

Online Appendix

**America, Jump-started: World War II R&D
and the Takeoff of the U.S. Innovation System**

Daniel P. Gross and Bhaven N. Sampat

A OSRD History and Data

A.1 Data collected from OSRD archival records

Data on OSRD contractors, contracts, inventions, and patents were obtained from archival records at the U.S. National Archives and Records Administration (NARA).¹ We have made a public-use version available at <https://doi.org/10.7910/DVN/POO1ZH>. NARA’s OSRD collection also includes narrative records (such as correspondence between Vannevar Bush and other OSRD administrators, contractors, and government agencies) that provide rich background on OSRD’s day-to-day operation and precedent-setting policy choices throughout the war.

For this paper, we make use of several sets of records from OSRD collection. The three key data sources on contracts, contractors, and inventions are (i) a collection of contract index cards, which identify each contract with the contract ID, contractor, OSRD division managing the contract, termination date, and obligated value through the end of life; (ii) a directory of contractors, by contract, with the contractor’s principal headquarters location; and (iii) a set of invention disclosure cards, which document inventions generated in the course of OSRD-supported research, with the associated contract, applicable patent clause (short form vs. long form), contractor, inventor(s), invention title, date reported, and if a patent application was filed, the application date and serial number. Examples of each are provided in Figures A.1 to A.3.

Figure A.1: Example Contract Index Card for OSRD Contract OEMsr-441

APPROPRIATION	OBL. NO.	AMOUNT OBLIGATED	SUPP. NO.	OBL. LIC
1120500	0-4761	\$ 30,000.00		✓
1120006(02).003	0-6310	200,000.00	1	-0-
1130500.031	JSR-1980	360,000.00	3	-0-
Time extension			2	
1143600.001	JSR-1184-44	150,000.00	4	-0-
1153600.001	5165	740,000.00	10	
1153600.001	5792	76,000.00		
	Time extension	95,000.00		
	Time extension			
	Time extension			
	Time extension			
Supersedes SR-171, 513, 514, 515, 4615 (210,000)				
Total 2/20/47				
CONTRACT NO.	CONTRACTOR	DIV.	TERM DATE	
sr-441	RCA	5.3	6/30/46	

¹See NARA Record Group 227, “Records of the Office of Scientific Research and Development”. The key record sets for this paper are the *Index to Contracts* (NC-138, Entry 27, Stack area 130, Row 20, Compartment 11, Shelf 1, Boxes 1-5), *Contractor Lists* (NC-138, Entry 16, Stack area 130, Row 20, Compartment 11, Shelf 4, Boxes 1-2), and *Register of Invention Disclosures* (NC-138, Entry 35, Stack area 130, Row 20, Compartment 38-39, Shelf 7/1, Boxes 1-8), though other records were used to complete and cross-validate these data.

Figure A.2: Example Contractor Directory Page for OEMsr-441

RADIO CORPORATION OF AMERICA, RCA VICTOR DIVISION **Division 5**

Contract No. **OEMsr-441**

Service Projects:

AC-36
AC-1
NA-116
NA-132
NA-136
NA-151
NA-190
NO-40
NO-115
NS-132
NS-136

Classification (a) Highest - Confidential
(b) Present - Confidential

Business Representative:

Mr. Meade Brunet, Vice President
Radio Corporation of America
RCA Victor Division
Camden, New Jersey

Figure A.3: Example Invention Disclosure Card for OEMsr-441

OSRD 5397

Contract No. **OEMsr-441** Long 5/3-sr441 Pat 43 (Project 415)

Contractor **Radio Corporation of America**

Inventor W. J. Poch

Title Deflection Oscillators

Received From John G. Roberts

Date 25 October 1945

To War 30 October 1945

To Navy 30 October 1945

To Inventor

Serial No. **607,111**

Filed 26 July 1945

Assignee Radio Corporation of America

License 17 October 1945 Filed - Patent Office File 6462 - 11/14/45

Received License Agreement

Recorded: Liber

Page

11/2/45 - Form #24 to ComPat
11/2/45 - War & Navy informed of LA
11/2/45 - LA returned to PA

OSRD Form P-1
7-39-44

Disclosure of Invention

D-1598

In addition to these records, we also collected three supplementary, independent lists of contractors and contracts which can be used to validate these data and fill in gaps. The first is a list of contractors, which lists contracts for each contractor provides the total obligations and counts the number of contracts with the short form vs. long form patent clause. The second is a list of contracts by contractor, and provides the associated OSRD division and patent clause. The third is a list of contracts by OSRD division. Using the collective set of records, we compile a master list of 2,288 OSRD contracts, 573 of which were CMR contracts.² For comparison, internal correspondence

²This list of contracts comprises the union of all record sets. Although the vast majority of contracts appear across all sources, there are gaps (missing contracts) in each data source: the contract index cards contain 2,275 of the

from OSRD’s patent division dated October 30, 1945 counts 2,266 research contracts, with the remaining 22 contracts appearing in our data as administrative contracts for the purchase/lease of office space, equipment rentals, and publication services. As another point of comparison, the administrative history of the CMR published in 1948 (Andrus 1948) lists 571 of the 573 CMR contracts in our data (of the remaining two, one was last contract OSRD entered into and likely simply did not make it into this published list).³

Given that for a few contract-level variables we have multiple sources (namely, OSRD division and patent clause), we make an additional effort to cross-validate these variables and reconcile differences as we describe below. We also harmonize contractor names and match them to assignees in the patent record (see discussion of patent datasets in the next subsection).

- The OSRD division of each contract indicates the subject matter and can be reported in up to four sources: the contract index cards, the contractor directory, and two supplemental contract lists. For 2,117 out of 2,288 contracts in the master list, at least one source reports an OSRD division and all reporting sources agree. In 152 cases, they disagree, and for these cases we take the most commonly-reported division as the true division (in the nine cases where there is no single most commonly-reported division, we prioritize the division reported in the contract index cards and contractor directory). In the 19 remaining cases, the contract’s OSRD division is not reported in any of our sources and is unknown.
- The patent clause of each contract is typically either “short form” or “long form”, which gives the government or the contractor (respectively) the first rights to patent inventions developed under the contract. The short form clause was used predominantly with academic contractors, and the long form clause with industrial contractors (as a means to incentivize their participation in the war effort). The patent clause of each contract is reported in one of the supplemental contract lists as well as on invention disclosure cards – but only known for 2,020 contracts in the supplemental contract list, and for an additional five contracts via the invention disclosures. Determining the applicable patent clause is further complicated by the fact that some contracts (i) had their patent clauses changed mid-contract, and/or (ii) had custom patent clauses (which were usually variants on the standard short form and long form clauses).⁴ Custom patent clauses and patent clause amendments, as well as the

2,288 total known contracts; the contractor directory contains 2,259 contracts (the 573 CMR contracts were not included in this directory per se, but the contractor information was available in separately-maintained CMR contract ledgers); the supplementary contractor list contains 2,192 contracts; and the supplementary contract list contains 2,112 contracts. These latter two lists were compiled before the end of the war, and most of contracts missing from these lists are missing because they post-date the sample. For the 13 contracts missing from the contract index cards, the contractor and OSRD division are available from other sources, but the contract value and termination date are unknown. For the 29 contracts missing from the contractor directory, we infer contractor locations from other contracts with the same contractor where available, and otherwise through manual research.

³As a final piece of validation: OSRD contract IDs were written with three different prefixes (NDCrc, OEMsr, OEMcmr) followed by an identification number. The maximum number of each series in the data is NDCrc-208, OEMsr-1507, and OEMcmr-573, and these numbers add to a total of 2,288.

⁴For example, all long form atomic energy research contracts (let under OSRD’s S-1 division) were converted to

date of amendment, are noted in this list and thus measured in our data. In addition, several contracts are tagged as having a patent clause of “Purchase Contract” or “Overhead” (these are typically administrative contracts, previously described), or even “None”. The information on patent clauses from this contract list and in the invention disclosure cards overwhelmingly agrees, but where it conflicts, we use the information from the contract list, which appears to have been created specifically for this purpose.

After dropping administrative and cancelled contracts, there are 2,254 contracts in our data, made to 461 unique contractors, with a total value of \$462 million in 1940-45 dollars (equivalent to roughly \$7.4 billion today). For comparison, in the official administrative history of OSRD, Stewart (1948) claims that OSRD contractors had spent “approximately 457 million dollars through November 30, 1945” – a difference of less than one percent. We observe the contractor and OSRD division for every contract, the patent clause for 2,006 of the contracts (89%), and the obligated value and termination date for 2,208 of the contracts (98%).

We also observe the inventions produced under these contracts, which contractors were contractually required to disclose and OSRD subsequently catalogued. As mentioned above, these index cards list the contract that the invention was developed under and also include information on patent applications, which we use to link OSRD contracts to granted patents – these patents are what we denote in the paper as “OSRD patents”. We discuss how we do so below.

