Online Appendix to "Measuring Technological Innovation over the Long Run" by Kelly, Papanikolaou, Seru, and Taddy

We briefly overview our conversion of unstructured patent text data into a numerical format suitable for statistical analysis. To begin, we build our collection of patent documents from two sources. The first is the USPTO patent search website, which records all patents beginning from 1976. Our web crawler collected the text content of patents from this site, which includes patent numbers 3,930,271 through 9,113,586. The records in this sample are comparatively easy to process as they are available in HTML format with standardized fields.

For patents granted prior to 1976, we collect patent text from our second main datasource, Google's patent search engine. For the pre-1976 patent records, we recover all of the fields listed above with the exception of inventor/assignee addresses (Google only provides their names), examiner, and attorney. Some parts of our analysis rely on firm-level aggregation of patent assignments. We match patents to firms by firm name and patent assignee name. Our procedure broadly follows that of Kogan et al. (2017) with adaptations for our more extensive sample. In addition to the citation data we scrape from Google, we obtain complementary information on patent citations from Berkes (2016) and the USPTO. The data in Berkes (2016) includes citations that are listed inside the patent document and which are sometimes missed by Google. Nevertheless, the likelihood of a citation being recorded is significantly higher in the post-1945 than in the pre-1945. When this consideration is relevant, we examine results separately for the pre- and post-1945 periods.

To represent patent text as numerical data, we convert it into a *document term matrix* (DTM), denoted C. Columns of C correspond to words and rows correspond patents. Each element of C, denoted c_{pw}, counts the number of times a given one-word phrase (indexed by w) is used in a particular patent (indexed by p), after imposing a number of filters to remove stop words, punctuation, and so forth. We provide a detailed step-by-step account of our DTM construction in Appendix . Our final dictionary includes 1,685,416 terms in the full sample of over nine million patents.

The next section provides additional details on the data construction, including the process through which we convert the text of patent documents to a format that is amenable to constructing similarity measures.

A. Text Data Collection, Additional Details

The Patent Act of 1836 established the official US Patent Office and is the grant year of patent number one.[1] We construct a dataset of textual content of US patent granted during the 180 year period from 1836-2015. Our dataset is built on two sources.

The first is the USPTO patent search website. This site provides records for all patents beginning in 1976. We designed a web crawler collect the text content of patents over this period, which includes patent numbers 3,930,271 through 9,113,586. We capture the following fields from each record:

1. Patent number (WKU)
2. Application date
3. Granted date
4. Inventors
5. Inventor addresses
6. Assignees
7. Assignee addresses
8. Family ID
9. Application number
10. US patent class
11. CPC patent class
12. Intl. patent class
13. Backward citations
14. Examiner
15. Attorney
16. Abstract
17. Claims
18. Description

The only information available from USPTO that we do not store are image files for a patent's "figure drawing" exhibits.

For patents granted prior to 1976, the USPTO also provides bulk downloads of .txt files for each patent. The quality of this data is inferior to that provided by the web search interface in three ways. First, the text data is recovered from image files of the original patent documents using OCR scans. OCR scans often contain errors. These generally arise from imperfections in the original images that lead to errors in the OCR's translation from image to text. Going backward in time from 1976, the quality of OCR scans deteriorates rapidly due to lower quality typesetting. Second, the bulk download files do not use a standardized format which makes it difficult to parse out the fields listed above.

Rather than using the USPTO bulk files, we collect text of pre-1976 patents from our second main datasource, Google's patent search engine. Like post-1976 patents from USPTO, Google provides patent records in an easy-to-parse HTML format that we collect with our web crawler. Furthermore, inspection of Google records versus 1) OCR files from the USPTO and 2) pdf images of patents that are the source of the OCR scans, reveals that in this earlier period Google's patent text is more accurate than the OCR text in USPTO bulk data. From Google's pre-1976 patent records, we recover all of the fields listed above with the exception of inventor/assignee addresses (Google only provides their names), examiner, and attorney.

B. Cleaning Post-1976 USPTO Data

Next, we conduct a battery of checks to correct data errors. For the most part, we are able to capture and parse of patent text from the USPTO web interface without error. When

[1]The first patent was granted in the US in 1790, but of the patents granted prior to the 1836 Act, all but 2,845 were destroyed by fire.

there are errors, it is almost always the case that the patent record was incompletely captured, and this occurs for one of two reasons. The first reason is that the network connection was interrupted during the capture and the second is that the patent record on the UPSTO website is itself incomplete (in comparison with PDF image files of the original document, which are also available from USPTO via bulk download).

Our primary data cleaning task was to find and complete any partially captured patent records. First, we find the list of patent numbers (WKUs) that are entirely missing from our database, and re-run our capture program until all have been recovered. Many of the missing records that we find are explicitly labeled as "WITHDRAWN" at the USPTO. [2] Next, we identify WKUs with an entirely missing value for the abstract, claims, or description field. Fortunately, we find this to be very infrequent, occurring in less than one patent in 100,000, making it easy for us to correct this manually.

Next, a team of research assistants (RA's) manually checked 3,000 utility patent records, 1,000 design patent records, and 1,000 plant patents records against their PDF image files. The RA task is to identify any records with missing or erroneous information in the reference, abstract, claims, or description fields. To do this, they manually read the original pdf image for the patent and our digitally captured record. We identify patterns in partial text omission and update our scraping algorithm to reflect these. We then re-ran the capture program on all patents and confirmed that omissions from the previous iteration were corrected.

C. Cleaning Pre-1976 Google Data

Fortunately, we find no instances of missing WKU's or incomplete text from Google web records. Next, we assess the accuracy of Google's OCR scans by manually re-scanning a random sample of 1,000 pre-1976 patents using more recent (and thus more accurate) ABBYY OCR software than was used for most of Google's image scans. We compare the ABBYY scan to the pdf image to confirm the scan content is complete, the compare the frequency of garbled terms in our scan versus that OCR text from Google. The distribution of pairwise cosine similarities in our ABBYY text and Google's OCR is reported below.

[2] Withdrawn information can be found at `https://www.uspto.gov/patents-application-process/patent-search/withdrawn-patent-numbers`.

Cosine Similarity	
mean	0.957
std	0.073
P1	0.701
P5	0.863
P10	0.900
P25	0.951
P50	0.977
P75	0.991
P90	0.996
P95	0.998
P99	0.999
N	1000

Only 10% of sampled Google OCR records have a correlation with ABBYY below 90%.

Next, we manually compare both our OCR scans and those from Google against the pdf image. We find that garble rate for ABBYY OCRed is 0.025 on average, with standard deviation of 0.029. We find that Google has only slightly more frequent garbling than our ABBYY scans. Of the term discrepancies in the two sets of scans, around 52% of these correspond to a garbled ABBYY records and 83% to a garbled Google record. We ultimately conclude that Google's OCR error frequency is acceptable for use in our analysis.

D. Conversion from Textual to Numeric Data

We convert the text content of patents into numerical data for statistical analysis. To do this, we use the NLTK Python Toolkit to parse the "abstract," "claims," and "description" sections of each patent into individual terms. We strip out all non-word text elements, such as punctuation, numbers, and HTML tags, and convert all capitalized characters to lowercase. Next, we remove all occurrences of 947 "stop words," which include prepositions, pronouns, and other words that carry little semantic content.[3]

The remaining list of "unstemmed" (that is, without removing suffixes) unigrams amounts to a dictionary of 35,640,250 unique terms. As discussed in Gentzkow, Kelly, and Taddy (2017),

[3]We construct our stop word list as the union of terms in the following commonly used lists:

http://www.ranks.nl/stopwords
https://dev.mysql.com/doc/refman/5.1/en/fulltext-stopwords.html
https://code.google.com/p/stop-words/
http://www.lextek.com/manuals/onix/stopwords1.html
http://www.lextek.com/manuals/onix/stopwords2.html
http://www.webconfs.com/stop-words.php
http://www.text-analytics101.com/2014/10/all-about-stop-words-for-text-mining.html
http://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_170.html
https://pypi.python.org/pypi/stop-words
https://msdn.microsof,t.com/zh-cn/library/bb164590
http://www.nltk.org/book/ch02.html (NLTK list)

an important preliminary step to improve signal-to-noise ratios in textual analysis is to reduce the dictionary by filtering out terms that occur extremely frequently or extremely infrequently. The most frequently used words show up in so many patents that they are uninformative for discriminating between patent technologies. On the other hand, words that show up in only a few patents can only negligibly contribute to understanding broad technology patterns, while their inclusion increases the computational cost of analysis.[4]

We apply filters to retain influential terms while keeping the computational burden of our analysis at a manageable level, and focus on the number of distinct patents and calendar years in which terms occur. A well known attribute of text count data is its sparsity—most terms show up very infrequently—and the table shows that this pattern is evident in patent text as well. We exclude terms that appear in fewer than twenty out of the more than nine million patents in our sample. These eliminate 33,954,834 terms, resulting in a final dictionary of 1,685,416 terms.[5]

After this dictionary reduction, the entire corpus of patent text is reduced in a $D \times W$ numerical matrix of term counts denoted C. Matrix row d corresponds to patent (WKU) d. Matrix column w corresponds the w^{th} term in the dictionary. Each matrix element c_{dw} the count of term w in patent d.

Pairwise similarities constitute a high-dimensional matrix of approximate dimension 9 million $\times$ 9 million. To reduce the computational burden, we set similarities below 5% to zero. This affects 93.4% of patent pairs. Patents with such low text similarity are, for all intents and purposes, completely unrelated, yet introduce a large computational load in the types of analyses we pursue. Replacing these approximate zeros with similarity scores of exactly zero achieves large computational gains by allowing us to work with sparse matrix representations that require substantially less memory. Our empirical findings are insensitive to this threshold as they are driven primarily by the highest similarity pairs. In experiments with similarity cutoffs ranging from 5% to 10%, we find results that are quantitatively indistinguishable.

E. Matching Patents to Firms

Much of our analysis relies on firm-level aggregation of patent assignments. We match patents to firms by merging firm names and patent assignee names. Our procedure broadly follows that of Kogan et al. (2017) with adaptations for our more extensive sample. It combines matching algorithms with extensive manual checking.

