Let $\mathbf{y} = \mathbf{Y} - \overline{\mathbf{Y}}$, where $\overline{\mathbf{Y}}$ is the sample mean of $\mathbf{Y}$. Define $\mathbf{x}^*$ similarly. We arrive at a simplified specification in deviation form, where we suppress the time subscript to avoid clutter.

$$\boldsymbol{y} = \alpha_1^* \boldsymbol{x}^* + \boldsymbol{\epsilon} \tag{3}$$

Let a_1 be the OLS estimate of α_1 in 1, and a_1^* be the corresponding estimate in 3. Then

$$\begin{aligned} a_1^* &= \frac{Cov(\mathbf{x}^*, \mathbf{y})}{Var(\mathbf{x}^*)} = \frac{Cov(\mathbf{x}, \mathbf{y}) + Cov(\mathbf{m}, \mathbf{y})}{Var(\mathbf{x}^*)} \\ &= \frac{Cov(\mathbf{x}, \mathbf{y})}{Var(\mathbf{x})}\frac{Var(\mathbf{x})}{Var(\mathbf{x}^*)} + \frac{Cov(\mathbf{m}, \mathbf{y})}{Var(\mathbf{x}^*)} \\ &= a_1 \frac{Var(\mathbf{x})}{Var(\mathbf{x}^*)} + \frac{Cov(\mathbf{m}, \mathbf{y})}{Var(\mathbf{x}^*)} \end{aligned} \tag{4}$$

$$E(a_1^*) = \alpha_1 \frac{Var(\mathbf{x})}{Var(\mathbf{x}^*)} + \frac{Cov(\mathbf{m}, \mathbf{y})}{Var(\mathbf{x}^*)} \tag{5}$$

Equation 5 highlights the simplest form of expressing the sources of variation in OLS estimates due to measurement error.[17] Measurement error is typically assumed to be uncorrelated with the dependent variable, $\mathbf{y}$, which implies that the second term on the right hand side of equation 5 is zero. If the measurement error is also uncorrelated with the independent variable,$\mathbf{x}$, so that $Var(\mathbf{x}^*) = \mathbf{Var}(\mathbf{x}) + Var(\mathbf{m})$, then $\frac{Var(\mathbf{x})}{Var(\mathbf{x}^*)} < 1$. Under these assumptions, $a_1{}^* < a_1$, we get the classical result that measurement error leads to attenuation bias.

However, depending on the source, measurement error $\mathbf{m}$ may be correlated with either or both $\mathbf{x}$ and $\mathbf{y}$. The nature of the possible bias differs between the two terms in equation 5. The first term deals with the magnitude of the coefficient estimate, and will inflate the magnitude of the coefficient if $Var(\mathbf{x}^*) < \mathbf{Var}(\mathbf{x})$, or deflate it if $Var(\mathbf{x}^*) > \mathbf{Var}(\mathbf{x})$. On the other hand, the second term can either bias up (positive bias), if $Cov(\mathbf{m}, \mathbf{y}) > \mathbf{0}$, or down (negative bias) if $Cov(\mathbf{m}, \mathbf{y}) < \mathbf{0}$

6.2. Matching errors, and failure to dynamically reassign patents

6.2.1. Dynamic reassignment bias

We work through a simple case to provide an intuitive understanding of the possible bias when patents are not dynamically reassigned. Mergers and acquisitions, and divestitures are an important source of failure of dynamic reassignment. Suppose a subsidiary is sold by firm s to firm b, but the stock of patents is not adjusted. Thus firm s's stock of patents per dollar of assets (X_s) is higher than the correct number by c, and firm b's is lower. Let S represent the set of sellers, and B represents the set of buying firms. Further, suppose that there are K buying and selling firms out of a sample of N, and let $\gamma = \frac{K}{N}$. Measurement error can be represented as

[17] We ignore the complications in replacing sample moments with population moments in moving from 4 to 5

$$m_i = \begin{cases} c & \text{if } i \in S \\ -c & \text{if } i \in B \\ 0 & \text{otherwise.} \end{cases}$$