Using these index cards alone, we identify 7,879 reported inventions. Each card has an identifying number in the top left corner, numbering from 1 to 8040 (e.g., see Figure A.3), with some also suffixed with letters. In all, the numbering suggests there may have been as many as 8,056 inventions, and the 177 (=8056-7879) missing index cards could either be unobserved inventions or numbers which were skipped or discarded. To fill in these potential gaps, we do a wider search for data on OSRD-funded inventions across other archival collections at NARA, and find additional lists of OSRD-supported inventions in the records of the U.S. Army Judge Advocate General’s Office (Record Group 153). Using these records, we recover data for five of the 177 missing index cards, and identify an additional 26 unnumbered inventions which were reported after the index card file was no longer maintained. For the inventions on which patent applications were filed, we use the serial number to link these to granted patents, as we explain in more detail in the next subsection. In all, we have 7,910 inventions, on which 3,382 patent applications were filed, and 2,659 patents granted (which we identify by linking serials to granted patents). 2,657 of these patents were granted by 1980, when our sample for this paper ends.

For completeness, we also search for continuations, divisions, and continuations-in-part of OSRD patent applications. To do so, we parse the first 800 characters of the text of patents filed between 1935 and 1969 (application series 2, 3, and 4) to identify patents which mention an OSRD serial, and manually check these cases to determine whether they were continued or divided from earlier

short form in 1942, when the research program was spun off into a weapons development program (the Manhattan Project), to ensure that the government controlled the intellectual property.

OSRD-supported filings. Through this process we find an additional 104 OSRD-supported patents, bringing our total to 2,763 patents (2,761 granted by 1980).

Our final source of data on OSRD-supported invention is a distinct list of patent applications and granted patents related to the Manhattan Project, obtained by FOIA (from the U.S. Department of Energy) by researchers at the Woodrow Wilson Center and available at its digital archives.⁵ This document lists over 1,000 U.S. government-supported, nuclear energy-related patent applications (and >850 grants) from the World War II period, along with the contracts under which they were produced, including many OSRD contracts. Most of these inventions were placed into secrecy at the time of filing (see Gross 2023) and were not in the NARA records, likely due to their sensitive nature, and through this list we identify another 374 OSRD-supported patents (373 granted by 1980), bringing our total to 3,137 patents (3,134 granted by 1980).

A.2 Descriptive characterization of OSRD R&D investment

Administrative records from OSRD offer additional insight into the nature of the work it funded, and the firms and institutions it mobilized into the war effort. In prior work we have used these data to study how OSRD organized research for war and managed the World War II research effort (Gross and Sampat 2022c). Here we reproduce some of this evidence, while also adding more, to bring color to the specific features of the OSRD-led research effort.

Borrowing from Stewart (1948), Table A.1 shows OSRD’s organizational structure in the form it eventually evolved into as the scope of its work grew. As we explain in the paper, OSRD was the parent agency of two organizations, NDRC (which managed the technological research effort—e.g. radar, atomic fission, and more) and CMR (which managed the medical research effort—e.g., penicillin, antimalarials). NDRC grew to have into 19 core divisions and seven special sections and panels. These units covered a wide range of subjects and varied equally widely in scale. The table shows total contract authorizations only for 1943 onwards, when NDRC organized into this structure. The two largest divisions were Radar (14) and Rocket Ordnance (3), with the majority of funding going to MIT and CalTech, respectively, to support major research labs. NDRC also directed the atomic fission research program until it was transferred to the Army in mid-1943, which operated at a similar scale to the radar program while under OSRD.

CMR, in contrast to NDRC, was more contained with six divisions of roughly equal size (in dollars), and as we describe in (Gross and Sampat 2022c), operated very differently. Despite having less than one-tenth the budget of NDRC, CMR was nevertheless similarly important to the war effort, as infectious disease and other ailments had in previous wars killed more soldiers than battlefield wounds, making military medicine a key R&D priority.

Table A.2 illustrates the geographic concentration of OSRD research, listing the top 10 states with OSRD contracts (left panel) and OSRD patents (right panel). Massachusetts, California, New York, and New Jersey rank highly on both inputs and outputs, and together comprise around 65

⁵See <https://digitalarchive.wilsoncenter.org/document/165247>.

to 75% of both contract obligations and patents. Table A.3 lists the top 10 university and firm contractors. The major university contractors hosted large, central laboratories which led specific research programs, such as the Radiation Laboratory at MIT for radar (Gross and Roche 2023) and the Jet Propulsion Laboratory at CalTech for rockets and projectiles.

Table A.1: OSRD Divisions, Panels, and Special Sections (1941-1947)
[reproduced from Gross and Sampat (2022c)]

<i>National Defense Research Committee (NDRC)</i>		Contract Authorizations
Division/Section	Name/Description	(\$, '000s) (1943-1947)
1	Ballistics	5,327.2
2	Effects of Impact and Explosion	2,701.4
3	Rocket Ordnance	85,196.5
4	Ordnance Accessories	20,014.3
5	New Missiles	12,881.2
6	Subsurface Warfare	33,883.5
7	Fire Control	7,711.7
8	Explosives	11,079.9
9	Chemistry	4,698.2
10	Absorbents and Aerosols	3,524.2
11	Chemical Engineering	9,216.2
12	Transportation Development	2,199.4
13	Electrical Communication	2,073.9
14	Radar	104,533.4
15	Radio Coordination	26,343.0
16	Optics	5,923.9
17	Physics	7,655.3
18	War Metallurgy	3,794.4
19	Miscellaneous Weapons	2,416.1 *
AMP	Advanced Mathematics Panel	2,522.9
APP	Applied Psychology Panel	1,542.5 *
COP	Committee on Propagation	453.0 *
TD	Tropical Deterioration	232.4 *
SD	Sensory Devices	272.5 *
S-1	Atomic Fission	18,138.2 *
T	Proximity Fuzes	26,400.0 *
<i>Total</i>		400,735.1
<i>Committee on Medical Research (CMR)</i>		Contract Authorizations
Division	Name/Description	(\$, '000s) (1941-1947)
1	Medicine	3,873.3
2	Surgery	2,847.6
3	Aviation Medicine	2,466.5
4	Physiology	3,981.5
5	Chemistry	2,383.9
6	Malaria	5,501.9
-	Miscellaneous	3,635.3
<i>Total</i>		24,689.9

Notes: NDRC authorizations from January 1, 1943 onwards, except where noted below. CMR authorizations reported for the entire history of CMR.

* Authorizations for Division 19 from April 1, 1943; APP, from September 18, 1943; COP, from January 22, 1944; TD, from May 18, 1944; SD, from November 1, 1945. Authorizations for Sections S-1 and T are from June 27, 1940 onwards, with Section S-1 terminating in September 1943.

Table A.2: Top 10 states with OSRD contracts, by contract obligations
[reproduced from Gross and Sampat (2022c)]

Top states by share of OSRD contracts				Top states by share of OSRD patents			
Rank	County	OSRD contracts, 1941-1948		Rank	County	OSRD patents, 1941-1948	
		Value	Share of total			Number	Share of total
1	Massachusetts	\$147.9 mil.	32.0%	1	Massachusetts	597	21.4%
2	California	\$97.3	21.1%	2	New York	516	18.5%
3	New York	\$91.3	19.8%	3	New Jersey	509	18.3%
4	Illinois	\$22.7	4.9%	4	California	247	8.9%
5	District of Columbia	\$17.1	3.7%	5	Illinois	159	5.7%
6	Pennsylvania	\$14.7	3.2%	6	Pennsylvania	113	4.1%
7	Maryland	\$13.0	2.8%	7	Maryland	92	3.3%
8	New Jersey	\$12.3	2.7%	8	Ohio	64	2.3%
9	Ohio	\$8.6	1.9%	9	District of Columbia	61	2.2%
10	Michigan	\$6.9	1.5%	10	Delaware	54	1.9%
	Total	\$431.8	93.5%		Total	2412	86.6%

Notes: Table lists the top states by total obligations (left panel) and patents (right panel). Percentages measure each state's percent of total OSRD obligations/patents.

Table A.3: Top OSRD contractors, by contract obligations
[reproduced from Gross and Sampat (2022c)]

Top 10 Universities			Top 10 Firms		
Contractor	Total oblig.	Pct. of total	Contractor	Total oblig.	Pct. of total
Massachusetts Inst. of Tech.	\$106.8 mil.	23.1%	Western Electric Co.	\$15.2 mil.	3.3%
California Inst. of Tech.	\$76.6	16.6%	General Electric Co.	\$7.6	1.6%
Harvard University	\$29.1	6.3%	Radio Corp. of America	\$6.0	1.3%
Columbia University	\$27.1	5.9%	E. I. Dupont De Nemours & Co.	\$5.4	1.2%
University of California	\$14.6	3.2%	Monsanto Chemical Co.	\$4.5	1.0%
Johns Hopkins University	\$10.8	2.3%	Eastman Kodak Co.	\$4.3	0.9%
George Washington University	\$6.9	1.5%	Zenith Radio Corp.	\$4.2	0.9%
University of Chicago	\$5.7	1.2%	Westinghouse Elect. & Mfg. Co.	\$3.9	0.8%
Princeton University	\$3.6	0.8%	Remington Rand, Inc.	\$3.7	0.8%
University of Pennsylvania	\$2.9	0.6%	Sylvania Electric Products, Inc.	\$3.1	0.7%
Total	\$284.0	61.5%	Total	\$57.8	12.5%

Notes: Table lists the top firms and universities with OSRD contracts by total obligations. Percentages measure each contractor's percent of total OSRD research spending.