The first step is extracting assignee names from patent records. For post-1976 data we use information from the USPTO web search to identify assignee names. Due to the high data quality in this sample, assignee extraction is straightforward and highly accurate. For

[4]Filtering out infrequent words also removes garbled terms, misspellings, and other errors, as their irregularity leads them to occur only sporadically.

[5]The table also shows that there are some terms that appear in almost all patents. Examples of the most frequently occurring words (that are not in the stop word lists) are "located," "process," and "material." Because these show up in most patents they are unlikely to be informative for statistical analysis. These terms are de-emphasized in our analysis through the $TFIDF$ transformation.

pre-1976, we use assignee information from Google patent search. While it is easy to locate the assignee name field thanks to the HTML format, Google's assignee names are occasionally garbled by the OCR.

Next, we clean the set of extracted assignee names. There are 766,673 distinct assignees in patents granted since 1836. Most of the assignees are firm names and those that are not firms are typically the names of inventors. We clean assignee name garbling using fuzzy matching algorithms. For example, the assignee "international business machines" also appears as an assignee under the names "innternational business machines," "international businesss machines," and "international business machiness." Garbled names are not uncommon, appearing for firms as large as GE, Microsoft, Ford Motor, and 3M.

We primarily rely on Levenshtein edit distance between assignees to identify and correct erroneous names. There are two major challenges to overcome in name cleaning. The first choosing a distance threshold for determining whether names are the same. As an example, the assignees "international business machines" (recorded in 103,544) and "ibm" (recorded in 547 patents) have a large Levenshtein distance. To address cases like this, we manually check the roughly 3,000 assignee names that have been assigned at least 200 patents, correcting those that are variations on the same firm name (including the IBM, GE, Microsoft, Ford, and 3M examples). Next, for each firm on the list of most frequent assignees, we calculate the Levenshtein distance between this assignee name and the remaining 730,000+ assignee names, and manually correct erroneous names identified by the list of assignees with short Levenshtein distances.

The second challenge is handling cases in which a firm subsidiary appears as assignee. For example, the General Motors subsidiary "gm global technology operations" is assigned 8,394 patents. To address this, we manually match subsidiary names from the list of top 3,000+ assignees to their parent company by manually searching Bloomberg, Wikipedia, and firms' websites.

After these two cleaning steps, and after removing patents with the inventor as assignee, we arrive at 3,036,859 patents whose assignee is associated with a public firm in CRSP/Compustat, for a total of 7,467 distinct cleaned assignee firm names. We standardized these names by removing suffixes such as "com," "corp," and "inc," and merge these with CRSP company names. Again we manually check the merge for the top 3,000+ assignees, and check that name changes are appropriately addressed in our CRSP merging step. Finally, we also merge our patent data with Kogan et al. (2017) patent valuation data for patents granted between 1926 and 2012.

F. Breakthrough Innovation and Measured Productivity

Here, we relate our innovation indices to measured productivity.

1. Data Sources

The US population data is from the U.S. Census Bureau.[6] The aggregate TFP data are from Basu et al. (2006).[7] The industry-level productivity series is from the Bureau of Labor Statistics.[8] The historical productivity data are from Kendrick (1961).[9] Last, the patent value metrics are from Kogan et al. (2017).[10]

2. Aggregate Productivity

For the post-war sample, we use the aggregate TFP measure constructed by Basu et al. (2006), which is available over the 1948-2018 period. For the earlier sample, we measure productivity using output per hour data collected by Kendrick (1961), which is available for the 1889 to 1957 period. Following Jorda (2005), we estimate:

$$\frac{1}{\tau}\left(x_{t+\tau}-x_t\right)=a_0+a_\tau\,\log \mathrm{BreakthroughIndex}_t+c_\tau\,\mathbf{Z}_t+u_{t+\tau}, \tag{1}$$

where x_t is log productivity, $\mathrm{BreakthroughIndex}_t$ refers to our innovation index, and $\mathbf{Z}_t$ is a vector of controls that includes the log number of patents per capita and the level of productivity. We consider horizons of up to $\tau = 10$ years and adjust the standard errors for serial correlation using the Newey-West procedure with $\tau + 1$ lags. All independent variables are normalized to unit standard deviation. To ensure that we are not capturing pre-existing trends, we also examine negative values of τ.

Panel A of Figure A.3 presents the results of estimating (1) for the post-war sample. We see that a one-standard deviation increase in our index is associated with 0.5 percent faster annual TFP growth, with some delay. This is substantial given that the standard deviation in measured TFP growth over this period is 1.8%. Panel B shows the results for the earlier sample. Again, we see that a one-standard deviation increase in our innovation index is associated with an increase in labor productivity growth of approximately 1.5–2% per year—compared to an annual standard deviation of 5.2% for labor productivity growth.

[6] We splice together three time series: i) pre-1900 data: U.S. Census Bureau, Decennial Census, retrieved from the U.S. Census Bureau; `https://www.census.gov/population/www/censusdata/files/table-2.pdf`, (accessed August 1, 2016); ii) data from 1900 to 1999: U.S. Bureau of Economic Analysis, Population [POPTHM], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/POPTHM`, (accessed August 1, 2016); and iii) post-1999 data: U.S. Census Bureau, National Population [POPH], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/POPH`, (accessed August 1, 2016).

[7] See Susanto Basu, John Fernald, and Miles Kimball. 2006. "Are Technology Improvements Contractionary?" American Economic Review. `https://www.frbsf.org/economic-research/files/quarterly_tfp.xlsx`, (accessed September 15, 2017).

[8] Bureau of Labor Statistics, Industry Multi-factor Productivity, retrieved from `https://download.bls.gov/pub/time.series/ip/`, (accessed September 19, 2017).

[9] Kendrick, J. W. (1961). Productivity Trends in the United States. National Bureau of Economic Research, Inc. Data hand collected from Tables D-II and D-V.

[10] Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017). Technological innovation, resource allocation, and growth. The Quarterly Journal of Economics; `https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth` (accessed October 1, 2019).

3. Sector-level Productivity

First, we examine how the distribution across technology class of breakthrough patents varies over time. Panel A of Appendix Figure A.2 shows the technology classes in which breakthrough inventions originated has varied substantially over the last 170 years. By contrast, Panel B shows that the composition of technology classes among all patents has remained relatively stable over time.

We next construct indices of innovation at the sector level. One issue that arises is how to map patents to industries in a way that is independent of the presence of an explicit assignee, since clean asignee identity and names are notoriously difficult to pin down. To address this, we exploit the mapping between patent technology classifications (CPC) and industry classifications constructed by Goldschlag et al. (2016). Because this is a probabilistic mapping (there is no one-to-one correspondence between CPC and industry codes), we assign a fraction of each patent to industry codes based on the given probability weights associated with its (4-digit) CPC technology classification.[11]

Panel A of Figure A.4 presents our results for the period from 1987 to the present. We use estimates of multi-factor productivity at the NAICS 4-digit level from the Bureau of Labor Statistics (BLS), which covers 86 manufacturing industries.[12] We then estimate a panel analogue of equation (1),

$$\frac{1}{\tau}\left(x_{i,t+\tau} - x_{i,t}\right) = a_0 + a_\tau \, \log \text{BreakthroughIndex}_{i,t} + c_\tau \, \mathbf{Z}_{i,t} + u_{i,t+\tau}, \tag{2}$$

except that now $\mathbf{Z}_{i,t}$ also includes time and industry fixed effects. Standard errors are clustered by industry. Given the shorter length of this sample, we consider horizons of $\tau = 1 \ldots 5$ years.

We find a strongly statistically positive relation between our innovation index and future productivity growth—while the relation with past productivity growth is insignificant. In terms of magnitudes, a one-standard deviation increase in our innovation index is associated with approximately 1–1.2% higher productivity growth per year, over the next 5 years.

Panel B performs a similar exercise for the earlier sample. We use the labor productivity data collected by Kendrick (1961), which covers 62 manufacturing industries for the years 1899, 1909, 1919, 1937, 1947, and 1954. Since the data is only available at discrete periods, we modify our approach: for each period $(t, t+\tau)$, we regress the annualized difference in log labor productivity on the log of the accumulated level of innovation (number of breakthrough patents) in $t \pm 2$ years. We again see a strong and statistically significant relation between our industry innovation indices and measured productivity: a one standard deviation increase in

[11] Two caveats are in order. First, this mapping is based on post-1970 data, whereas our analysis spans the entire period since the 1840s. Hence, there might be measurement error in our index since we assign a fraction of patents to each of the industries that map to a CPC classification based on the weights estimated from only part of the sample. Second, this mapping is primarily available for manufacturing industries—which are however the industries that patent most heavily.

[12] Susanto Basu, John Fernald, and Miles Kimball. 2006. "Are Technology Improvements Contractionary?" American Economic Review. `https://www.frbsf.org/economic-research/files/quarterly_tfp.xlsx`, (accessed September 15, 2017).

our innovation index is associated with a 1.4% higher growth rate in measured productivity over the next period.

For comparison, we also construct a corresponding index based on citations (measured over a 10 year horizon). Examining Panels A and B of Appendix Figure A.5, we see that there is no statistically significant relation between the citations-based index and industry productivity in either sample period.

References

Basu, S., J. G. Fernald, and M. S. Kimball (2006). Are technology improvements contractionary? *American Economic Review 96*(5), 1418–1448. Retrieved from `https://www.frbsf.org/economic-research/files/quarterly_tfp.xlsx`, (accessed September 15, 2017).

Berkes, E. (2016). Comprehensive universe of U.S. Patents (cusp): Data and Facts. Working paper, Northwestern University. Unpublished Data

Bureau of Labor Statistics, Industry Multi-factor Productivity, retrieved from `https://download.bls.gov/pub/time.series/ip/`, (accessed September 19, 2017).

Google Patents Search Engine, retrieved from https://patents.google.com/ (accessed July 2016)

Goldschlag, N., T. J. Lybbert, and N. J. Zolas (2016). An 'algorithmic links with probabilities' crosswalk for uspc and cpc patent classifications with an application towards industrial technology composition. CES Discussion Paper 16-15, U.S. Census Bureau. retrieved from `https://sites.google.com/site/nikolaszolas/PatentCrosswalk` (accessed January 13, 2017).