Equation 5 points to two key quantities: $Var(\mathbf{x}^*)$ and $Cov(\mathbf{m}, \mathbf{y})$. Let $\overline{Y_s}$ represent the mean for $i \in S$, and likewise for $\overline{Y_b}$, $\overline{X_s}$, and $\overline{X_b}$

$$Var(\mathbf{x}^*) = Var(\mathbf{x}) + Var(\mathbf{m}) + 2Cov(\mathbf{x}, \mathbf{m})$$

$$Cov(\mathbf{x}, \mathbf{m}) = \frac{1}{N}\left(c\sum_{i \in S} X_i - c\sum_{i \in B} X_i\right) = \gamma c\left(\overline{X_s} - \overline{X_b}\right)$$

$$Cov(\mathbf{y}, \mathbf{m}) = \frac{1}{N}\left(c\sum_{i \in S} Y_i - c\sum_{i \in B} Y_i\right) = \gamma c\left(\overline{Y_s} - \overline{Y_b}\right) \tag{6}$$

As long as sellers are less patent-intensive than buyers, $Var(\mathbf{x}^*) > Var(\mathbf{x})$. Therefore, from equation 5, $|E(a_1^*)| < |\alpha_1|$, implying that the OLS estimate is biased towards zero. However, if sellers have lower Tobin Q than buyers, $Cov(\mathbf{y}, \mathbf{m}) < \mathbf{0}$. If $\alpha_1 > 0$, as one might expect, this would reinforce the downward bias. If sellers are on average more patent-intensive than buyers, then equation 6 implies that it is possible that the $Var(\mathbf{x}^*) < Var(\mathbf{x})$. Thus, the first term of 5 may be greater than α_1 in magnitude. Notice that the quantitative importance of the bias, if any, grows over time because γ, the proportion of firms involved increases.[18] In summary, if patents are not dynamically reassigned, the estimation bias will depend upon the difference between the average Tobin Q of buyers and sellers, the difference the average patent intensity of sellers and buyers, the magnitude of the mismatch, c, and on the share of firms, γ involved in transactions

6.2.2. Bias from omitted matches

We match more patents, not fewer patents. That is, most of the errors in NBER are errors of omission, and relatively few are errors of commission (assigning patents to a firm that do not belong to that firm). Suppose C_i patents are incorrectly unmatched to firm i, $i \in M$, where M is the set of firms that suffer from such errors of omission. Let $c_i = \frac{C_i}{A_i}$, where A_i be the assets of firm i. For simplicity, let $c_i = c$. The measurement error m can be represented as

$$m_i = \begin{cases} -c & \text{if } i \in M \\ 0 & \text{otherwise.} \end{cases}$$

Following equations 6, and abusing the notation to let $\gamma = \frac{M}{N}$ we get

$$Cov(\mathbf{x}, \mathbf{m}) = -\gamma c\sum_{i \in M}\left(X_i - \overline{X}\right) \tag{7}$$

$$Cov(\mathbf{y}, \mathbf{m}) = -\gamma c\sum_{i \in M}\left(Y_i - \overline{Y}\right) \tag{8}$$

If mismatched firms have higher Tobin Q than average, then equation 8 implies a negative bias. However, Equation 7 implies that if firms with missing patents are more patent-intensive than average, then it may be that $Var(\mathbf{x}^*) > Var(\mathbf{x})$. In that case, instead of attenuation bias, we would have amplification bias. If $\alpha_1 > 0$, the negative bias works against the amplification bias. The outcomes under different combinations of assumptions can be worked out similarly.

The foregoing analysis is, of course, simplified.[19] Despite its simplicity, it shows that the direction and magnitude of bias is not straightforward: It will depend on a variety of factors, which themselves depend on the source of measurement error, its magnitude, and the characteristics of the firms subject to the error. Importantly, the different sources of bias reflected in equation 5 may point in opposing directions. Simply put, the bias resulting from measurement error due to errors of omission and imperfect dynamic assignment is mostly an empirical matter.

Table 8
Summary statistics for main variables.