B Additional Data Sources

B.1 Construction of U.S. patent datasets

B.1.1 Base data

The construction of the patent datasets used in this paper begins with the USPTO historical master file (Marco et al. 2015), which provides a master list of utility patents with grant dates, patent class/subclass (USPC), and two-digit NBER category (Hall et al. 2001). In building this paper’s dataset, we restrict the sample to patents granted between January 1, 1920 and December 31, 1979—although most of the paper invokes only a subset of these, emphasizing the sample filed between 1930 and 1970. For all granted patents in this set, we obtain additional patent characteristics from the following sources:

- FreePatentsOnline.com (FPO): serial numbers, filing dates, and the network of forward and backward citations (front-page citations only)
- Derwent Innovation (2018) database (DI): assignee names⁶
- Petralia et al. (2016a) HistPat, Berkes (2018) CUSP, and Bergeaud et al. (2022) PatentCity datasets: inventor country, state, county⁷
- USPTO Historical Government Register: administrative data on government-funded invention (owned or licensed), collected since 1944⁸

The DI assignee names are (mostly) standardized and were later found to match those in Google Patents data, which are freely available through Google BigQuery. These data are mostly complete, but a small number of patents are missing filing dates and assignees. Table B.1 shows the number patents with missing data, by decade of grant. For the period sampled in this paper (1930-1970), approximately 2.3% of patents are missing a filing date and 2.3% missing an assignee (note: these percentages are calculated for patents granted between 1930 and 1970, whereas the paper uses the sample of patents known to have been filed between 1930 and 1970).

⁶Note that serial numbers, filing dates, and the network of patent citations were also retrieved from the Derwent database for comparison against the FPO data, as a validation exercise. The two data sources overwhelmingly agreed, and where they disagreed, spot checks revealed that FPO was consistently the more accurate of the two, and when there was an error in the FPO data, it typically reflected the occasional typographical error on the printed patent publication itself, such as two flipped digits, or a digit one unit off the correct value. Given their reliability, the data for this paper thus use serial numbers, filing dates, and citations from FPO.

⁷Also see Petralia et al. (2016), as well as Andrews (2019) for additional discussion of historical patent geography data. We evaluate the quality of these data, and use all three sources to produce a composite measure of inventor locations, below. We are grateful to the authors for sharing their data.

⁸We introduce these data, and compare them to Fleming et al. (2019), below.

Table B.1: Number of patents with missing data, by decade

Decade of grant	Patents	No filing date		No assignee data	
		Number	Percent	Number	Percent
1920-1929	414901	25738	6.2%	25918	6.2%
1930-1939	442842	11102	2.5%	11221	2.5%
1940-1949	307630	5470	1.8%	5546	1.8%
1950-1959	425985	12461	2.9%	12661	3.0%
1960-1969	567761	11203	2.0%	11363	2.0%
1970-1979	689027	2	0.0%	73	0.0%
Total	2848146	65976	2.3%	66782	2.3%

Notes: Table shows counts of patents with missing data, and their fraction of all patents, by decade.

Patented, OSRD-funded inventions are identified in the OSRD archival records by the serial number of the patent application. It is thus critical to have accurate data on serial numbers. We manually reviewed and validated the application-level data (serials and filing dates) from FPO for the period around World War II by checking patents with serial numbers or filing dates which are out of sequence. The important feature of the USPTO’s application numbering system for our purposes here is that applications are organized into application “series”, which span several years, and identified by a serial number within that series, generally issued in the order in which patent applications arrive at the USPTO, with serial numbers never exceeding six digits. Application series increment, and serial numbers reset, at the beginning of a year in which the serial numbers from the previous series are expected to surpass 1,000,000. Series 2 begins January 1, 1935 and ends December, 1947 and is the focus of this data cleaning effort. We take all patents identified by FPO as belonging to Series 2 and sort these patents by serial. We then look for patents where the previous and next serial have the same filing date but the given patent has a different filing date, and then manually validate the serial and filing date for these patents. Out of over 370,000 patents in Series 2, corrections were made to 279 serials and 188 filing dates. Although these corrections are valuable for matching patents to serials in OSRD records, the low error rate for this sample also indicates that such errors are not widespread in the data.

B.1.2 Harmonizing assignee names

Although the assignee names from DI are largely already standardized, closer examination reveals that there are still variants on individual assignee names (e.g., BELL TELEPHONE LABOR INC with >10,000 patents, and BELL TELPHONE LAB INC, BELL TEL PHONE LAB INC, and BELL TEIEPHONE LAB INC with 1 patent each). We undertake several procedures to further harmonize assignee names. We begin by sorting a list of assignees in alphabetical order, and for each assignee recording other nearby assignees up to 9 positions before/after in the sorted list. We then calculate the edit distance between the given assignee name and each of these nearby assignee names. When this edit distance is less than 25% of the length of the longer name in each pair, We flag that pair as a candidate for manual review. We then review all such matches for several categories of assignees, and standardize names when a match is found:

- Assignees with ≥ 15 patents between 1930 and 1960

- Assignees which were OSRD contractors
- Assignees identified as government agencies (see next section)
- Assignees identified as universities or hospitals (see next section)
- Assignees which were synthetic rubber manufacturers
- Assignees which were spinouts from Standard Oil

This process is repeated (because each round of harmonization may bring new assignees into the set with ≥ 15 patents between 1930 and 1960) until no new matches are found.

This harmonization is neither perfect nor exhaustive, but it is believed to be effective for the purposes of this paper. It is also worth noting that for the vast majority of assignee names which were standardized by this procedure, there was clearly a primary spelling for that assignee in the original DI data, with hundreds or thousands of associated patents in the case of large assignees, and at worst a handful of secondary spellings with one or two associated patents—such that the actual effects of both (i) performing this harmonization for the priority assignees above, and of (ii) not performing it for non-priority assignees, are likely minimal.

B.1.3 Determining assignee types

Assignees are then classified into four categories—firms, universities and hospitals, government agencies, and individuals—through a combination of rule-based and manual classification. We begin by classifying assignees as firms when the assignee name includes any of roughly 120 words which indicate firms (e.g., CO, CORP, INC, LTD, SPA, GMBH, etc., as well as technical words such as AERO, AUTO, CHEM, ENG, MACHINE, OIL, PROD, TECH, WORKS; full list available on request). We then manually classify remaining assignees with ≥ 15 patents between 1930 and 1960, as well as assignees whose name includes any of the following strings:

- COLLEGE, INST, UNIV, HOSP, RES FOUND
- US, CANADA, UK, FRANCE, GERMANY, SWITZERLAND, AUSTRALIA, JAPAN, ISRAEL, and assorted other countries
- ATOM (to identify international atomic energy commissions)

Assignees with >200 patents in the 1920-1979 period which are thus far unclassified are then classified as firms. Any remaining unclassified assignees are classified as individuals.

This classification procedure was developed over several years, and although—like the name harmonization—it is neither perfect nor exhaustive, random spot checks suggest it is overwhelmingly effective at categorizing assignees into the right bins. In total, 60.1% of patents with an assignee in the 1920-1979 sample are assigned to a firm, 0.2% to a university, 0.8% to a government agency, and 39.1% to an individual (numbers sum to $>100\%$ because 5% of patents have multiple assignees, and 0.2%

have assignees in multiple categories). Using administrative data, we will see below that the fraction we measure through names as assigned to a government entity is an undercount, primarily because the DI data sometimes undermeasure patent assignment.

B.1.4 Patent geography data

Measuring historical patents’ inventor locations with completeness and accuracy is critical to our analysis. For modern, post-1976 patents, the USPTO provides inventor locations as part of the electronic record, but for pre-1976 patents, locations are only available from the text of the patent itself, presenting a formidable measurement challenge. Over the past few years, several researchers have invested in harvesting inventor and assignee locations from patent text for the universe of historical U.S. patents, typically rely on a mix of OCR with rule-based regex string parsing, machine learning, and manual correction (see Andrews 2019, for a summary). The most recent such effort has been produced by Bergeaud and Verluise (2022), which expands on prior work by measuring the location of historical patents issued by European patent authorities.