Jorda, O. (2005, March). Estimation and inference of impulse responses by local projections. *American Economic Review 95*(1), 161–182.

Kendrick, J. W. (1961). *Productivity Trends in the United States.* National Bureau of Economic Research, Inc. Data hand collected from Tables D-II and D-V.

Kogan, L., D. Papanikolaou, A. Seru, and N. Stoffman (2017). Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics 132*(2), 665–712. retrieved from `https://github.com/KPSS2017/Technological-Innovation-Resource-Allocation-and-Growth` (accessed October 1, 2019).

U.S. Census Bureau, National Population [POPH], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/POPH`, (accessed August 1, 2016).

U.S. Bureau of Economic Analysis, Population [POPTHM], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/POPTHM`, (accessed August 1, 2016).

U.S. Census Bureau, Decennial Census, retrieved from the U.S. Census Bureau; `https://www.census.gov/population/www/censusdata/files/table-2.pdf`, (accessed August 1, 2016).

Patent Search System, https://www.uspto.gov/patents-application-process/search-patents (accessed July 2016)

Appendix Material

Table A.1: Important Patents

Patent	Year	Inventor	Invention	Citations	Percentile Ranks						Source
					No Adjustment			Remove year FE			
					Quality	Citations		Quality	Citations		
				(total)	(0-10)	(0-10)	(total)	(0-10)	(0-10)	(total)	
1647	1840	Samuel F. B. Morse	Morse Code	2	0.03	-	0.29	0.03	0.64	0.81	Reference
3237	1843	Nobert Rillieux	Sugar Refining	0	0.80	-	-	0.84	0.64	0.44	Reference
3316	1843	Samuel F. B. Morse	telegraphy wire	0	0.97	-	-	0.99	0.64	0.44	Reference
3633	1844	Charles Goodyear	Vulcanized Rubber	3	0.99	-	0.38	0.98	0.64	0.88	Reference
4453	1846	Samuel F. B. Morse	telegraph battery	0	1.00	-	-	0.99	0.64	0.44	Reference
4750	1846	Elias Howe, Jr.	Sewing Machine	1	1.00	-	0.17	0.99	0.64	0.70	Reference
4834	1846	Benjamin Franklin Palmer	Artificial Limb	0	0.99	-	-	0.87	0.64	0.44	Reference
4848	1846	Charles T. Jackson	Anesthesia	0	0.98	-	-	0.75	0.64	0.44	Reference
4874	1846	Christian Frederick Schonbein	Guncotton	0	0.97	-	-	0.69	0.64	0.44	Reference
5199	1847	Richard M. Hoe	Rotary Printing Press	0	0.99	-	-	0.80	0.64	0.42	Reference
5711	1848	M. Waldo Hanchett	Dental Chair	1	1.00	-	0.17	0.99	0.64	0.70	Reference
5942	1848	John Bradshaw	Sewing Machine	0	1.00	-	-	0.98	0.64	0.44	Reference
6099	1849	Morey/Johnson	Sewing Machine	1	1.00	-	0.17	0.99	0.64	0.69	Reference
6281	1849	Walter Hunt	Safety Pin	0	1.00	-	-	0.94	0.64	0.42	Reference
6439	1849	John Bachelder	Sewing Machine	0	1.00	-	-	0.97	0.64	0.42	Reference
7296	1850	D.M. Smith	Sewing Machine	0	1.00	-	-	1.00	0.64	0.40	Reference
7509	1850	J. Hollen	Sewing Machine	0	1.00	-	-	1.00	0.64	0.40	Reference
7931	1851	Grover and Baker	Sewing Machine	0	1.00	-	-	0.99	0.64	0.40	Reference
8080	1851	John Gorrie	Ice Machine	0	0.99	-	-	0.35	0.64	0.40	Reference
8294	1851	Isaac Singer	Sewing Machine	0	1.00	-	-	0.98	0.64	0.40	Reference
9300	1852	Lorenzo L. Langstroth	Beehive	1	1.00	-	0.17	0.85	0.64	0.69	Reference
13661	1855	Isaac M. Singer	Shuttle Sewing Machine	1	0.95	-	0.17	0.03	0.50	0.63	Reference
15553	1856	Gail Borden, Jr.	Condensed Milk	0	0.99	-	-	0.92	0.63	0.34	Reference
17628	1857	William Kelly	Iron and Steel Manufacturing	0	0.99	-	-	0.85	0.54	0.35	Reference
18653	1857	H.N. WadsworthÊ	Toothbrush	6	0.98	-	0.58	0.61	0.54	0.94	Reference
23536	1859	Martha Coston	System of Pyrotechnic Night Signals	1	0.97	-	0.17	0.60	0.64	0.58	Reference
26196	1859	James J. Mapes	Artificial Fertilizer	1	0.99	-	0.17	0.94	0.64	0.58	Reference
31128	1861	Elisha Graves Otis	Elevator	1	0.98	-	0.17	0.78	0.41	0.46	Reference
31278	1861	Linus Yale, Jr.	Lock	10	0.96	-	0.72	0.60	0.41	0.94	Reference
31310	1861	Samuel Goodale	Moving Picture Peep Show Machine	0	0.99	-	-	0.95	0.41	0.18	Reference
36836	1862	Richard J. Gatling	Machine Gun	3	0.95	0.22	0.38	0.21	0.83	0.82	Reference
43465	1864	Sarah Mather	Submarine Telescope	0	0.94	-	-	0.03	0.41	0.40	Reference
46454	1865	John Deere	Plow	0	0.99	-	-	0.60	0.43	0.41	Reference

Table A.1: Important Patents (cont)

Patent	Year	Inventor	Invention	Citations	Percentile Ranks						Source
					No Adjustment			Remove year FE			
					Quality	Citations		Quality	Citations		
				(total)	(0-10)	(0-10)	(total)	(0-10)	(0-10)	(total)	
53561	1866	Milton Bradleyy	Board Game	2	1.00	-	0.29	1.00	0.51	0.81	Reference
59915	1866	Pierre Lallement	Bicycle	0	0.99	-	-	0.86	0.51	0.41	Reference
78317	1868	Alfred Nobel	Dynamite	4	0.65	-	0.46	0.09	0.64	0.92	Reference
79265	1868	C. Latham Sholes	Typewriter	1	0.93	-	0.17	0.81	0.64	0.69	Reference
79965	1868	Alvin J. Fellows	Spring Tape Measure	2	0.82	-	0.29	0.36	0.64	0.82	Reference
88929	1869	George Westinghouse	Air Brake	1	0.84	-	0.17	0.79	0.64	0.69	Reference
91145	1869	Ives W. McGaffey	Vacuum Cleaner	4	0.74	0.22	0.46	0.58	0.85	0.92	Reference
110971	1871	Andrew Smith Hallidie	Cable Car	1	0.80	0.22	0.17	0.79	0.83	0.67	Reference
113448	1871	Mary Potts	Sad Iron	3	0.67	-	0.38	0.55	0.41	0.87	Reference
127360	1872	J.P. Cooley, S. Noble	Toothpick-making machine	0	0.75	-	-	0.68	0.39	0.39	Reference
129843	1872	Elijah McCoy	Improvements in Lubricators for Steam-Engines	1	0.73	-	0.17	0.64	0.39	0.66	Reference
135245	1873	Louis Pasteur	Pasteurization	0	0.54	-	-	0.26	0.37	0.38	Reference
141072	1873	Louis Pasteur	Manufacture of Beer and Treatment of Yeast	1	0.48	-	0.17	0.18	0.37	0.66	Reference
157124	1874	Joseph F. Glidden	Barbed Wire	1	0.94	-	0.17	0.95	0.38	0.65	Reference
161739	1875	Alexander Graham Bell	Telephone	7	0.99	-	0.62	0.99	0.38	0.96	Reference
171121	1875	George Green	Dental DrillÊ	2	0.97	0.22	0.29	0.98	0.83	0.79	Reference
174465	1876	Alexander Graham Bell	Telephone	6	1.00	0.37	0.58	1.00	0.92	0.95	Reference
178216	1876	Alexander Graham Bell	Telephone	0	1.00	-	-	1.00	0.39	0.38	Reference
178399	1876	Alexander Graham Bell	Telephone	2	0.99	0.22	0.29	0.99	0.83	0.79	Reference
186787	1877	Alexander Graham Bell	Electric Telegraphy	0	1.00	-	-	1.00	0.37	0.37	Reference
188292	1877	Chester Greenwood	Earmuffs	17	0.92	-	0.84	0.91	0.37	0.99	Reference
194047	1877	Nicolaus August Otto	Internal Combustion Engine	1	0.73	-	0.17	0.48	0.37	0.65	Reference
200521	1878	Thomas Alva Edison	Phonograph	12	0.91	0.37	0.77	0.84	0.91	0.98	Reference
201488	1878	Alexander Graham Bell	Telephone	2	1.00	-	0.29	1.00	0.36	0.78	Reference
203016	1878	Thomas Alva Edison	Speaking Telephone	15	1.00	0.37	0.82	1.00	0.91	0.99	Reference
206112	1878	Thaddeus Hyatt	Reinforced Concrete	0	0.79	-	-	0.47	0.36	0.36	Reference
220925	1879	Margaret Knight	Paper-Bag Machine	4	0.88	0.49	0.46	0.64	0.95	0.90	Reference
222390	1879	Thomas Alva Edison	Improvement in carbon telephones	16	1.00	-	0.83	1.00	0.36	0.99	Reference
223898	1880	Thomas Alva Edison	First Incandescent Light	20	0.99	-	0.87	0.99	0.41	0.99	Reference
224573	1880	Emile Berliner	Microphone	0	0.95	-	-	0.89	0.41	0.36	Reference
228507	1880	Alexander Graham Bell	Electric Telephone	3	1.00	0.37	0.38	1.00	0.92	0.85	Reference
237664	1881	Frederic E. Ives	Halftone Printing Plate	1	0.90	0.22	0.17	0.72	0.83	0.64	Reference
304272	1884	Ottmar Mergenthaler	Linotype	0	0.90	-	-	0.93	0.40	0.35	Reference

Table A.1: Important Patents (cont)