VARIABLES	(1) Obs.	(2) Mean	(3) Std. Dev.	(4) Distribution 10th	(5) Distribution 50th	(6) Distribution 90th
Panel A: ABS (1980–2015)						
R&D stock ($mm)	57,837	425	2,339	1.04	34.64	540
R&D expenditure ($mm)	57,837	95	498	0.41	7.99	126
Tobin's Q	57,837	5	6	0.36	1.80	20
Assets ($mm)	57,837	1,612	9,687	1.99	56.34	2,239
Sales ($mm)	57,837	2,143	11,054	2.62	116.23	3,465
Patent stock (ABS)	57,837	256	1,621	1.00	11.00	300
Patent flow (ABS)	57,837	23	135	0.00	1.00	33
Panel B: NBER '06 & ABS Match (1980–2006)						
R&D stock ($mm)	43,448	304	1,675	0.84	26.84	402
R&D expenditure ($mm)	43,448	69	365	0.35	6.19	94
Tobin's Q	43,448	4	6	0.33	1.70	17
Assets ($mm)	43,448	1,179	6,970	2.06	47.73	1,709
Sales ($mm)	43,448	1,688	8,424	2.95	100.94	2,818
Patent stock (ABS)	43,448	170	962	1.00	9.00	211
Patent flow (ABS)	43,448	18	89	0.00	1.00	28
Patent stock (NBER '06)	43,448	147	918	0.00	5.00	154
Patent flow (NBER '06)	43,448	15	85	0.00	1.00	20
Panel C: NBER '01 & NBER '06 & ABS Match (1980–1999)						
R&D stock ($mm)	29,075	242	1,335	0.65	19.94	323
R&D expenditure ($mm)	29,075	56	293	0.31	4.79	77
Tobin's Q	29,075	4	6	0.27	1.39	13
Assets ($mm)	29,075	991	4,997	2.26	44.74	1,621
Sales ($mm)	29,075	1,522	6,844	3.51	102.98	2,857
Patent stock (ABS)	29,075	131	649	1.00	7.00	180
Patent flow (ABS)	29,075	16	70	0.00	1.00	26
Patent stock (NBER '06)	29,075	112	616	0.00	4.00	132
Patent flow (NBER '06)	29,075	13	67	0.00	1.00	19
Patent stock (NBER '01)	29,075	84	528	0.00	0.00	91
Patent flow (NBER '01)	29,075	10	59	0.00	0.00	13

Notes: This table presents summary statistics of the main variables examined in the paper. Panel A consists of observations from the ABS sample; Panel B consists of observations from a sample matched on GVKEY-year pair between ABS and NBER '06; and Panel C consists of observations from a sample matched on cusip-year pair across ABS, NBER '01, and NBER '06. Patent stock is constructed by summing up the number of patents a firm owns prior to, and up to, a given year. Patent flow is the number of patents a firm produces in a given year. R&D stock, Tobin's Q, and sales are based on the ABS sample, where R&D stock is calculated using a perpetual inventory method with a 15 percent depreciation rate (Hall et al., 2005) and Tobin's Q as market value over assets.

6.3. Market value equation

In what follows, we focus the possible biases in the estimate of the coefficient of patent stock in a Tobin Q regression originating from different sources of measurement error. We include only companies that appear in the relevant datasets being compared, so that differences in estimates are not driven by changes in the sample composition (that is, extending the coverage of firms being matched in a given year, rather than the quality of the matches). Table 8 presents descriptive statistics

[18] Recall from section 5, the number of mismatched firms grows over time, from 142 in 1980 to 404 in 2006

[19] For instance, we have ignored other control variables.

Table 9
Tobin's Q and Patent Stock: NBER '01, NBER '06 and ABS.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable:	ln(Tobin's Q)								
Patent Data:	NBER '01			NBER '06			ABS		
	Pooled	Within-firms	Mismatched firms	Pooled	Within-firms	Mismatched firms	Pooled	Within-firms	Mismatched firms
Patent stock$_{t-2}$/ $Assets_t$	0.006	0.043	0.010	0.072**	0.044**	0.028	0.090**	0.049**	0.043*
	(0.013)	(0.029)	(0.046)	(0.007)	(0.015)	(0.034)	(0.006)	(0.013)	(0.018)
R&D stock$_{t-2}$ / $Assets_t$	0.162**	0.124**	0.123**	0.143**	0.117**	0.121**	0.122**	0.104**	0.100**
	(0.004)	(0.011)	(0.015)	(0.005)	(0.012)	(0.015)	(0.005)	(0.012)	(0.017)
Industry fixed effects	Yes	No	No	Yes	No	No	Yes	No	No
Firm fixed effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV sample average	3.365	3.365	3.205	3.365	3.365	3.205	3.365	3.365	3.205
Number of firms	2,764	2,764	1,162	2,764	2,764	1,162	2,764	2,764	1,162
Observations	19,897	19,897	9,463	19,897	19,897	9,463	19,897	19,897	9,463
R-squared	0.41	0.73	0.72	0.42	0.73	0.72	0.42	0.73	0.72

Notes: The sample consists of firm-year level observations matched on cusip-year pair across NBER '01, NBER '06, and ABS samples. The sample for mismatched-firm analysis includes firms whose cumulative patents in ABS is at least 10% above or below the number in NBER '06. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by cusip. ** p<0.01, * p<0.05.