Because efforts to measure inventor locations from patent text are potentially susceptible to significant errors, we combine three sources—the Petralia et al. (2016a) HistPat, Berkes (2018) CUSP, and Bergeaud and Verluise (2022) PatentCity datasets—to develop and validate a composite measure of first inventor location in which we have high confidence, with the inventor’s country, and if in the U.S., state and county. We begin by cleaning and harmonizing the inventor location data in each data source (country codes, state abbreviations, and county names). In some cases, FIPS codes are provided, but where needed, we assign FIPS codes to counties on state and name. We then restrict to first-listed inventors, calling their location the location of invention (a standard practice in research using patent data). We move forward with these data.

Our next step is to compare data sources, to evaluate how complete they are and how often they differ versus agree. Table B.2 provides summary statistics, including the rate at which they locate the patents in our sample, the rate at which they co-occur with and agree with other data sources. The simple takeaway is that CUSP and PatentCity are much more complete than HistPat, but any two sources agree on first inventor location for only 80% of patents.

Table B.2: Completeness and agreement of patent location data sources

Source	Overall Located	When HistPat located		When CUSP Located		When PatentCity located	
	(1)	Located	Matches	Located	Matches	Located	Matches
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HistPat	0.85	1.00	1.00	0.84	0.81	0.85	0.83
CUSP	0.93	0.93	0.81	1.00	1.00	0.94	0.85
PatentCity	0.94	0.94	0.83	0.94	0.85	1.00	1.00

Notes: Table summarizes the geolocation rates of patents by data source, and the pairwise rate of agreement across sources. For each row, the table shows the fraction of patents in our 1920-1979 sample for which it provides a first inventor location (Column 1). It then shows how often a location is provided when the column-wise data source also provides a location, and when so, the fraction of patents for which they agree (Columns 2-3, 4-5, and 6-7). Table illustrates the degree to which the data sources differ and we aim to reconcile.

These differences motivate our effort to aggregate these location measures to extract signal from the noise. To support this effort, we manually collected two ground-truth validation samples (patents

for which we hand-coded their locations), against which we could benchmark the performance of different approaches to aggregation. One validation sample was generated from the USPTO’s Index of Patents, an annual publication that lists all patents issued in a given year, along with inventor and assignee information. We selected two pages at random from each of the 1930, 1940, and 1950 editions and transcribed all patents therein, with approximately 350 patents in total.⁹ In a second validation sample, we select 200 patents at random from each of the years 1930, 1940, 1950, 1960, and 1970 and hire workers on Mechanical Turk to extract inventor and assignee locations, which we then review for consistency. Through iterative comparisons to these validation samples, we ultimately arrived at an approach where we assign each patent in our sample a location as follows. Where a majority of our data sources agree, we assign a patent the consensus or majority location. This rule accounts for 90.6% of patents. When multiple sources provide a location but our sources disagree, we apply the following rules, in sequence:

1. If the PatentCity relevance score == 1, use PatentCity Location. Else:
2. If the HistPat accuracy score == ‘High’, use HistPat location. Else:
3. If the HistPat accuracy score == ‘Medium’, use HistPat location. else:
4. If CUSP has a county, use CUSP location. Else:
5. If PatentCity has a county, use PatentCity location. Else:
6. If CUSP has a state, use CUSP location. Else:
7. If PatentCity has a state, use PatentCity location. Else:
8. If CUSP has a country, use CUSP location. Else:
9. If PatentCity has a country, use PatentCity location. Else:
10. Leave as missing.

The end result is the assignment of a location to 99.9% of patents in our sample, whose data sources break down as shown in Table B.3. When compared against our validation samples, we find that 98% to 99% of patents where all three sources agree on a location have the correct location, and 90% to 93% of those where two of three sources agree on a location have the correct location. However, when our sources disagree, or only one sources provides a location, it is correct only 45% to 50% of the time. To limit potential mismeasurement, throughout the paper we restrict our analytical sample to patents where at least two sources agree on location. In Figure B.1 we show the share of patents with 3, 2, 1, and 0 data sources, by year, where we can see that between 80% and 95% of patents each year have agreement among 2+ location data sources.

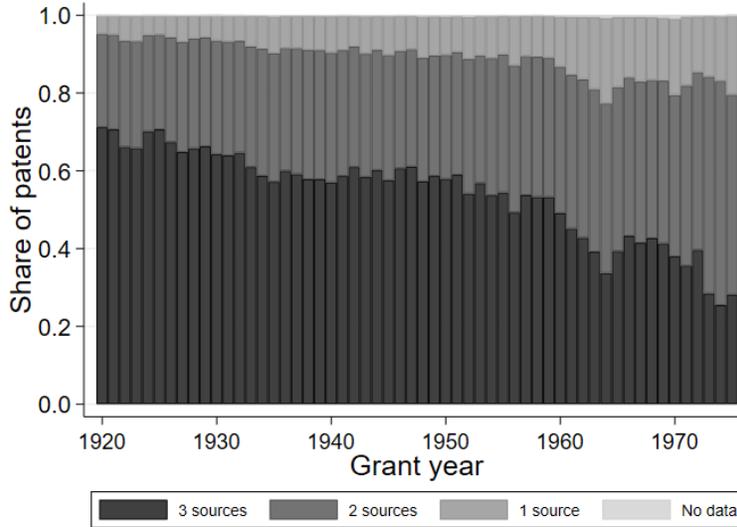
⁹The 1960 edition did not provide inventor locations.

Table B.3: Summary of final patent location measure

Source	Count	Percent	Cumulative
All sources (agreement)	1,600,150	56.2	56.2
2 of 3 sources (majority)	980,271	34.4	90.6
Single source (conflict)	265,040	9.3	99.9
No sources (no data)	2,685	0.1	100.0

Notes: Table summarizes the location data sources of all patents in our sample.

Figure B.1: Consensus across sources of patent geography data



Notes: Figure shows the annual fraction of granted U.S. patents from 1920 to 1975 for which we have a first inventor location where we have agreement (i) across all three data sources, (ii) across two sources, (i) no agreement across sources, and (iv) no data. Post-1976, USPTO began maintaining electronic records, such that inventor locations are broadly well-measured for this period.

In a final test, we also compare our data to administrative patent counts at the state-year level from the USPTO’s 1977 Technology Assessment & Forecast (USPTO 1977).¹⁰ Though accurately measuring counties is most crucial for our purposes in this paper, this comparison can be insightful as to whether our measurement is broadly on track. Figure B.2 plots state-year totals from each of our data sources (including our composite measure) against the USPTO state totals for the 1930 to 1960 period. All series are broadly consistent with each other and with administrative totals, but specific series seem to face measurement challenges in specific states (e.g., HistPat in WV, CUSP in FL, and PatentCity in RI, though it appears the USPTO data also have irregularities in RI). Our composite measure minimizes the RMSE against USPTO totals, being roughly 35% lower than that of PatentCity, 40% lower than HistPat, and 85% lower than CUSP.

¹⁰We thank Seamans et al. (2018) for pointing us to this source and sharing their data.