Patent	Year	Inventor	Invention	Citations	Percentile Ranks						Source
					No Adjustment			Remove year FE			
					Quality	Citations		Quality	Citations		
				(total)	(0-10)	(0-10)	(total)	(0-10)	(0-10)	(total)	
312085	1885	Edward J. Claghorn	Seat Belt	13	0.34	-	0.79	0.30	0.38	0.98	Reference
322177	1885	Sarah Goode	Folding Cabinet Bed	3	0.53	-	0.38	0.60	0.38	0.84	Reference
347140	1886	Elihu Thomson	Electric Welder	16	0.58	0.88	0.83	0.58	1.00	0.99	Reference
349983	1886	Gottlieb Daimler	Four Stroke Combustion Engine	4	0.98	-	0.46	0.99	0.39	0.89	Reference
371496	1887	Dorr E. Felt	Adding Machine	6	0.71	0.57	0.58	0.73	0.97	0.94	Reference
372786	1887	Emile Berliner	Phonograph Record	4	0.73	0.49	0.46	0.75	0.95	0.89	Reference
373064	1887	Carl Gassner, Jr.	Dry Cell Battery	3	0.28	-	0.38	0.11	0.38	0.84	Reference
382280	1888	Nikola Tesla	A. C. Induction Motor	2	0.80	0.22	0.29	0.89	0.83	0.76	Reference
386289	1888	Miriam Benjamin	Gong and Signal Chair for Hotels	0	0.50	-	-	0.53	0.41	0.34	Reference
388116	1888	William S. Burroughs	Calculator	3	0.76	-	0.38	0.85	0.41	0.84	Reference
388850	1888	George Eastman	Roll Film Camera	1	0.85	-	0.17	0.93	0.41	0.62	Reference
395782	1889	Herman Hollerith	Computer	1	0.54	0.22	0.17	0.66	0.83	0.61	Reference
400665	1889	Charles M. Hall	Aluminum Manufacture	2	0.85	0.22	0.29	0.94	0.83	0.76	Reference
415072	1889	Starley/Owen	Tandem Bicycle	1	0.63	-	0.17	0.77	0.42	0.61	Reference
430212	1890	Hiram Stevens Maxim	Smokeless Gunpowder	0	0.54	-	-	0.73	0.46	0.34	Reference
430804	1890	Herman Hollerith	Electric Adding Machine	2	0.80	0.22	0.29	0.93	0.84	0.76	Reference
447918	1891	Almon B. Strowger	Telephone Exchange	81	0.56	-	0.98	0.81	0.48	1.00	Reference
453550	1891	John Boyd Dunlop	Pneumatic Tyres	1	0.78	0.22	0.17	0.94	0.84	0.61	Reference
468226	1892	William Painter	Bottle Cap	7	0.73	-	0.62	0.94	0.34	0.94	Reference
472692	1892	G.C. Blickensderfer	Typewriting Machine	4	0.28	0.37	0.46	0.58	0.91	0.88	Reference
492767	1893	Edward G. Acheson	Carborundum	12	0.24	-	0.77	0.53	0.44	0.98	Reference
493426	1893	Thomas Alva Edison	Motion Picture	1	0.77	-	0.17	0.95	0.44	0.60	Reference
504038	1893	Whitcomb L. Judson	Zipper	6	0.24	-	0.58	0.53	0.44	0.93	Reference
536569	1895	Charles Jenkins	Phantoscope	0	0.87	-	-	0.96	0.34	0.31	Reference
549160	1895	George B. Selden	Automobile	0	0.69	-	-	0.87	0.34	0.31	Reference
558393	1896	John Harvey Kellogg	Cereal	3	0.60	-	0.38	0.67	0.49	0.83	Reference
558719	1896	C.B. Brooks	Street Sweeper	2	0.61	0.37	0.29	0.68	0.92	0.75	Reference
558936	1896	Joseph S. Duncan	Addressograph	3	0.33	0.22	0.38	0.25	0.84	0.83	Reference
586193	1897	Guglielmo Marconi	Radio	4	0.83	0.57	0.46	0.87	0.97	0.88	Reference
589168	1897	Thomas A. Edison	Motion Picture Camera	0	0.63	-	-	0.61	0.49	0.31	Reference
608845	1898	Rudolf Diesel	Diesel Engine	8	0.77	-	0.66	0.76	0.47	0.95	Reference
621195	1899	Ferdinand Graf Zepplin	Dirigible	1	0.72	-	0.17	0.52	0.35	0.57	Reference
644077	1900	Felix Hoffmann	Aspirin	1	0.71	-	0.17	0.41	0.46	0.58	Reference

Table A.1: Important Patents (cont)

Patent	Year	Inventor	Invention	Citations	Percentile Ranks						Source
					No Adjustment			Remove year FE			
					Quality	Citations		Quality	Citations		
				(total)	(0-10)	(0-10)	(total)	(0-10)	(0-10)	(total)	
661619	1900	Valdemar Poulsen	Magnetic Tape Recorder	15	0.84	0.69	0.82	0.74	0.98	0.98	Reference
708553	1902	John P. Holland	Submarine	1	0.75	-	0.17	0.54	0.45	0.57	Reference
743801	1903	ÊMary Anderson	Windscreen Wiper	2	0.35	-	0.29	0.07	0.51	0.73	Reference
745157	1903	Clyde J. Coleman	Electric Starter	1	0.91	-	0.17	0.91	0.51	0.57	Reference
764166	1904	Albert Gonzales	Railroad Switch	0	0.67	-	-	0.59	0.52	0.30	Reference
766768	1904	Michael J. Owens	Glass Bottle Manufacturing	7	0.76	0.64	0.62	0.74	0.98	0.94	Reference
775134	1904	KC Gillette	Razor (with removable blades)	4	0.92	0.49	0.46	0.95	0.95	0.87	Reference
808897	1906	Willis H. Carrier	Air Conditioning	21	0.66	0.22	0.88	0.72	0.84	0.99	Reference
815350	1906	John Holland	Submarine	0	0.71	-	-	0.78	0.54	0.28	Reference
821393	1906	Orville Wright	Airplane	19	1.00	0.22	0.86	1.00	0.84	0.99	Reference
841387	1907	Lee De Forest	Triode Vacuum Tube	5	0.29	0.22	0.52	0.23	0.85	0.90	Reference
921963	1909	Leonard H. Dyer	Automobile Vehicle	0	0.59	-	-	0.77	0.57	0.26	Reference
942809	1909	Leo H. Baekeland	Bakelite	3	0.89	0.22	0.38	0.97	0.84	0.80	Reference
970616	1910	Thomas A Edison	helicopter (never flown)	2	0.91	-	0.29	0.98	0.61	0.71	Reference
971501	1910	Fritz Haber	Ammonia Production	1	0.97	0.22	0.17	0.99	0.85	0.54	Reference
1000000	1911	Francis Holton	Non-Puncturable Vehicle Tire	2	0.83	-	0.29	0.93	0.60	0.71	Reference
1005186	1911	Henry Ford	Automotive Transmission	3	0.59	-	0.38	0.76	0.60	0.80	Reference
1008577	1911	Ernst F. W. Alexanderson	High Frequency Generator	6	0.50	0.69	0.58	0.65	0.99	0.92	Reference
1030178	1912	Peter Cooper Hewitt	Mercury Vapor Lamp	1	0.85	-	0.17	0.95	0.55	0.54	Reference
1082933	1913	William D. Coolidge	Tungsten Filament Light Bulb	28	0.73	0.57	0.92	0.90	0.97	0.99	Reference
1102653	1914	Robert H. Goddard	Rocket	58	0.42	0.49	0.97	0.62	0.95	1.00	Reference
1103503	1914	Robert Goddard	Rocket Apparatus	29	0.36	0.49	0.92	0.53	0.95	0.99	Reference
1113149	1914	Edwin H. Armstrong	Wireless Receiver	11	0.87	0.22	0.75	0.97	0.85	0.97	Reference
1115674	1914	Mary P. Jacob	Brassiere	1	0.53	-	0.17	0.76	0.59	0.53	Reference
1180159	1916	Irving Langmuir	Gas Filled Electric Lamp	13	0.78	0.64	0.79	0.94	0.98	0.97	Reference
1203495	1916	William D. Coolidge	X-Ray Tube	11	0.77	0.49	0.75	0.93	0.95	0.96	Reference
1211092	1917	William Coolidge	X-Ray Tube	7	0.94	0.22	0.62	0.99	0.84	0.92	Reference
1228388	1917	Frederick C Bargar	Fire Extinguisher	2	0.51	-	0.29	0.78	0.53	0.68	Reference
1254811	1918	Charles F. Kettering	Engine Ignition	1	0.50	-	0.17	0.78	0.60	0.51	Reference
1279471	1918	Elmer A. Sperry	Gyroscopic Compass	9	0.94	0.22	0.69	0.99	0.85	0.95	Reference
1360168	1920	Ernst Alexanderson	Antenna	4	0.92	-	0.46	0.98	0.62	0.83	Reference
1394450	1921	Charles P Strite	Bread Toaster	2	0.60	-	0.29	0.85	0.62	0.66	Reference
1413121	1922	John Arthur Johnson	Adjustable Wrench	0	0.05	-	-	0.06	0.63	0.20	Reference

Table A.1: Important Patents (cont)