Table 10
Tobin's Q and Patent Stock: NBER '06 and ABS.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	ln(Tobin's Q)							
Dataset (sample years):	NBER '06 (1980–2006)			ABS (1980–2006)			ABS(1980–2015)	ABS(2007–2015)
	Pooled	Within-firms	Mismatched firms	Pooled	Within-firms	Mismatched firms	Within-firms	Within-firms
Patent stock$_{t-2}$ /$Assets_t$	0.060**	0.031**	0.016	0.080**	0.042**	0.042**	0.043**	0.056**
	(0.004)	(0.009)	(0.019)	(0.004)	(0.008)	(0.012)	(0.007)	(0.014)
R&D stock$_{t-2}$ /$Assets_t$	0.137**	0.119**	0.123**	0.117**	0.107**	0.103**	0.110**	0.133**
	(0.003)	(0.007)	(0.010)	(0.003)	(0.008)	(0.011)	(0.006)	(0.015)
Industry fixed effecs	Yes	No	No	Yes	No	No	No	No
Firm fixed effects	No	Yes	Yes	No	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV sample average	4.125	4.125	3.757	4.125	4.125	3.757	4.348	5.001
Number of firms	3,644	3,644	1,567	3,644	3,644	1,567	4,190	1,861
Observations	31,439	31,439	15,810	31,439	31,439	15,810	43,470	9,111
R-squared	0.43	0.71	0.69	0.43	0.71	0.69	0.69	0.84

Notes: The sample consists of firm-year level observations matched on GVKEY-year pair between NBER '06 and ABS samples. The sample for mismatched-firm analysis includes firms whose cumulative patents in ABS is at least 10% above or below the number in NBER '06. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by GVKEY. ** p<0.01, * p<0.05.

for the main variables used in the estimation. We follow Bloom et al. (2013) and Arora et al., 2021 and estimate the following Tobin's Q specification:

$$\ln\left(\frac{Value_{it}}{Assets_{it}}\right) = \alpha_0 + \alpha_1\frac{Patent\ stock_{it-2}}{Assets_{it-2}} + \alpha_2\frac{R\&D\ stock_{it-2}}{Assets_{it-2}} + +\boldsymbol{\eta}_i + \boldsymbol{\tau}_t + \epsilon_{it} \tag{9}$$

Tobin's Q is market value over assets. *Patent stock*$_{it-2}$ and *R&D stock*$_{it-2}$ are measured as the stocks of patents and perpetual R&D stock, respectively.[20] We are interested in how the estimated market value of patents, $\widehat{\alpha_1}$, differs between the NBER datasets and ours. Tables 9 and 10 present the estimation results of comparing $\widehat{\alpha_1}$ across the different datasets.

Table 9 compares $\widehat{\alpha_1}$ across NBER '01, NBER '06, and ABS (our sample) for the period 1980–1999 (the sample period used in NBER '01). To isolate the effect of measurement error, only firm-years observations common across databases are included (Panel C in Table 8 provides summary statistics for these firms). Columns 1–3 present the estimates for 1980–2001 for NBER '01. Column 1 controls for industry fixed effect in a pooled regression, Column 2 add firms fixed effects, and Column 3 includes only mismatched firms – whose cumulative patents for the entire sample period in ABS is at least 10% above or below that in NBER '06. The estimated $\widehat{\alpha_1}$ ranges from 0.006 to 0.043 but is not statistically significantly different from zero. Columns 4–6 repeat the same specifications using the NBER '06 dataset. The estimates range from 0.028 to 0.072, and $\widehat{\alpha_1}$ is statistically significant except for mismatched firms. Comparing Columns 3 to Column 6 indicates that when focusing only on mismatched firms, $\widehat{\alpha_1}$ increases from 0.010 to 0.028, but remains statistically indistinguishable from zero. Columns 7–9 replicate the same specifications using ABS data. The estimated $\widehat{\alpha_1}$ ranges from 0.090 to 0.043, and all estimates are statistically significantly different from zero. In particular, the estimate for mistmatched firms is 0.043, close to the estimate for the sample as a whole, 0.049. In other words, the mismatched firms appear to be representative of the ABS dataset, but not so for the NBER datasets.