Figure B.2: State-year patent totals, HistPat vs. Administrative data, 1930-1960

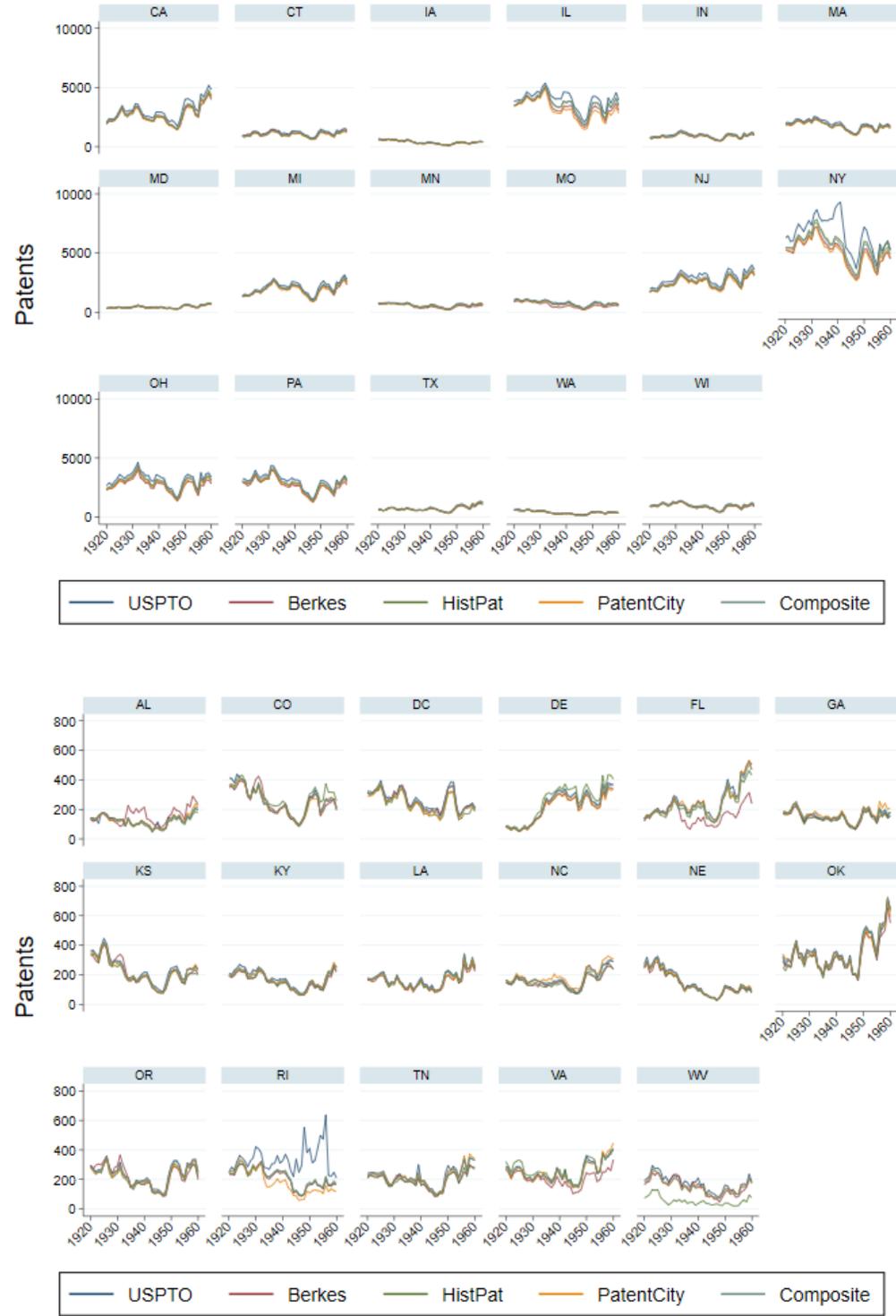
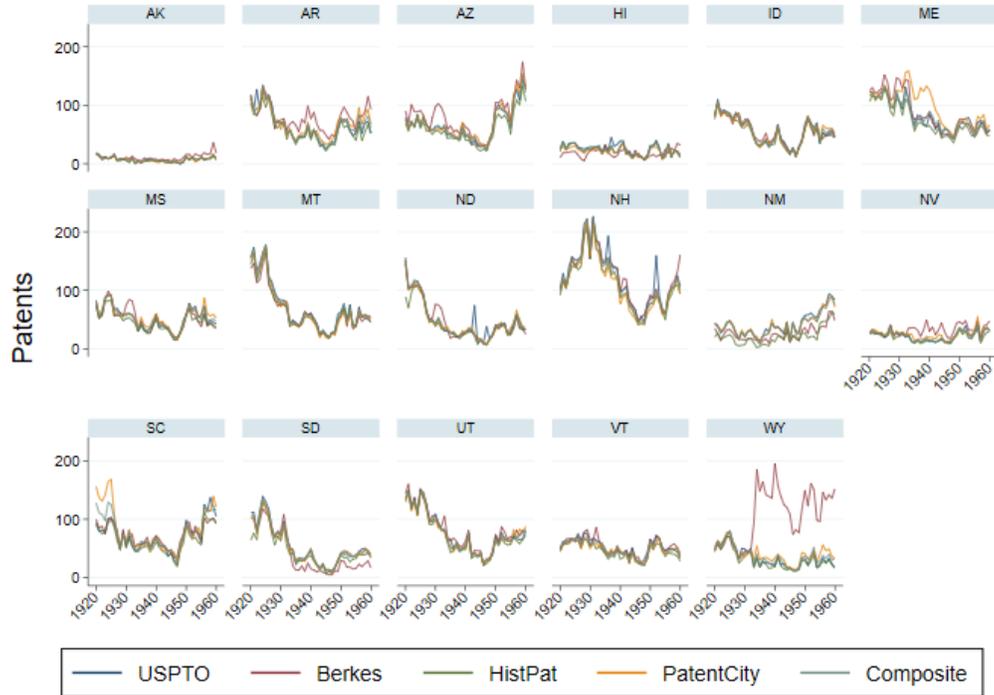


Figure B.2: State-year patent totals, HistPat vs. Administrative data, 1930-1960



Notes: Figure compares USPTO administrative totals of patent counts by state and year to those in each of our patent geography data sources (Petralia et al. 2016, Berkes 2018, Bergeaud and Verluise 2022), and to the composite measure we construct for this paper. Sample divided into the top third of states by average annual patents over the sampled period, middle third, and bottom third for ease of viewing.

Fixing county definitions

A final concern for location measurement is the fact that county definitions change over time, due to counties being combined, divided, or simply shifting borders—though most of these changes occur before 1940 (with the exception of independent cities in Virginia being carved out of, or merged into, their surrounding counties across most of the postwar period). To fix county boundaries across our sample period, we group up counties into the smallest stable units. We use two resources for tracking county boundary changes: the Atlas of Historical County Boundaries maintained by the Newberry Library, which provides a comprehensive list of county boundary changes for every state over the entire history of the U.S., and the U.S. Census Bureau’s list of substantial changes to county boundaries since 1970. Note that we apply these changes to other county-level data sources as well, to ensure geographic consistency across all of our analysis.

B.2 Data on government-supported patents

A first-order question we wrestle with in the paper is to what degree the effects of World War II R&D we observe are driven by continued public investment in the same locations and technologies as those funded in the war, and (or) whether they occur organically, driven by Marshallian increasing returns to scale. To answer this question, we not only want to relate wartime government-funded R&D to postwar R&D, but also to compare places and technologies which were shocked by World War II and received more versus less postwar public research investment. Doing so requires comprehensive data on U.S. government (henceforth, USG) funded invention.

Recent efforts to comprehensively measure USG-supported patents have done so by algorithmically reading patent text for patent assignments to government entities and for government interest statements or other in-text mentions of government support (Fleming et al. 2019). With considerable effort, albeit also with the imprecisions of algorithmic measurement in a very large corpus, this work captures the set of patents where government interests can be discerned from the patent document. But this approach also bears important limitations that could result in a significant undercount, especially in the historical period, and present problems for our exercise. The first issue is that a large number of government-supported, patented inventions are produced by contractors and grantees in the course of supported work, who sometimes retain title to these inventions with the funding agency holding a paid-up license for government use. The Department of Defense, for example, has historically been a major funder of applied research and invention but has not taken title to patents, instead letting contractors have title while retaining a royalty-free license.¹¹ Though many such patents write government interest statements into their text, these are systematically required by agencies only after the Bayh-Dole Act of 1980, and even then, compliance and enforcement have been haphazard (Rai and Sampat 2012). Additionally, many patents by government employee-inventors have historically not been assigned nor acknowledge government employment, and are thus undetectable by reading and processing patent text. The text-based approach will thus tend to undermeasure USG-supported invention.¹²

In light of these measurement challenges, we chose to pursue administrative data on government-funded invention. Since 1944, USPTO has maintained a “Register of Government Interest in Patents”, pursuant to Executive Order 9424 (18 Feb. 1944), which contains information on patents resulting from government grants and contracts. This Register has been used in research before: Watson and Holman (1964) study the so-called “Government Register” to report on the U.S. government patent portfolio, describing an index card series on which government interests were recorded. To our knowledge, it has not been revisited since.

We located these records at the U.S. National Archives (NARA) and arranged to have them scanned

¹¹Watson and Holman (1964) note that individual agencies are usually either “title-policy” or “license-policy”, and document that at least through the mid-1960s, more than 2/3 of federally-funded patents were licensed, not owned, by the agencies that funded the underlying R&D.

¹²As a more pedantic matter, as is we are able to identify roughly 1.5% more government-assigned patents in the Fleming et al. (2019) data than the authors identify, using the assignee data in their replication package, though this difference is too small to meaningfully alter our work.

and transcribed.¹³ These records comprised 127,852 index cards, of which examples are shown in Figure B.3. The cards provide several pieces of information: identifying information (patent number and issue date, serial number and filing date, title), the assignor (i.e., contractor) and inventor, the sponsoring government agency (e.g., Army, Navy, Air Force, AEC, NASA), and the government interest (title vs. license, as well as Act of 1883 or U.S.C. 266, which are legal statutes which assign the government rights to employee inventions). Though the collection of these data began in 1944 under the E.O., the records include patents filed and/or issued as early as the 1890s, and as late as the early 1990s, when USPTO completed its transition to electronic records.¹⁴

We undertake several steps to clean and regularize these data, including: hand-checking the values of numeric fields with non-numeric characters, correcting errors in transcription as well as on the original cards; confirming that all identifying information is internally consistent, and manually resolving inconsistencies; and harmonizing government agency names and spellings, aggregating them up to modern cabinet-level departments where possible (e.g., Army, Navy, Air Force, War Department, NSA all become DOD; AEC becomes DOE; HEW, PHS, NIH all become HHS; etc.). The Government Register also identifies a number of OSRD patents, as the example in Figure B.3 illustrates, albeit far fewer than the OSRD records do, as many of the OSRD-funded inventions end up in these records marked as assigned or licensed to the armed services. We thus rely on OSRD's official, archival records to measure OSRD-funded invention, and on the Government Register to measure USG-funded invention in the wider population of patents.