Patent	Year	Inventor	Invention	Citations	Percentile Ranks						Source
					No Adjustment			Remove year FE			
					Quality	Citations		Quality	Citations		
				(total)	(0-10)	(0-10)	(total)	(0-10)	(0-10)	(total)	
1420609	1922	Glenn H. Curtiss	Hydroplane	2	0.68	-	0.29	0.89	0.63	0.65	Reference
1573846	1926	Thomas Midgley, Jr.	Ethyl Gasoline	3	0.36	0.22	0.38	0.78	0.84	0.72	Reference
1682366	1928	Charles F. Brannock	Foot Measuring Device	4	0.10	0.22	0.46	0.38	0.84	0.78	Reference
1699270	1929	John Logie Baird	Television / TV	11	0.55	-	0.75	0.91	0.48	0.94	Reference
1773079	1930	Clarence Birdseye	Frozen Food	10	0.53	0.22	0.72	0.92	0.84	0.93	Reference
1773080	1930	Clarence Birdseye	Frozen Food	18	0.60	-	0.86	0.94	0.45	0.97	Reference
1773980	1930	Philo T. Farnsworth	Television	29	0.83	0.91	0.92	0.98	1.00	0.99	Reference
1800156	1931	Erik Rotheim	Aerosol Spray Can	30	0.66	0.22	0.93	0.96	0.84	0.99	Reference
1821525	1931	Nielsen Emanuel	Hair Dryer	11	0.14	-	0.75	0.63	0.44	0.93	Reference
1835031	1931	Herman Affel	Coaxial cable	15	0.52	0.64	0.82	0.93	0.98	0.96	Reference
1848389	1932	Igor Sikorsky	Helicopter	5	0.41	-	0.52	0.91	0.42	0.78	Reference
1867377	1932	Otto F Rohwedder	Bread-Slicing Machine	2	0.09	-	0.29	0.57	0.42	0.52	Reference
1925554	1933	John Logie Baird	Color Television	1	0.38	-	0.17	0.90	0.37	0.33	Reference
1929453	1933	Waldo Semon	Rubber	56	0.83	0.96	0.97	0.99	1.00	1.00	Reference
1941066	1933	Edwin H. Armstrong	FM Radio	0	0.55	-	-	0.95	0.37	0.10	Reference
1948384	1934	Ernest O. Lawrence	Cyclotron	96	0.39	0.22	0.99	0.91	0.82	1.00	Reference
1949446	1934	William Burroughs	Adding and Listing Machine	1	0.09	0.22	0.17	0.62	0.82	0.31	Reference
1980972	1934	Lyndon Frederick	Krokodil	1	0.66	-	0.17	0.97	0.36	0.31	Reference
2021907	1935	Vladimir K. Zworykin	Television	18	0.56	0.37	0.86	0.95	0.91	0.95	Reference
2059884	1936	Leopold D. Mannes	Color Film	15	0.26	0.57	0.82	0.80	0.96	0.93	Reference
2071250	1937	Wallace H. Carothers	Nylon	231	0.79	0.97	1.00	0.98	1.00	1.00	Reference
2087683	1937	PT Farnsworth	Image Dissector	1	0.58	-	0.17	0.93	0.27	0.23	Reference
2153729	1939	Ernest H. Volwiler	Pentothal (General Anesthetic)	2	0.66	-	0.29	0.94	0.21	0.38	Reference
2188396	1940	Waldo Semon	Rubber	59	0.91	0.64	0.97	0.99	0.94	0.99	Reference
2206634	1940	Enrico Fermi	Radioactive Isotopes	99	0.78	0.99	0.99	0.97	1.00	1.00	Reference
2230654	1941	Roy J. Plunkett	TEFLON	49	0.48	0.93	0.96	0.91	0.99	0.99	Reference
2258841	1941	Jozsef Bir— Laszlo	Fountain Pen	20	0.05	0.84	0.87	0.31	0.98	0.94	Reference
2292387	1942	Markey/Antheil	Secret Communication System	71	0.33	0.37	0.98	0.86	0.74	0.99	Reference
2297691	1942	Chester F. Carlson	Xerography	738	0.11	0.91	1.00	0.58	0.99	1.00	Reference
2329074	1943	Paul Muller	DDT - Insecticide	48	0.15	0.97	0.96	0.68	1.00	0.98	Reference
2390636	1945	Ladislo Biro	Ball Point Pen	27	0.31	0.94	0.92	0.71	0.99	0.95	Reference
2404334	1946	Frank Whittle	Jet Engine	35	0.17	0.94	0.94	0.31	0.99	0.97	Reference
2436265	1948	Allen Du Mont	Cathode Ray Tube	18	0.54	0.88	0.86	0.64	0.98	0.91	Reference

Table A.1: Important Patents (cont)

Patent	Year	Inventor	Invention	Citations	Percentile Ranks						Source
					No Adjustment			Remove year FE			
					Quality	Citations		Quality	Citations		
				(total)	(0-10)	(0-10)	(total)	(0-10)	(0-10)	(total)	
2451804	1948	Donald L. Campbell	Fluid Catalytic Cracking	9	0.63	0.77	0.69	0.76	0.94	0.77	Reference
2495429	1950	Percy Spencer	Microwave	15	0.25	0.80	0.82	0.20	0.96	0.89	Reference
2524035	1950	John Bardeen	Transistor	132	0.79	1.00	0.99	0.90	1.00	1.00	Reference
2543181	1951	Edwin H. Land	Instant Photography	116	0.61	0.99	0.99	0.76	1.00	1.00	Reference
2569347	1951	William Shockley	Junction Transistor	140	0.64	1.00	0.99	0.79	1.00	1.00	Reference
2642679	1953	Frank ZamboniÊ	Resurfacing Machine	16	0.41	0.57	0.83	0.53	0.85	0.89	Reference
2668661	1954	George R. Stibitz	Modern Digital Computer	14	0.96	0.77	0.80	0.99	0.94	0.86	Reference
2682050	1954	Andrew Alford	Radio Navigation System	3	0.68	-	0.38	0.81	0.09	0.39	Reference
2682235	1954	Richard Buckminster Fuller	Geodesic Dome	86	0.57	0.96	0.99	0.69	1.00	0.99	Reference
2691028	1954	Frank B. Colton	First Oral Contraceptive	4	0.90	-	0.46	0.97	0.09	0.48	Reference
2699054	1955	Lloyd H. Conover	Tetracycline	38	0.91	0.95	0.95	0.96	0.99	0.97	Reference
2708656	1955	Enrico Fermi	Atomic Reactor	196	0.96	1.00	1.00	0.99	1.00	1.00	Reference
2708722	1955	An Wang	Magnetic Core Memory	76	0.82	0.99	0.98	0.90	1.00	0.99	Reference
2717437	1955	George De Mestral	Velcro	258	0.39	0.98	1.00	0.35	1.00	1.00	Reference
2724711	1955	Gertrude Elion	Leukemia-fighting drug 6-mercaptopurine	1	0.78	0.22	0.17	0.88	0.26	0.13	Reference
2752339	1956	Percy L. Julian	Preparation of Cortisone	11	0.87	0.84	0.75	0.92	0.97	0.81	Reference
2756226	1956	Ernst Brandl, Hans Margreiter	Oral Penicillin	7	0.71	0.69	0.62	0.76	0.91	0.67	Reference
2797183	1957	Hazen/ Brown	Nystatin	13	0.90	0.57	0.79	0.95	0.84	0.85	Reference
2816721	1957	R. J. Taylor	Rocket Engine	25	0.71	0.90	0.91	0.74	0.98	0.95	Reference
2817025	1957	Robert Adler	TV remote control	27	0.74	0.94	0.92	0.77	0.99	0.95	Reference
2835548	1958	Robert C. Baumann	Satellite	16	0.79	0.87	0.83	0.83	0.98	0.89	Reference
2866012	1958	Charles P. Ginsburg	Video Tape Recorder	30	0.80	0.93	0.93	0.85	0.99	0.96	Reference
2879439	1959	Charles H. Townes	Maser	24	0.73	0.91	0.90	0.77	0.99	0.94	Reference
2929922	1960	Arthur L. Shawlow	Laser	122	0.82	1.00	0.99	0.87	1.00	1.00	Reference
2937186	1960	Burckhalter/Seiwald	Antibody Labelling Agent	8	0.82	0.22	0.66	0.88	0.28	0.72	Reference
2947611	1960	Francis P. Bundy	Diamond Synthesis	62	0.71	0.37	0.98	0.75	0.70	0.99	Reference
2956114	1960	Charles P. Ginsburg	Wideband Magnetic Tape System	11	0.73	0.77	0.75	0.78	0.94	0.81	Reference
2981877	1961	Robert N. Noyce	Semiconductor Device	152	0.95	1.00	1.00	0.98	1.00	1.00	Reference
3057356	1962	Greatbatch Wilson	Pacemaker	127	0.86	0.91	0.99	0.92	0.99	1.00	Reference
3093346	1963	Maxime A. Faget	First Manned Space Capsule-Mercury	19	0.86	0.85	0.86	0.92	0.97	0.91	Reference
3097366	1963	Paul Winchell	Artificial Heart	23	0.45	0.69	0.89	0.36	0.91	0.93	Reference
3118022	1964	Gerhard M. Sessler	Electret Microphone	39	0.69	0.82	0.95	0.73	0.96	0.97	Reference
3156523	1964	Glenn T. Seaborg	Americium (Element 95)	1	0.84	-	0.17	0.90	0.13	0.13	Reference

Table A.1: Important Patents (cont)