[20] To facilitate comparison between the samples, the patent stock includes only patents granted under the current UO

Table 10 compares NBER '06 to our dataset by including only firm-year observations that appear in both for the period 1980–2006.[21] The estimate of $\widehat{\alpha_1}$ obtained from NBER '06 is lower, especially for the sample of mismatched firms (0.016 versus 0.042 from Columns 3 and 6). Column 7 shows that $\widehat{\alpha_1}$ remains very similar to its 1980–2006 value when using our expanded sample of 1980–2015. Overall, improvements in measurement result in higher estimated private value of patent stock, particularly for firms whose patents are measured with error.

6.4. *Patent production function*

If measurement error in patents is systematically correlated with R&D stock, measurement error will also bias the estimates of elasticity of patenting with respect to R&D in patent production function estimates. We estimate the following patent production function to assess how the elasticity of patents with respect to R&D expenditures changes across the different samples. As before, to understand the implications of measurement error, we use firm-year observations common across all datasets, so that changes in the composition of the sample are not at work.

$$\ln(Patents)_{it} = \beta_0 + \beta_1 \ln(R\&D\ stock_{it-2}) + \beta_2 \ln(Assets)_{it-2} + \boldsymbol{\eta_i} + \boldsymbol{\tau_t} + \epsilon_{it} \tag{10}$$

Our interest is in $\widehat{\beta_1}$, elasticity of patents with respect to R&D. As discussed, a primary source of measurement error is in patents themselves. Unlike the Tobin's Q estimates, here, measurement error is in the dependent variable, which results in larger standard errors of the estimate, or a bias that depends directly on the correlation between the measurement error and some independent variable, or both. For patents flows, the most likely source of error is mismatching. Incomplete dynamic reassignment mostly affects lagged patent stocks rather than the flow of patents.

Table 11 presents the estimation results.[22] Columns 1–6 compare NBER '01, NBER '06, and ABS for the period 1980-99. Comparing Columns 1, 3, and 5, we see that the estimated elasticity of patents with respect to R&D stock is similar across the datasets for the entire sample. However, considering only the mismatched firms, we see a marked increase in the measured estimated elasticity, consistent with a negative correlation between the measurement error and R&D: Firms for which NBER failed to match patents are also likely to have higher R&D stock. This is consistent with the results in Table 3, which showed that the absolute difference between ABS patents and NBER '06 patents is positively related to R&D stock. Because matching errors are mostly errors of omission rather than commission, it follows that high R&D stock firms are likely to have more missing patents. In turn, this would lead to a downward bias in the estimated elasticity of patents with respect to R&D.

7. Summary and conclusions

This paper reports on updates and improvements to the well-known NBER patent database. We extend the database from 2006 to 2015. We also improve the accuracy of the matches by accounting for changes in company names, and changes to corporate boundaries through mergers and acquisitions of firms or their subsidiaries. Doing so enables us to improve the accuracy of the match between firms and the patents they own. This results in an increase in the number of patents we match to sample firms. It also enables us to dynamically assign patent stocks to firms as the patent assignees change owners. We find that approximately 40% of the sample firms are mismatched firms – whose cumulative patents for the entire sample period in ABS is at least 10% above or

[21] Appendix Table C2 presents results by industrial sector.

[22] Appendix Table C3 presents pooled results.

Table 11
Patent production function estimates across datasets.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent variable:	ln(1+ Number of patents)											
	1980–1999						1980–2006				1980–2015	2007–2015
	NBER '01		NBER '06		ABS		NBER '06		ABS		ABS	
Sample:	All	Mismatched	All	Mismatched	All	Mismatched	All	Mismatched	All	Mismatched	All	All
ln(R&D stock)	0.164**	0.092**	0.250**	0.140**	0.301**	0.231**	0.271**	0.139**	0.339**	0.255**	0.345**	0.283**
	(0.025)	(0.026)	(0.031)	(0.034)	(0.030)	(0.035)	(0.028)	(0.028)	(0.025)	(0.029)	(0.024)	(0.050)
ln(Assets)	0.068**	0.042**	0.102**	0.064**	0.127**	0.122**	0.095**	0.063**	0.119**	0.118**	0.121**	0.064**
	(0.012)	(0.010)	(0.013)	(0.014)	(0.014)	(0.018)	(0.010)	(0.011)	(0.010)	(0.014)	(0.009)	(0.012)
Firm fixed-effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
DV sample average	10.088	6.314	13.597	8.652	15.947	13.597	15.059	9.722	17.933	15.264	23.285	39.548
Number of firms	3,188	1,295	3,188	1,295	3,188	1,295	3,935	1,660	3,935	1,660	4,347	2,220
Observations	28,785	13,146	28,785	13,146	28,785	13,146	42,949	20,805	42,949	20,805	57,082	14,133
R-squared	0.90	0.89	0.82	0.79	0.85	0.84	0.79	0.75	0.84	0.83	0.83	0.93