¹³See NARA Record Group 241, "Records of the Patent and Trademark Office". The data are from records titled *Index to Patent Assignments by Government Licensees, January 1, 1890–December 31, 1955*. This title is a misnomer, as the records measure patents in which the U.S. federal government has an interest (i.e., title or license), and the records continue into the 1990s. See <https://catalog.archives.gov/id/159071266>.

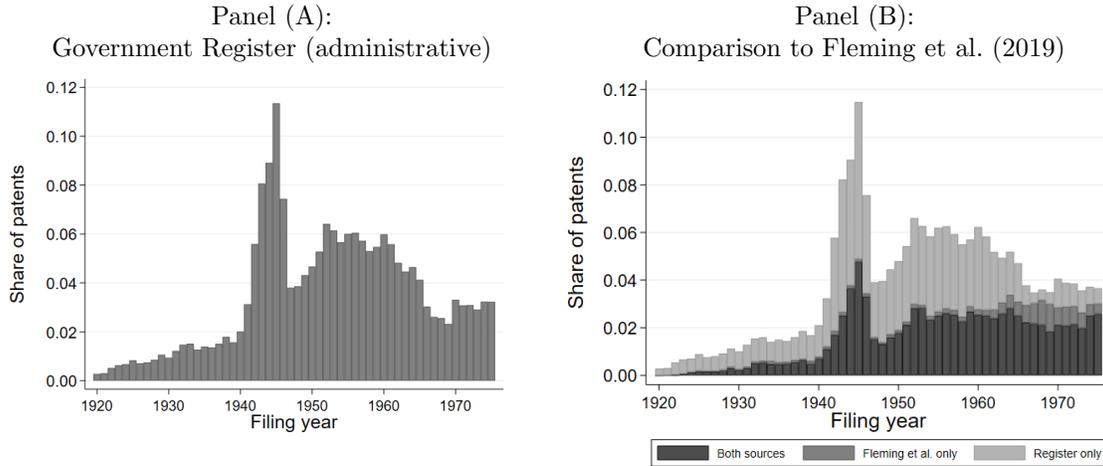
¹⁴Analogous, modern data are recorded in the USPTO's Patent Assignment Dataset, where all assignments of interests (title, license, and more) are recorded. See Marco et al. (2015b). Conveyances of title and license can also be searched online at <https://assignment.uspto.gov/patent/index.html#/patent/search>.

Figure B.3: Example Government Register index cards



In the figures below we illustrate what these data have to offer. Using the Register data, in Figure B.4, Panel (A) we first measure and plot the government-funded share of all U.S. patents by filing year, from 1920 to 1975, where we can see the USG share reaching its peak in World War II (note: when the series is extended into the 2000s, using modern data, we find that this share continues dropping). In Panel (B) we compare our data against those of Fleming et al. (2019), measuring what fraction of patents are in one, the other, or both sources. For most of the period we study, the Government Register data more than doubles the known number of government interest patents, with the Fleming et al. (2019) set fully subsumed by the Register data—though in the late 1960s and early 1970s, some patents begin to appear in Fleming et al. (2019) which are not present in the Register, potentially because it is less complete for these years.

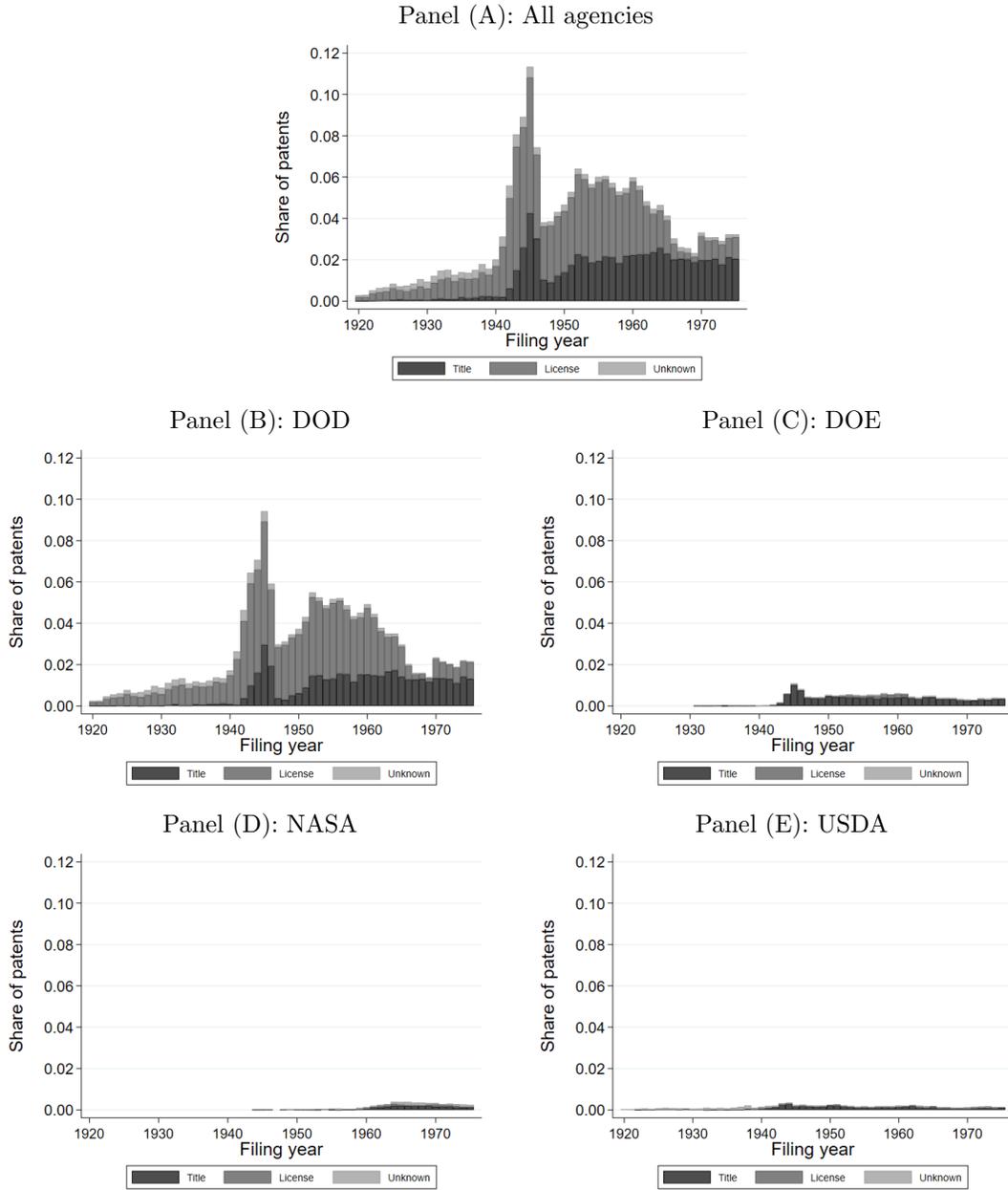
Figure B.4: Government-funded share of U.S. patent applications, 1920-1975



Notes: Figure shows the government-funded fraction of U.S. patents from 1920 to 1975 (Panel A) and compares the annual government share of patents in our newly-collected administrative data to Fleming et al. (2019) (Panel B).