Patent	Year	Inventor	Invention	Citations	Percentile Ranks						Source
					No Adjustment			Remove year FE			
					Quality	Citations		Quality	Citations		
				(total)	(0-10)	(0-10)	(total)	(0-10)	(0-10)	(total)	
3174267	1965	Edward C Bopf, Deere & Co	Cotton Harvester	4	0.43	0.57	0.46	0.32	0.84	0.47	Reference
3220816	1965	Alastair Pilkington	Manufacture of Flat Glass	25	0.77	0.37	0.91	0.83	0.69	0.94	Reference
3287323	1966	Stephanie Kwolek, Paul Morgan	Kevlar	1	0.63	-	0.17	0.69	0.08	0.12	Reference
3478216	1969	George Carruthers	Far-Ultraviolet Camera	3	0.63	0.37	0.38	0.82	0.70	0.39	Reference
3574791	1971	Patsy Sherman	Scotchguard	81	0.54	0.84	0.98	0.79	0.97	0.99	Reference
3663762	1972	Edward Joel Amos Jr	Cellular Telephone	112	0.59	0.93	0.99	0.84	0.99	1.00	Reference
3789832	1974	Raymond V. Damadian	MRI	59	0.44	0.89	0.97	0.81	0.98	0.98	Reference
3858232	1974	William Boyle	Digital EyeÊ	51	0.38	0.97	0.97	0.76	1.00	0.98	Reference
3906166	1975	Martin Cooper	Cellular Telephone	219	0.39	0.93	1.00	0.78	0.99	1.00	Reference
4136359	1979	Stephen Wozniak, Apple	Microcomputer for use with video display	37	0.77	0.69	0.95	0.95	0.87	0.94	Reference
4229761	1980	Valerie Thomas	Illusion Transmitter	3	0.84	-	0.38	0.97	0.05	0.21	Reference
4237224	1980	Boyer/Cohen	Molecular chimeras	301	1.00	1.00	1.00	1.00	1.00	1.00	Reference
4363877	1982	Howard M. Goodman	Human Growth Hormone	51	1.00	0.85	0.97	1.00	0.95	0.96	Reference
4371752	1983	Gordon Matthews	Digital Voice Mail System	223	0.82	0.98	1.00	0.92	1.00	1.00	Reference
4399216	1983	Richard Axel	Co-transformation	482	0.99	0.98	1.00	1.00	1.00	1.00	Reference
4437122	1984	Walsh/Halpert	bitmap (raster) graphics	178	1.00	0.96	1.00	1.00	0.99	1.00	Reference
4464652	1984	Apple	Lisa Mouse	112	0.85	0.98	0.99	0.92	1.00	0.99	Reference
4468464	1984	Boyer/Cohen	Molecular chimeras	109	1.00	0.91	0.99	1.00	0.97	0.99	Reference
4590598	1986	Gordon Gould	Laser	20	0.76	0.49	0.87	0.62	0.33	0.80	Reference
4634665	1987	Richard Axel	Co-transformation	183	1.00	0.82	1.00	1.00	0.88	1.00	Reference
4683195	1987	Kary B. Mullis	polymerase chain reaction	2884	1.00	1.00	1.00	1.00	1.00	1.00	Reference
4683202	1987	(several)	polymerase chain reaction	3328	0.99	1.00	1.00	0.99	1.00	1.00	Reference
4736866	1988	Leder/Stewart	transgenic (genetically modified) animals	370	1.00	0.99	1.00	1.00	1.00	1.00	Reference
4744360	1988	Patricia Bath	Cataract Laserphaco Probe	81	0.94	0.97	0.98	0.89	0.99	0.98	Reference
4816397	1989	Michael A. Boss	recombinant antibodies	567	0.99	0.97	1.00	0.99	0.99	1.00	Reference
4816567	1989	Shmuel Cabilly	immunoglobulins	1785	0.99	0.98	1.00	0.99	0.99	1.00	Reference
4838644	1989	Ellen Ochoa	Recognizing Method	22	0.96	0.88	0.89	0.94	0.92	0.81	Reference
4889818	1989	(several)	polymerase chain reaction	366	1.00	1.00	1.00	1.00	1.00	1.00	Reference
4965188	1990	(several)	polymerase chain reaction	1176	1.00	1.00	1.00	1.00	1.00	1.00	Reference
5061620	1991	(several)	Method for isolating the human stem cell	252	0.99	1.00	1.00	0.99	1.00	1.00	Reference
5071161	1991	Geoffrey L Mahoon	Airbag	23	0.87	0.94	0.89	0.68	0.96	0.81	Reference
5108388	1992	Stephen L. Troke	Laser Surgery Method	125	0.97	0.94	0.99	0.95	0.95	0.99	Reference
5149636	1992	Richard Axel	Co-transformation	6	1.00	0.49	0.58	1.00	0.22	0.36	Reference

Table A.1: Important Patents (cont)

Patent	Year	Inventor	Invention	Citations	Percentile Ranks						Source
					No Adjustment			Remove year FE			
					Quality	Citations		Quality	Citations		
				(total)	(0-10)	(0-10)	(total)	(0-10)	(0-10)	(total)	
5179017	1993	Richard Axel	Co-transformation	131	1.00	0.95	0.99	1.00	0.96	0.99	Reference
5184830	1993	Saturo Okada, Shin Kojo	Compact Hand-Held Video Game System	201	0.98	0.99	1.00	0.97	1.00	1.00	Reference
5194299	1993	Arthur Fry	Post-It Note	76	0.89	0.80	0.98	0.69	0.78	0.97	Reference
5225539	1993	Gregory P. Winter	Chimeric, humanized antibodies	671	1.00	1.00	1.00	1.00	1.00	1.00	Reference
5272628	1993	Michael Koss	Core Excel Function	94	1.00	0.98	0.99	0.99	0.99	0.98	Reference
5747282	1998	Mark H. Skolnick	isolating BRCA1 gene	15	0.95	0.69	0.82	0.91	0.32	0.67	Reference
5770429	1998	Nils Lonberg	human antibodies from transgenic mice	248	0.83	0.99	1.00	0.46	0.99	1.00	Reference
5837492	1998	(several)	isolating BRCA2 gene	5	0.82	0.37	0.52	0.44	0.08	0.26	Reference
5939598	1999	(several)	Transgenic mice	262	1.00	0.94	1.00	1.00	0.93	1.00	Reference
5960411	1999	Peri Hartman, Jeff Bezos	1-click buying	1387	1.00	1.00	1.00	1.00	1.00	1.00	Reference
6230409	2001	Patricia Billings	Geobond	7	0.77	0.69	0.62	0.74	0.36	0.46	Reference
6285999	2001	Larry Page	Google Pagerank	689	0.98	1.00	1.00	0.99	1.00	1.00	Reference
6331415	2001	Shmuel Cabilly	Antibody molecules	243	1.00	0.49	1.00	1.00	0.18	1.00	Reference
6455275	2002	Richard Axel	Co-transformation	7	0.93	0.49	0.62	0.97	0.19	0.52	Reference

Table A.2: Validation: Patent Importance and Forward Citations

Forward Citations	A. Contemporaneous Relation						B. Predictive Relation					
Measurement–Prediction	$(0,1) \rightarrow (0,1)$		$(0,5) \rightarrow (0,5)$		$(0,10) \rightarrow (0,10)$		$(0,1) \rightarrow 2+$		$(0,5) \rightarrow 6+$		$(0,10) \rightarrow 11+$	
Horizon	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
log(Patent Importance)	0.275***	0.139	1.008***	0.789***	1.058***	0.894***	1.029***	0.997***	0.692***	0.769***	0.344***	0.391***
	(0.062)	(0.074)	(0.037)	(0.053)	(0.015)	(0.025)	(0.059)	(0.064)	(0.053)	(0.075)	(0.027)	(0.045)
log(1 + Fwd. Citations)							0.615***	0.512***	0.588***	0.550***	0.546***	0.517***
							(0.017)	(0.015)	(0.016)	(0.015)	(0.015)	(0.014)
R^2	0.092	0.225	0.232	0.367	0.295	0.425	0.354	0.508	0.362	0.497	0.347	0.472
Observations	6,017,673	4,084,292	4,802,836	3,064,631	4,135,358	2,533,724	6,017,673	4,084,292	4,964,003	3,195,838	4,135,358	2,533,724
Grant Year FE	Y		Y		Y		Y		Y		Y	
Technology Class (CPC3)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Assignee $\times$ Year FE		Y		Y		Y		Y		Y		Y

Table reports the results of estimating the following specification at the patent level (indexed by j):

$$\log\left(1 + CITES_j\right) = \alpha + \beta \,\log q_j^{\tau} + \gamma\,\mathbf{Z}_j + \varepsilon_j.$$

In terms of the independent variables, we measure patent importance and citations over the τ years since the patent is filed. For the dependent variable, we measure forward citations over the same interval (Panel A) or from year $\tau + 1$ onwards (Panel B). The vector $\mathbf{Z}_j$ includes dummies controlling for technology class (defined at the 3-digit CPC level), grant year, and the interaction of assignee and year effects. Including assignee fixed effects reduces the number of observations since many patents have no assignees. We restrict attention to the sample of patents issued after 1947, as this is the period for which citations are recorded consistently by the USPTO. We cluster the standard errors by the patent grant year and report them in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Validation: Patent Importance and Market Values

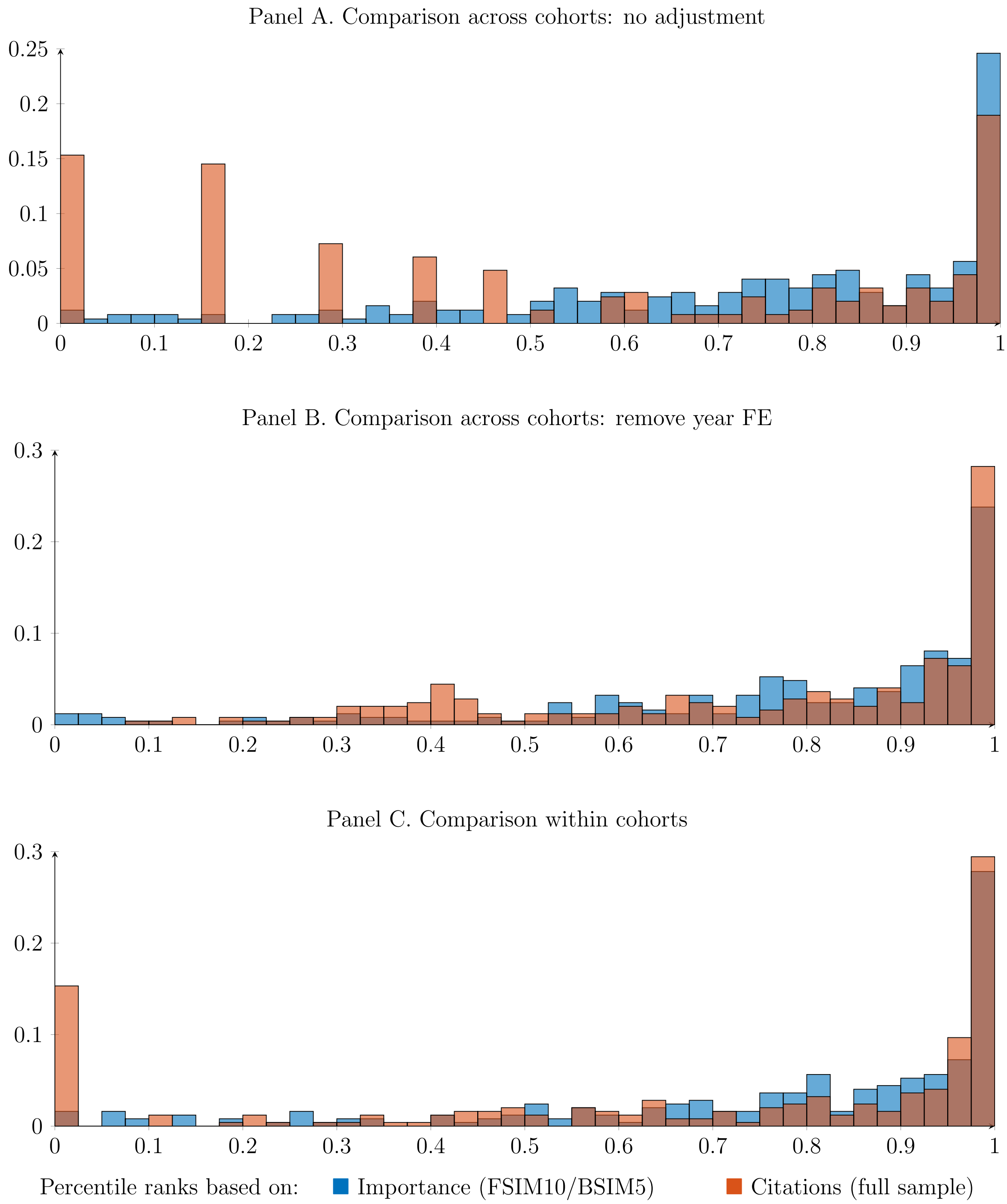
Horizon of Forward Similarity/Citations	(0-1yrs) (1)	(0-1yrs) (2)	(0-5yrs) (3)	(0-5yrs) (4)	(0-10yrs) (5)	(0-10yrs) (6)
log(Patent Importance)	0.0019**	0.0020**	0.0027**	0.0024*	0.0039**	0.0029*
	(0.0009)	(0.0010)	(0.0012)	(0.0012)	(0.0015)	(0.0016)
log(1 + Forward Citations)		-0.0003		0.0016**		0.0038***
		(0.0004)		(0.0006)		(0.0010)
Observations	2,097,985	2,097,985	1,737,732	1,737,732	1,424,928	1,424,928
R^2	0.949	0.949	0.947	0.947	0.939	0.939