Notes: The sample used for 1980–1999 consists of firm-year level observations matched on cusip-year pair across NBER '01, NBER '06, and ABS. The sample used for years 1980–2006 consists of firm-year level observations matched on GVKEY-year pair between NBER '06 and ABS. The sample used for 1980–2015 and 2007–2015 consists of firm-year observations from ABS. Standard errors (in brackets) are robust to arbitrary heteroscedasticity and allow for serial correlation through clustering by firm. ** p<0.01, * p<0.05.

below that in NBER ’06.

We explore the implications of the measurement error for two important relationships that the literature has investigated. The first is the market value of patents, and the second is the patent production function, which relates patent flow to R&D investment. We provide a simplified framework to guide intuition about possible bias. We find that measurement error results in modest under-estimates of patent value. We also find that estimates of the elasticity of patenting with respect to R&D are also under-estimated, especially for firms where measurement error is significant. Overall, these results are good news in that existing estimates reported in the literature are not biased in any significant way. However, additional research would be required to evaluate the implications for more restricted samples (where the fraction of firms affected by error may be larger). In addition, analyses of trends over time, based on the older data, may need to be re-examined, insofar as the incidence of measurement error may have increased over the sample period.

CRediT authorship contribution statement

Ashish Arora: Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **Sharon Belenzon:** Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft, Writing - review & editing. **Lia Sheer:** Conceptualization, Methodology, Data curation, Formal analysis, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

None.

Acknowledgements

We thank Jim Bessen, Wesley Cohen, Bronwyn Hall, Adam Jaffe, Manuel Trajtenberg, and seminar participants at NBER Innovation Information Initiative for helpful comments. We thank Bernardo Dionisi, Honggi Lee, Dror Shvadron, and JK Suh for excellent research assistance. We gratefully acknowledge support from NSF grant 1747316.

Appendix A. Name standardization

Prior to matching, we standardize firm names to reconcile company names that may be spelled differently across databases. 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, we standardize 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. This list is presented in Table A.1.

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 (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. 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”.

The last step in name standardization is to locate abbreviations that are commonly used by companies instead of their official names. 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.

Appendix B. Case Studies

B1. TIME-WARNER and AMERICAN ONLINE

Warner Communication was an independent and publicly traded company until its 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. This example shows that ownership can change without a name change, but also that name changes can take place without an ownership change.

A comparison with NBER ’06 reveals the following. First, Warner Communication and its related subsidiary 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. Consequently, the patent stock of Time-Warner (and later AOL Time-Warner) from patents of Warner communication and its subsidiaries (e.g., Warner Bros, WEA Manufacturing), are below the true value after 1989.

Second, Time-Warner’s patents from 1991 to 2000 (are matched incorrectly to GVKEY 25056, which during those years was solely AMERICAN-ONLINE INC. The current name associated with GVKEY 25056, TIME WARNER INC, was only adopted retroactively in 2003 when the “AOL” was dropped from the official name. Moreover, 152 AOL patents are not linked to any Compustat record. AOL-TIME WARNER related patents, on the other hand, are matched to a “Pro-Forma” Compustat record that is active for only two years 1999–2000: AOL TIME WARNER INC-PRO FORM (GVKEY 142022). All of which implies that AOL Time Warner’s flow of patents is smaller than the actual patents it owns.

We match AMERICAN-ONLINE INC (and later AOL) from 1980 until its spinout in 2009 to GVKEY 25056 and after to AOL INC (GVKEY 183920). Warner Communication is matched up to the merger with Time

Table A.1
Most frequent abbreviated words.