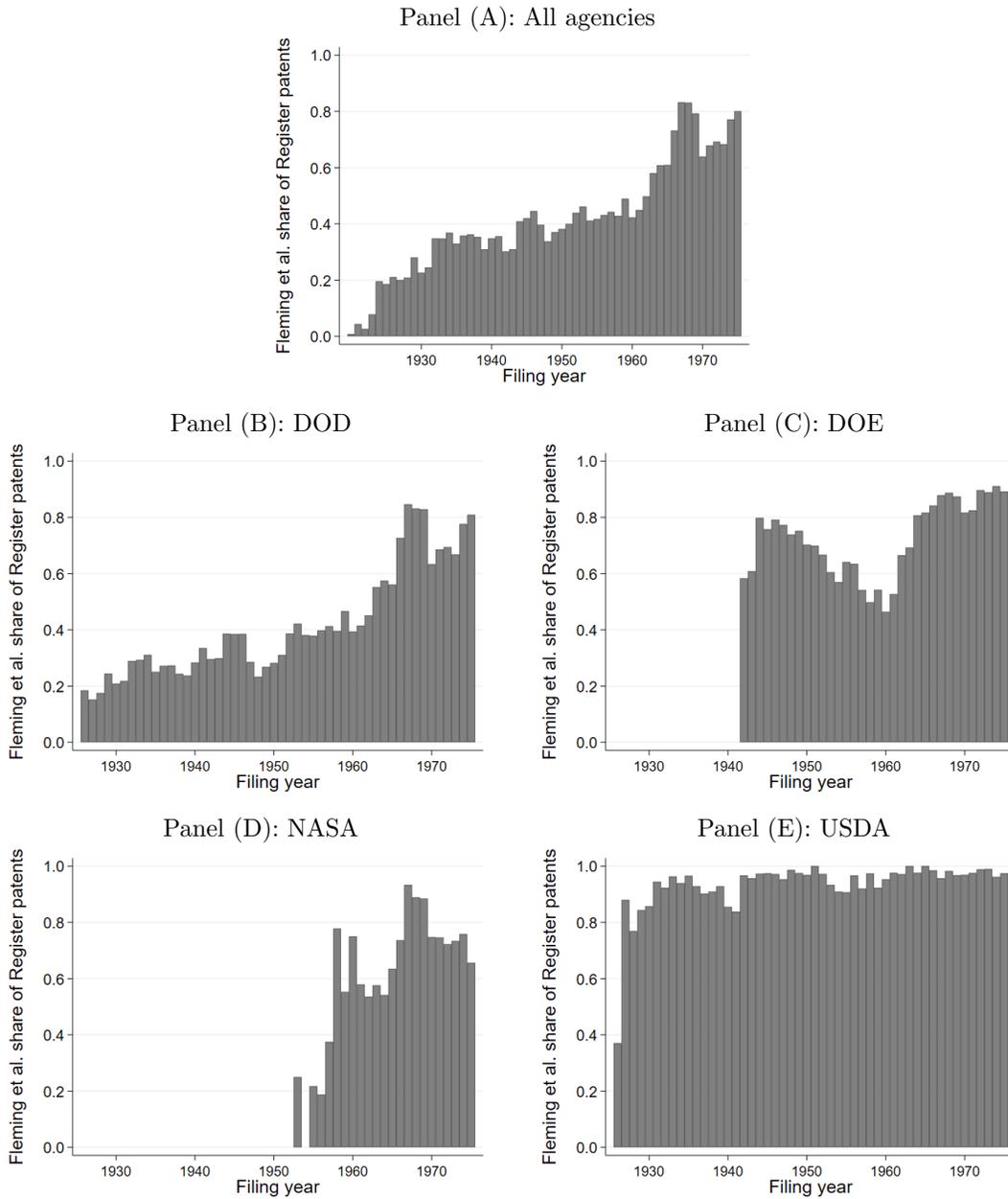
Figure B.5 examines the share of Register patents over time where the government has title, license, or an ambiguous interest, first overall (Panel A) and then for several of the larger R&D-funding agencies (Panels B to E). Here we see that license policy patents make up the vast majority of government-funded invention until the late 1960s, explaining the gap between the Register and Fleming et al. (2019) counts. We can also see that DOD is by far the driving force behind USG-supported invention: the next-largest agencies are only a small fraction of its size. In Figure B.6, we plot the share of Register patents identified as government assigned or supported (acknowledging), overall (Panel A) and by agency (Panels B to E). This share grows over time at DOD, fluctuates between 50% and 80% at DOE and NASA for most of the period, and is consistently near 100% at USDA. The completeness of the B.6 is largely driven by the degree to which each agency applies a title policy versus a license policy, and how this changed over time.

Figure B.5: Title vs. license policy share of U.S. patent applications, by agency, 1920-1975



Notes: Figure shows the government-funded fraction of U.S. patents from 1920 to 1975, overall (Panel A) and for the four highest patenting agencies (Panels B to E). Each subfigure shows the fraction of patents in our data that are title policy (i.e., government retains title) versus license policy (inventor allowed title, government retains paid-up license). License policy patents comprise two-thirds of government-funded invention but prior to these data could not be systematically measured.

Figure B.6: Share of patents captured by Fleming et al. (2019), by agency, 1920-1975



Notes: Figure shows the share of government-funded patents from 1920 to 1975 which appear in Fleming et al. (2019), overall (Panel A) and by agency (Panels B to E). These shares correlate with each agency’s use of a title patent policy, both in the cross-section and over time, since title patents are assigned to the U.S. government and easily identified in modern data. DOD was historically a license policy agency but increasingly used title policy over time, while USDA has always been a title policy agency.

B.3 Construction of foreign patent dataset

We complement the U.S. patent data with foreign authority patent data. Specifically, we collect data on all utility patents granted by the U.S., British, and French patent authorities between 1920 and 1979 from the EPO PATSTAT database (EPO 2017). Much like the U.S. data, these data include grant dates and patent class (IPC), which are the main information used in this paper, as well as filing date, number of inventors and original assignees, the cross-jurisdiction patent family size, and the patent family’s forward citations. In preparing these data, we restrict to patents from each patent authority (US, GB, FR) of inventions (as opposed to utility models in some jurisdictions, or design patents), and we restrict the U.S. patents in this dataset to those which are also in the USPTO dataset which we constructed and described above.

B.4 Other datasets

B.4.1 Historical U.S. County Business Patterns data

In the paper, we use U.S. County Business Patterns (CBP) data (U.S. Census Bureau 1980) to go beyond the patent record and study the effects of OSRD on local high-tech industries. The CBP measures establishments, employment, and wages at the level of counties and SIC industries, and was published at irregular intervals from 1946 to 1964 and annually thereafter. We collected data from several postwar CBP editions: 1947, 1959, 1970, and 1980. Early editions turned out to be too early and too aggregated for our purposes: too early because SIC 367 (“Electronic Components and Accessories”, a key industry in relation to the OSRD shock) was only created by the 1957 SIC classification, and too aggregated because they have too few 3- and 4-digit SICs—the level at which more precise links between technology areas and industries can be drawn. This is the case even up to the 1970 CBP. The 1980 CBP, however, reports lower-level SICs more widely. For this reason, and because it (i) is based on the 1972 SIC classification, consistent with the USPTO crosswalk we use to connect patent classes and industries, and (ii) matches the vintage of D&B data we use to study firm creation (discussed below), 1980 is our preferred edition.

One challenge to working with the CBP is that it historically suppressed employment and wages in small county-industry cells, to avoid identifiable disclosures. In these cases, there are multiple means of resolving these missing values (e.g., Eckert et al. 2021). We impute employment in these cells using data on the distribution of establishments across size bins, which is provided unsuppressed for all county-industries. To do so, we assign each establishment in the size distribution an employment level equal to the midpoint of its bin, and add them together.¹⁵ The correlation of these imputed values with reported values (where observed) is generally 0.8 or higher, although this correlation is in part driven by the skewed distribution of county size.

¹⁵These bins are as follows: 1-4, 5-9, 10-19, 20-49, 50-99, 100-249, 250-499, 500-999, 1000-1499, 1500-2499, 2500-4999, 5000+. For firms in the upper, unbounded bin, we assign them an employment level equal to that lower bound. An alternative approach is to regress total employment (where observed) on the establishment size distribution, and use the estimated parameters to predict employment for all county-industry cells—though in practice we found this to be less reliable than the midpoint-based approach, because it is estimated (vs. imputed).

Given our use of a USPTO crosswalk that maps SICs and patent classes to common industry codes (USPTO 2007), it is useful to describe this in more detail as well. This classification has a total of 41 product codes (which we call industry codes, in this paper, for convenience) that group to the following 15, not mutually exclusive industry categories:

- R1: Chemicals and Allied Products
- R2: Chemicals, except Drugs and Medicines
- R3: Basic Industrial Inorganic and Organic Chemistry
- R4: All Other Chemicals
- R5: Primary Metals
- R6: Machinery, except Electrical
- R7: Other Machinery, except Electrical
- R8: Electrical and Electronic Machinery, Equipment and Supplies
- R9: Electrical Equipment, except Communications Equipment
- R10: Other Electrical Machinery, Equipment and Supplies
- R11: Communications Equipment and Electronic Components
- R12: Transportation Equipment
- R13: Motor Vehicles and Other Transportation Equipment, except Aircraft
- R14: Other Transportation Equipment
- R15: All Industries

Our focus in Section 5 is on the “Electrical and Electronic Machinery, Equipment and Supplies” category, which includes the following industries:

USPTO Code	Description	Crosswalked SICs
35	Electrical transmission and distribution equipment	361, 3825
36	Electrical industrial apparatus	362
38	Household appliances	363
39	Electrical lighting and wiring equipment	364
40	Miscellaneous electrical machinery, equipment and supplies	369
42	Radio and television receiving equipment	365
43	Electronic components, accessories and communications equip.	366-367

Industry 43 (SICs 366 and 367) was the category with the most associated OSRD patents, and the second-highest OSRD rate (behind ordnance and explosives). In light of their importance, it is also useful to describe some of the industries and products within these 3-digit SICs. The bullet list below provides the 4-digit subindustries and examples of products they manufacture, from the 1972 SIC classification manual. Where products are not listed, it is because the products approximately match the industry description.