Table reports the results of estimating the following specification

$$\log \hat{V}_j = \alpha + \beta \, \log q_j^{\tau} + \gamma \, \mathbf{Z}_j + \varepsilon_j.$$

The regression relates the log of the Kogan et al. (2017) estimate of the market value of the patent to our (log) measures of patent importance, which combines the patent's impact and novelty, constructed in equation (10) in the paper. As controls $\mathbf{Z}_j$, we include dummies controlling for technology class (defined at the 3-digit CPC level), the logarithm of the firm's market capitalization and the interaction of firm (CRSP: permco) and grant year effects. In columns (3), (5), and (5) we include as additional controls the number of forward citations (measured over the same horizon as our importance measure). We cluster the standard errors by the patent grant year and report them in parentheses. Independent variables are normalized to unit standard deviation. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure A.1: Significant Patents: Importance vs Forward Citations



Distribution of patent percentile ranks based on our patent importance indicator (blue) measured over a horizon of 10 years and forward citations (light red) measured over the entire sample . A value of x% indicates that a given patent scores higher than x% of all other patents in the sample (panel A); same after removing year-fixed effects from importance and citations (Panel B); or computing percentile ranks relative to patents that are issued in the same year (panel C). The list of patents, along with their source, appears in Appendix Table A.1

Figure A.2: Breakdown of Innovation by Technology Classes

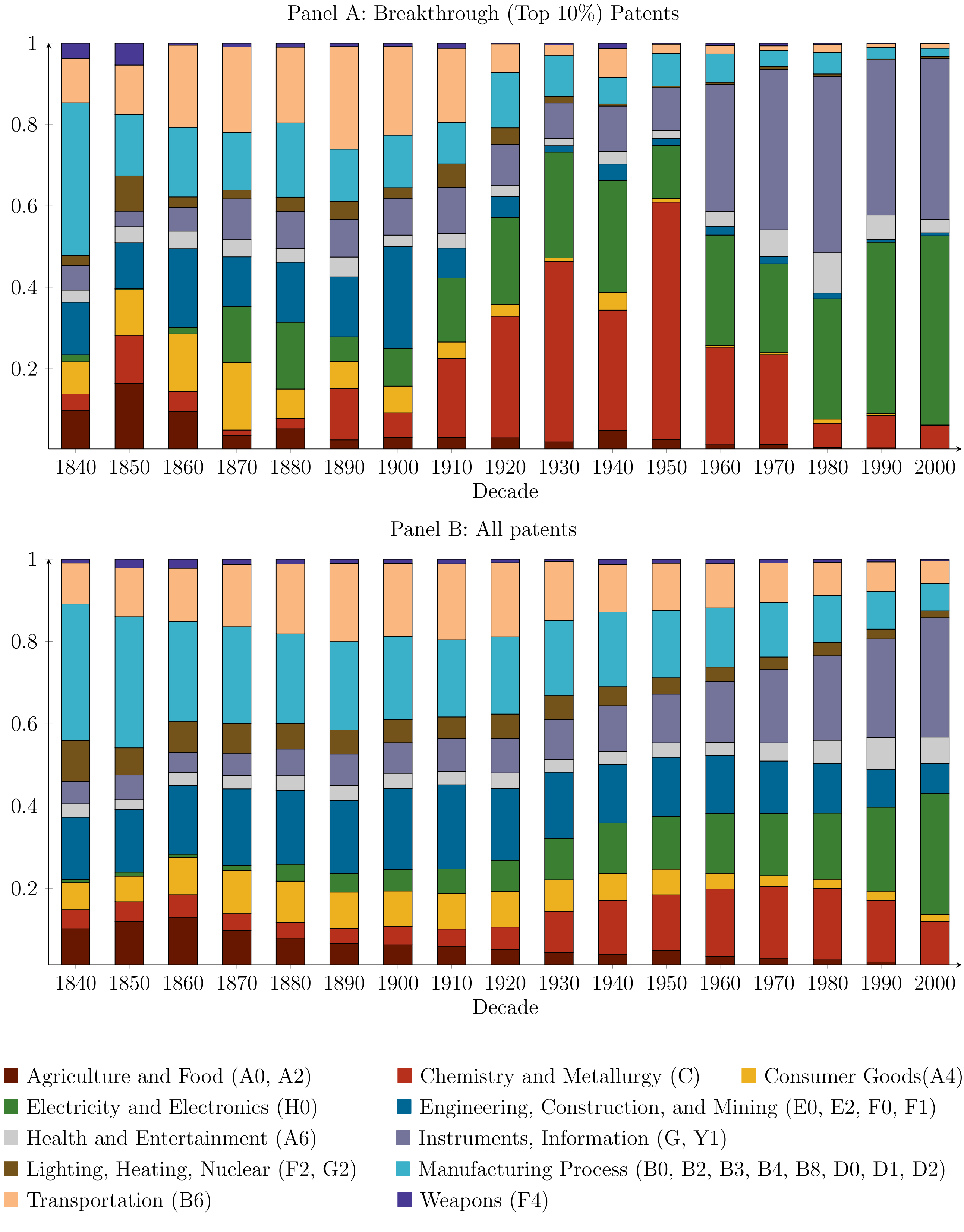
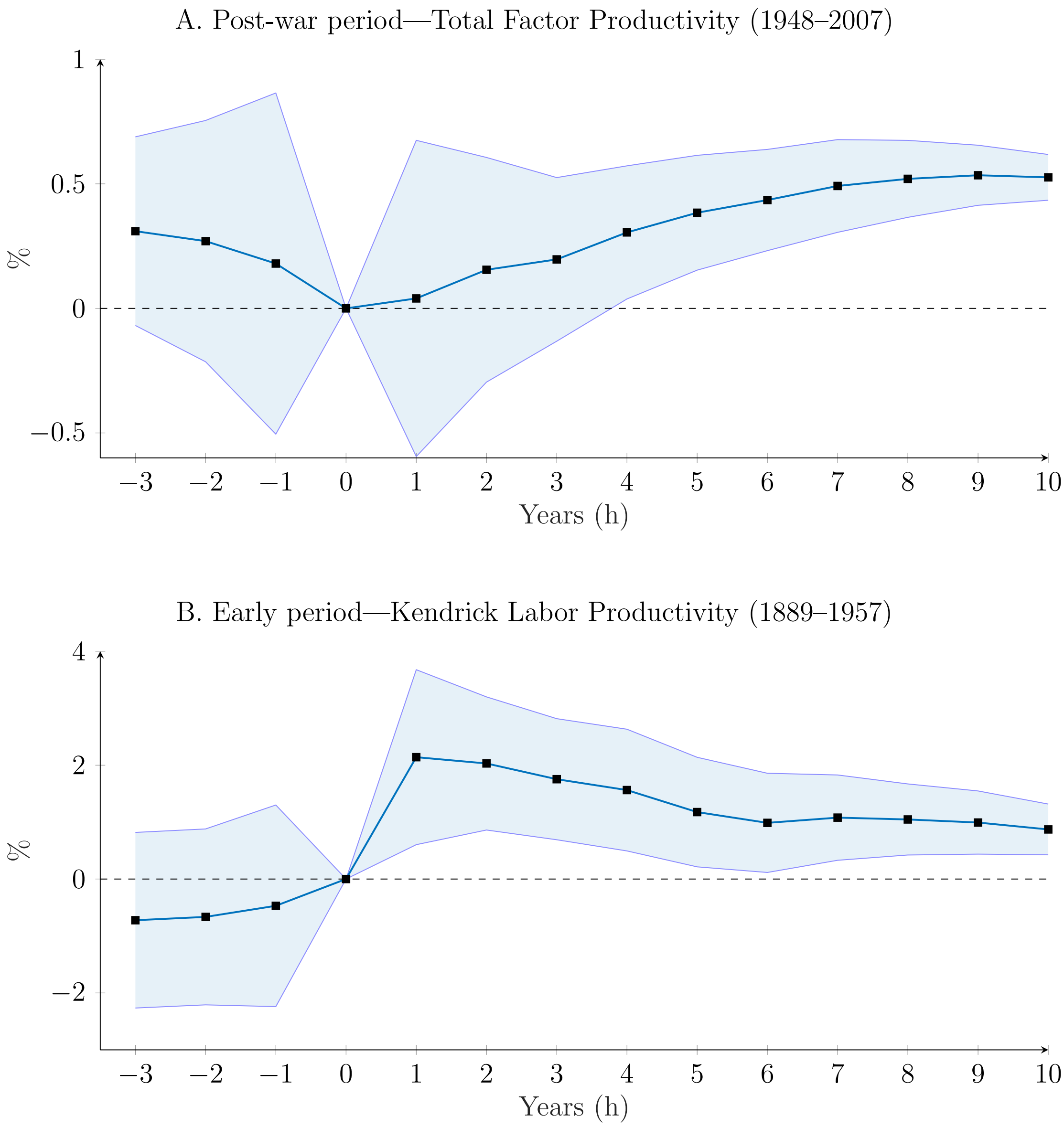
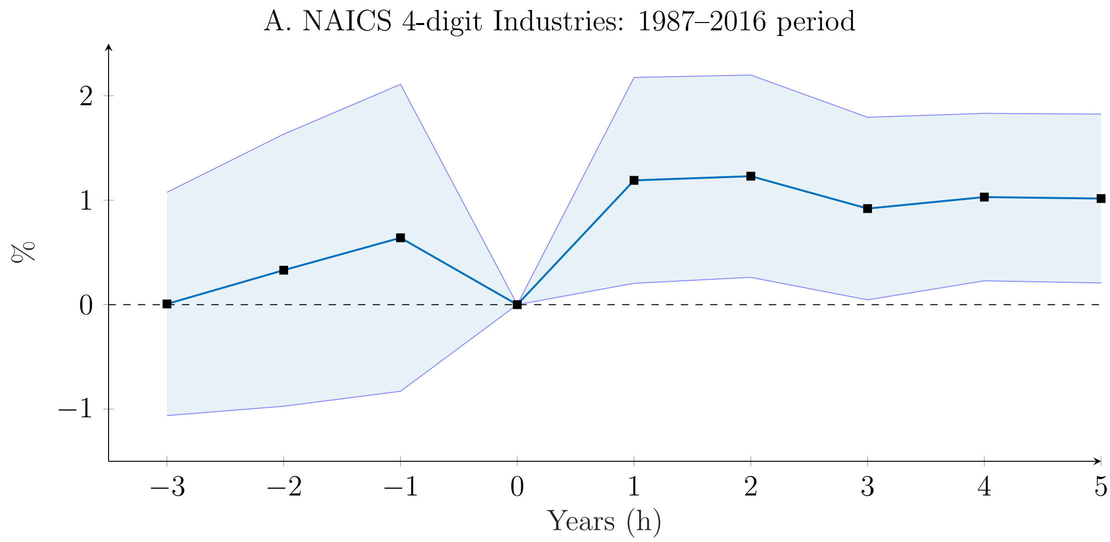