ADV	AEROSP	AGR	AMER	ANAL	ANALYT	ANIM	APPL	APPLICAT
ASSOC	AUTOMAT	BIOL	BIOMED	BIOPHARM	BIOSCI	BIOSURG	BIOSYS	BIOTEC
BIOTHERAPEUT	CHEM	CLIN	COMMUN	COMP	CORP	CTR	DEV	DIAGNOST
DYNAM	EDUC	ELECTR	ENGN	ENVIRONM	FAVORS	GEN	GENET	GRAPH
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

Inc to WARNER COMMUNICATIONS INC (GVKEY 11284) and later dynamically transferred to AOL -Time Warner GVKEY (25056) starting 2001. AOL -Time Warner is matched to AOL -TIME WARNER (GVKEY 25056) starting 2001 after the merger was approved. Time Inc itself is not included as an UO in our sample as it did not report any R&D, but it is included as a subsidiary name under the Time-Warner UO company.

B2. 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. Linking each patent to its correct financial record can be a concern for papers that link patents to market value.

In 1995 Pharmacia merged with Upjohn to form Pharmacia & Upjohn. In 2000, 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 newly 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 B.1 shows a complex pattern of unallocated patents, as well as misallocated patent stock and also summarizes how we deal with this case.

B3. NABISCO

Table B.2 shows that Nabisco has changed ownership four times over the sample period. In 1981 Nabisco merged with the publicly traded company Standard Brands to form Nabisco Brands. In 1985 R.J. Reynolds merged with Nabisco Brands to create RJR Nabisco, which eventually became Nabisco Group holding after the tobacco 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. Table B.2 lays out how patents are dynamically reassigned.

Examining NBER '06, Table B.3 shows that 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", prior 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 tobacco 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 patents.

Table B.1
PHARMACIA & UPJOHN and MONSANTO dynamic match.

Period	related GVKEY	Relevant Compustat name for period	Most recent Compustat name	Comments	Patent flow per our strategy	Patent flow per 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 agriculture 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 fillings 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)

*Notes:*This table presents the comparison between NBER '06 and our data for dynamic patent reassignment for Phamacia-Monsanto related patents at the GVKEY-Period level. Most recent Compustat name is based on Compustat 2018 file. Relevant Compustat name for the period is the historical firm name based on CRSP Monthly Stock file. Patent flow per our strategy is based on NBER raw patent match data for the relevant Compustat name excluding subsidiaries. Patent flow per original NBER match is based on NBER '06 data.

Table B.2
Nabisco dynamic match.

Years	related GVKEY	Original Compustat owner	Current Compustat name	Comments
1981–1985	7674	STANDARD BRANDS INC	NABISCO BRANDS INCO	1981: Standard Brands company merged with Nabisco Inc to form Nabisco Brands Inc
1986–1999	9113	R J REYNOLDS IND INC	NABISCO GROUP HOLDINGS CORP	1985: R.J. Reynolds Industries merged with Nabisco Brands to form R J R Nabisco Inc
2000	8543	PHILIP MORRIS COS INC	ALTRIA GROUP INC	2000: Nabisco was acquired by Phillip Morris
2001–2015	142953	KRAFT FOOD INC	MONDELEZ INTERNATIONAL INC	2001: Kraft together with Nabisco split from Phillip Morris

Notes: This table presents the dynamic reassignment for Nabisco related patents at the GVKEY-Period level. Current Compustat name is based on Compustat 2018 file. Original Compustat owner for the period is the historical firm name based on CRSP Monthly Stock file.

Table B.3
Data Entry for Nabisco in NBER '06.

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		

Notes: PDPCO is NBER's Patent Data Project (PDP) unique com y id. FIRSTYR is the first year GVKEY company has data. LASTYR is the last year a GVKEY company has data. PDPSEQ is the GVKEY sequence within PDPCO. If PDPSEQ=-1, the related GVKEY is disregarded. BEGYR is the beginning year for GVKEY within PDPCO. ENDYR is the last year for GVKEY within PDPCO. All patents related to Nabisco will be accounted under GVKEY 9113 from 1950 to 1999, while all other related GVKEYs are disregarded.

Furthermore, NBER '06 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).

B4. CHEMTURA CORPORATION

This case shows how having historical names helps 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.

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

The complete set of historical names enables us to locate all the relevant M&As throughout the years of the publicly traded firms that exist as an independently traded company, and 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 '06 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 one patent was matched for that name. In addition, none of the acquired publicly traded companies were dynamically transferred to CHEMTURA CORP post-acquisition. By considering all previous names related to GVKEY 3607: (i) Crompton & Knowles Corp; (ii) CK Witco Corp and (iii) Crompton Corp, 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).[23] 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 three patents are in 2006, the year after the company was acquired). Overall, we find a patent stock of 1,187 patents in 2006 for GVKEY 3607 as opposed to 1 patent in NBER.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.respol.2021.104217.

References

Acs, Z.J., Anselin, L., Varga, A., 2002. Patents and innovation counts as measures of regional production of new knowledge. Res Policy 31 (7), 1069–1085.
Arora, A., Belenzon, S., Rios, L.A., 2014. Make, buy, organize: The interplay between research, external knowledge, and firm structure. Strategic Management Journal 35 (3), 317–337.
Arora, A., Belenzon, S., Sheer, L., 2021. Knowledge spillovers and corporate investment in scientific research. American Economic Review 111 (3), 871–898.
Bena, J., Ferreira, M.A., Matos, P., Pires, P., 2017. Are foreign investors locusts? The long-term effects of foreign institutional ownership. J Financ Econ 126 (1), 122–146.
Bessen, J., 2008. The value of us patents by owner and patent characteristics. Res Policy 37 (5), 932–945.
Bessen, J., 2009. NBER PDP project user documentation. Unpublished documentation.
Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica 81 (4), 1347–1393.
Chen, S.-S., Chen, Y.-S., Liang, W.-l., Wang, Y., 2020. Public r&d spending and cross-sectional stock returns. Res Policy 49 (1), 103887.
de Leeuw, T., Gilsing, V., Duysters, G., 2019. Greater adaptivity or greater control? Adaptation of IOR portfolios in response to technological change. Res Policy 48 (6), 1586–1600.

[23] This example does not include subsidiaries and acquired companies, which are, however, included in ABS.

Dushnitsky, G., Lenox, M.J., 2005. When do incumbents learn from entrepreneurial ventures?: corporate venture capital and investing firm innovation rates. Res Policy 34 (5), 615–639.
Fabrizio, K.R., 2009. Absorptive capacity and the search for innovation. Res Policy 38 (2), 255–267.
Figueroa, N., Serrano, C.J., 2019. Patent trading flows of small and large firms. Res Policy 48 (7), 1601–1616.
Hall, B.H., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. RAND Journal of economics 16–38.
Hall, B. H., Jaffe, A. B., Trajtenberg, M., 2001. The NBER patent citation data file: Lessons, insights and methodological tools. NBER Working Papers: No. 8498.
Im, H.J., Shon, J., 2019. The effect of technological imitation on corporate innovation: Evidence from U.S. patent data. Res Policy 48 (9), 103802.
Kogan, L., Papanikolaou, D., Seru, A., Stoffman, N., 2017. Technological innovation, resource allocation, and growth. Q J Econ 132 (2), 665–712.
Cohen, W.M., Klepper, S., 1996. Firm size and the nature of innovation within industries: The case of process and product R&D. Rev Econ Stat 232–243.
Li, G.-C., Lai, R., D'Amour, A., Doolin, D.M., Sun, Y., Torvik, V.I., Amy, Z.Y., Fleming, L., 2014. Disambiguation and co-authorship networks of the us patent inventor database (1975–2010). Res Policy 43 (6), 941–955.
Marx, M., Fuegi, A., 2020. Reliance on science: Worldwide front-page patent citations to scientific articles. Strategic Management Journal 41 (9), 1572–1594.
Marx, M., Strumsky, D., Fleming, L., 2009. Mobility, skills, and the michigan non-compete experiment. Manage Sci 55 (6), 875–889.
McGahan, A.M., Silverman, B.S., 2006. Profiting from technological innovation by others: The effect of competitor patenting on firm value. Res Policy 35 (8), 1222–1242.
Mezzanotti, F., Simcoe, T., 2019. Patent policy and American innovation after ebay: An empirical examination. Res Policy 48 (5), 1271–1281.
Miller, D.J., Fern, M.J., Cardinal, L.B., 2007. The use of knowledge for technological innovation within diversified firms. Academy of Management Journal 50 (2), 307–325.
Raiteri, E., 2018. A time to nourish? evaluating the impact of public procurement on technological generality through patent data. Res Policy 47 (5), 936–952.
Silvestri, D., Riccaboni, M., Della Malva, A., 2018. Sailing in all winds: Technological search over the business cycle. Res Policy 47 (10), 1933–1944.
Wu, Y., 2010. What's in a name? What leads a firm to change its name and what the new name foreshadows. Journal of Banking & Finance 34 (6), 1344–1359.