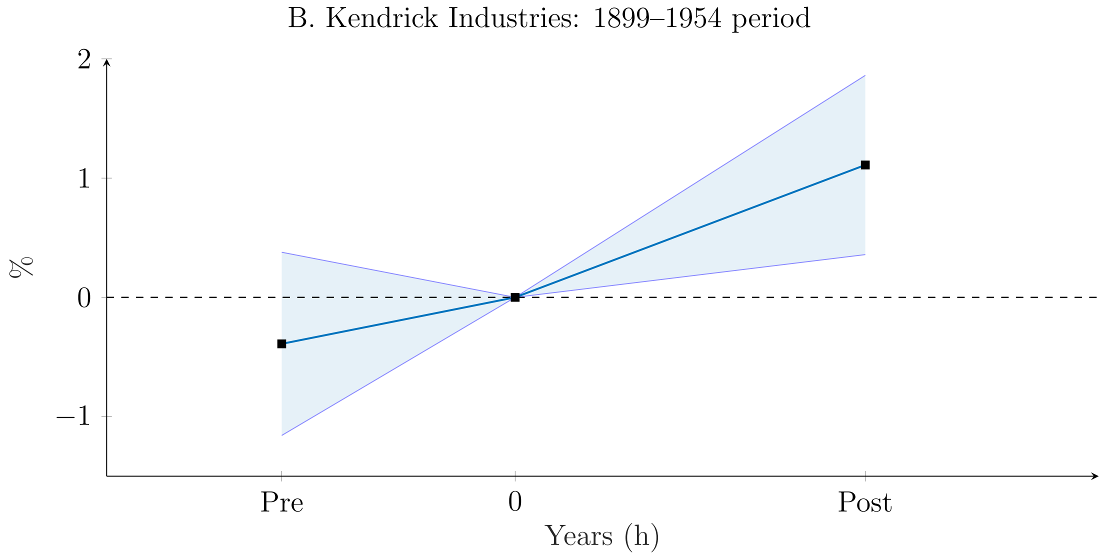
Figure A.3: Breakthrough Innovation and Aggregate TFP



Response of measured productivity to a unit standard deviation shock to our technological innovation index (in logs). In Panel A, productivity is measured using total factor productivity from Basu et al. (2006). In Panel B, productivity is measured by output per manhour in manufacturing (Kendrick, Table D-II). We include 90% confidence intervals, computed using Newey-West standard errors (with a maximum number of lags equal to one plus the number of overlapping observations). All specifications control for the lag level of productivity.

Figure A.4: Breakthrough Innovation and Industry TFP

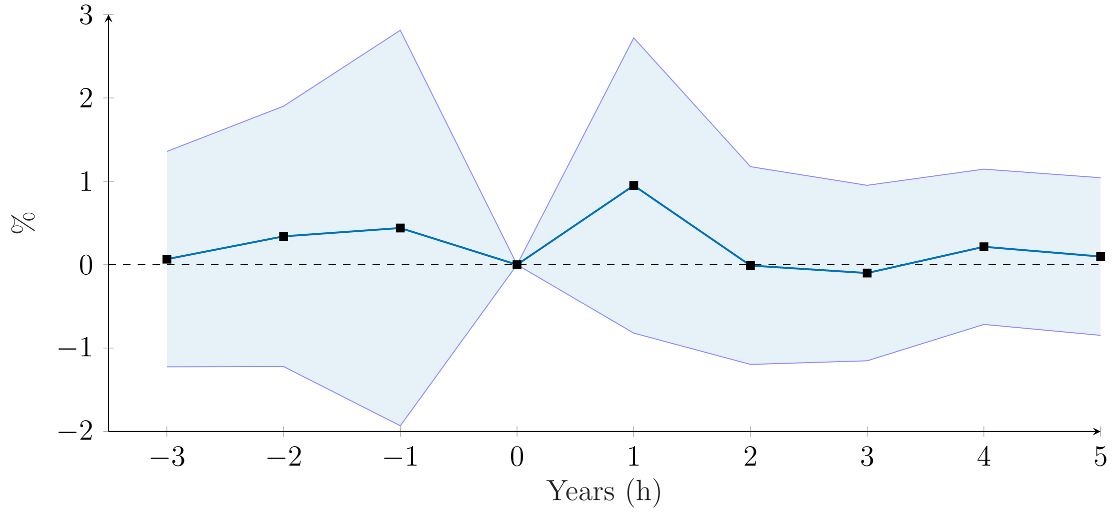




Response of industry total factor productivity to a unit standard deviation shock to our technological innovation index. Panel A presents results for 86 manufacturing industries at the NAICS 4-digit level. Productivity data is from the Bureau of Labor Statistics. Kendrick industries are from Table D-V, and productivity is output per manhour. The Kendrick data includes information for the level of labor productivity (output per manhour) for 62 manufacturing industries for the years 1899, 1909, 1919, 1937, 1947, and 1954. For each period (t, s), we regress the annualized difference in log labor productivity on the log of the accumulated level of innovation (number of breakthrough patents) in $t \pm 2$ years—controlling for time and industry dummies, the log number of patents during the same period, and the log level of productivity at t. Standard errors are clustered by industry. To construct industry innovation indices for NAICS industries, we use the probabilistic mapping from CPC codes to NAICS codes from Goldschlag et al. (2016). To construct innovation indices for the Kendrick industries, which are defined at the SIC code level, we use the concordance between 1997 NAICS and 1987 SIC codes from the Census Bureau. If NAICS industries map into multiple SIC codes, we assign an equal fraction to each.

Figure A.5: Breakthrough patents and Industry TFP—comparison to Citations

A. NAICS 4-digit Industries: 1987–2016 period



B. Kendrick Industries: 1899–1954 period

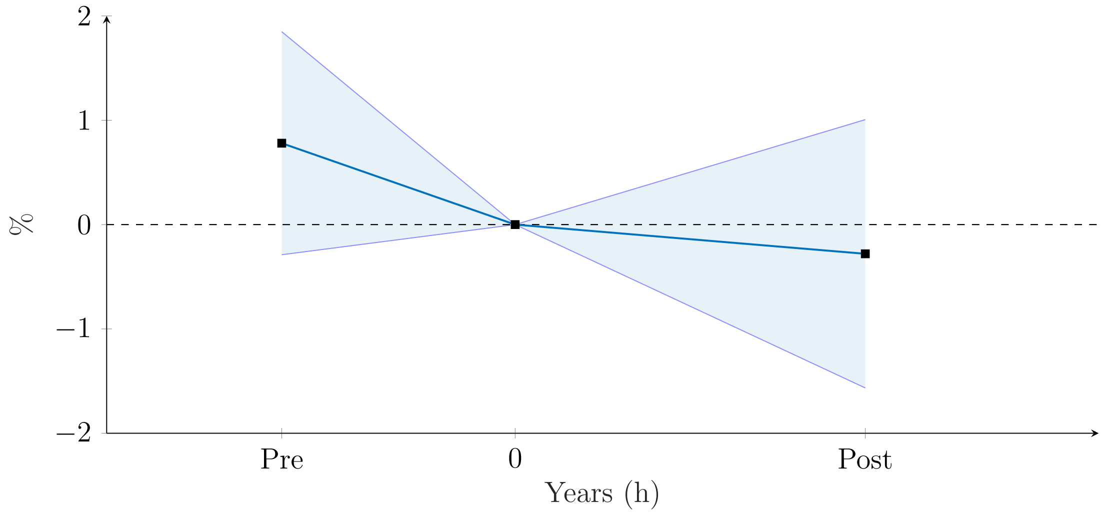


Figure performs the same exercise as Figure A.4, except that we now construct the industry innovation indices based on citation counts.