- 366: Communications Equipment
 - 3661: Telephone and Telegraph Apparatus
 - 3662: Radio and TV Transmitting, Signaling, and Detection Apparatus

- * Aircraft control systems, Amplifiers, Antennas, Digital encoders, Electronic control systems, Inertial guidance systems, Laser systems, Linear accelerators, Microwave communication equipment, Radar equipment, Sonar equipment, Transponders
- 367: Electronic Components and Accessories
 - 3671: Radio and TV Receiving Type Electron Tubes, except Cathode Ray
 - 3672: Cathode Ray TV Picture Tubes
 - 3673: Transmitting, Industrial and Special Purpose Electron Tubes
 - 3674: Semiconductors and Related Devices
 - 3675: Electronic Capacitors
 - 3676: Resistors for Electronic Applications
 - 3677: Electronic Coils, Transformers, and Other Inductors
 - 3678: Connectors for Electronic Applications
 - 3679: Electronic Components (n.e.c.)
 - * Antennas, Circuit boards, Magnetic recording tape, Oscillators, Relays

B.4.2 Historical firm creation from Dun & Bradstreet

We supplement these data with distinct data from Dun & Bradstreet (1980) (D&B), which has for two centuries collected data on U.S. firms for the purpose of assessing creditworthiness, making these data commercially available. Electronic records begin in roughly 1970, and modern D&B data cover millions of U.S. firms. These data are useful for our purposes because they measure firms' locations, 4-digit SIC industries, and founding years, enabling us to produce a measure of business creation by county, SIC, and year. An important caveat of these data is that our measurement is conditioned on both inclusion in the D&B sample (though this is very large) and survival to a given data year (this is more limiting, especially with the passage of time). Because it appears the D&B sample grew significantly over the 1970s, and in order to be able to include the 1970s in our analysis, we use the 1980 D&B data, balancing data quality against the survival of older cohorts. As with the CBP data, we focus our attention on firms in SIC 366 and 367, measuring the number of firms created in these SICs by county and decade.

B.4.3 Rosters of technical staff at OSRD-funded R&D labs

We also collect data on individuals who were at two major OSRD-funded R&D labs during the war: the MIT Radiation Laboratory (Rad Lab or RL), which was the epicenter of the Allied radar research effort during WW2, and the Harvard Radio Research Laboratory (RRL), which worked on radar countermeasures (i.e., stealth movement and enemy radar jamming). Both labs were large, employing thousands of scientists and engineers over the course of the war, and the Rad Lab was considered so successful that it became a model for post-war federally-funded research labs and is celebrated as an important part of the history of MIT and the broader region.

We use records from the Harvard and MIT university archives to compile rosters of technical staff from each of these labs. The MIT archive's collection of Rad Lab records contains three documents

useful towards this end: (i) a list of former RL staff members published in June 1946, which lists the name, field, highest degree, year of degree, and years spent at the RL; (ii) the most reliable known address of former staff members, as of March 1946; and (iii) the post-war job (or grad school) placement of former staff members, as of January 1946. We digitize these data sources, harmonizing name formats in the process so that we can successfully link individuals across them. Although some individuals appear in only a subset of these records, for most we have complete records. From the Harvard archive, we observe only RRL staff member addresses, as of March 1946, without the ancillary information on educational background or post-war job placement. We supplement this information by collecting data on all pre-, mid-, and post-war patents by these individuals, which we can use to study the effects that these individuals might have had on local inventive activity after the war in the fields where they were active.

B.4.4 Universities and PhD student production across the 20th century

We also collect data on historical PhD graduates of U.S. universities, which we use to identify major universities. We collect two sources to do so: a National Academy of Sciences report on doctorate production in U.S. universities (NAS 1963), which covers the 1920-1962 period, and the NSF Survey of Earned Doctorates (NSF 2022), which covers 1958+. We then count PhD graduates by university, field, and decade, and identify universities with the most postwar (1950-1969) graduates in the physical and biological sciences. We also borrow a measure of leading institutions from the NAS (1963) report. See Appendix C for further discussion.

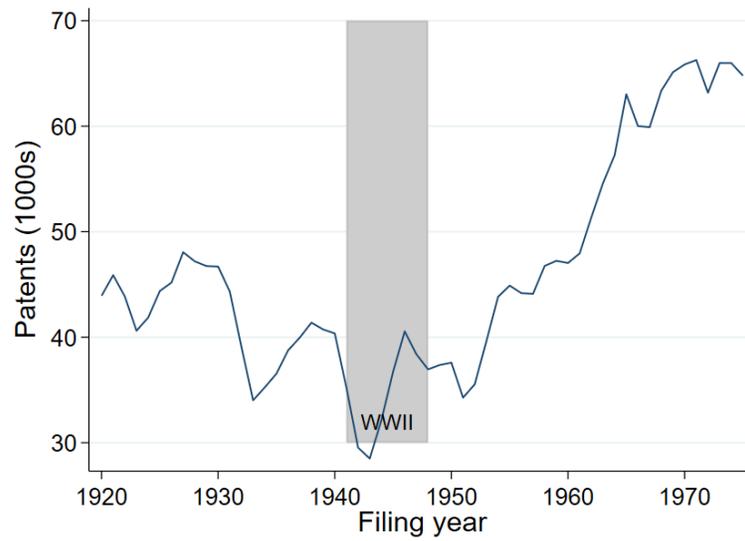
B.4.5 Population data, shapefiles for mapping, etc.

Where we normalize cluster patenting by county population, we use data from Forstall (1996) (obtained in 2018 from the U.S. Census Bureau’s website, at an address which no longer functions), which we interpolate across census decades for an annual measure. Where we map OSRD patenting or other variables, we use the 2000 TIGER/Line shapefile from Manson et al. (2022), filtered to the continental United States, as a base layer. As we noted earlier, county-level variables are grouped to their smallest stable geographic units throughout the paper.

C Supplementary Results

C.1 U.S. patenting in the 20th century

Figure C.1: Aggregate U.S. patenting, 1920 to 1975



Notes: Figure shows the time series of U.S. patents, by filing year, 1920 to 1975.

C.2 Top OSRD-supported technology fields

Table C.1: Top 10 technology areas with OSRD patents (denominator: all patents)

NBER	Description	Patents from OSRD contracts	Pct. of patents from OSRD, 1941-48	Max pct. OSRD in any year, 1941-48
44	Nuclear, X-rays	194	12.5%	24.8%
21	Communications	671	6.9%	16.6%
46	Semiconductor devices	15	5.4%	12.1%
42	Electrical lighting	241	4.2%	10.0%
22	Computer hardware/software	65	4.1%	8.5%
43	Measuring, testing	187	3.1%	6.8%
41	Electrical devices	308	2.5%	6.5%
45	Power systems	163	1.7%	4.2%
31	Drugs	27	1.7%	6.4%
49	Misc. (elec)	93	1.6%	3.7%

Notes: Table lists the top 2-digit NBER technology subfields (Hall et al. 2001), ranked by the fraction of 1941-1948 patents which were OSRD-funded, and the maximal fraction of OSRD-funded patents in any year.

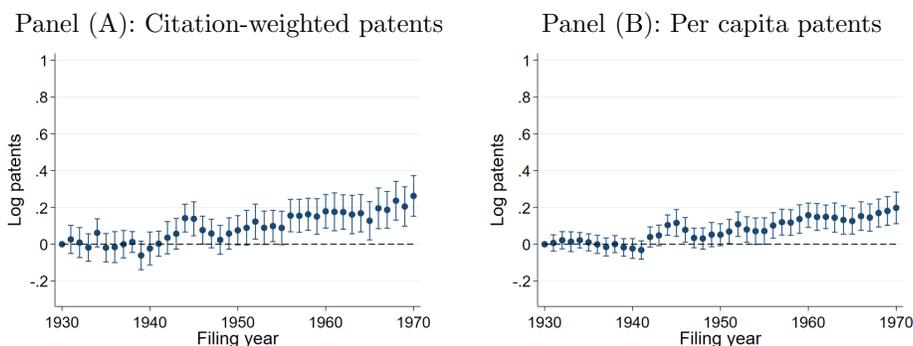
Table C.2: Top 10 patent classes with OSRD patents (denominator: all patents)

USPC	Description	Patents from OSRD contracts	Pct. of patents from OSRD, 1941-48	Max pct. OSRD in any year, 1941-48
376	Induced nuclear reactions	62	33.9%	81.8%
367	Acoustic wave systems/devices	95	20.4%	41.0%
102	Ammunition and explosives	174	16.6%	39.4%
343	Radio wave antennas	104	12.6%	33.9%
380	Cryptography	29	12.0%	21.4%
250	Nuclear energy	120	12.0%	23.8%
342	Directive radio wave systems/devices (radar)	200	10.9%	21.9%
333	Wave transmission lines and networks	108	10.5%	20.0%
327	Misc. electrical devices, circuits, and systems	88	10.3%	25.8%
708	Electrical computers	15	9.4%	22.2%

Notes: Table lists the top patent classes ranked by the fraction of 1941-1948 patents which were OSRD-funded, and the maximal fraction of OSRD-funded patents in any year.

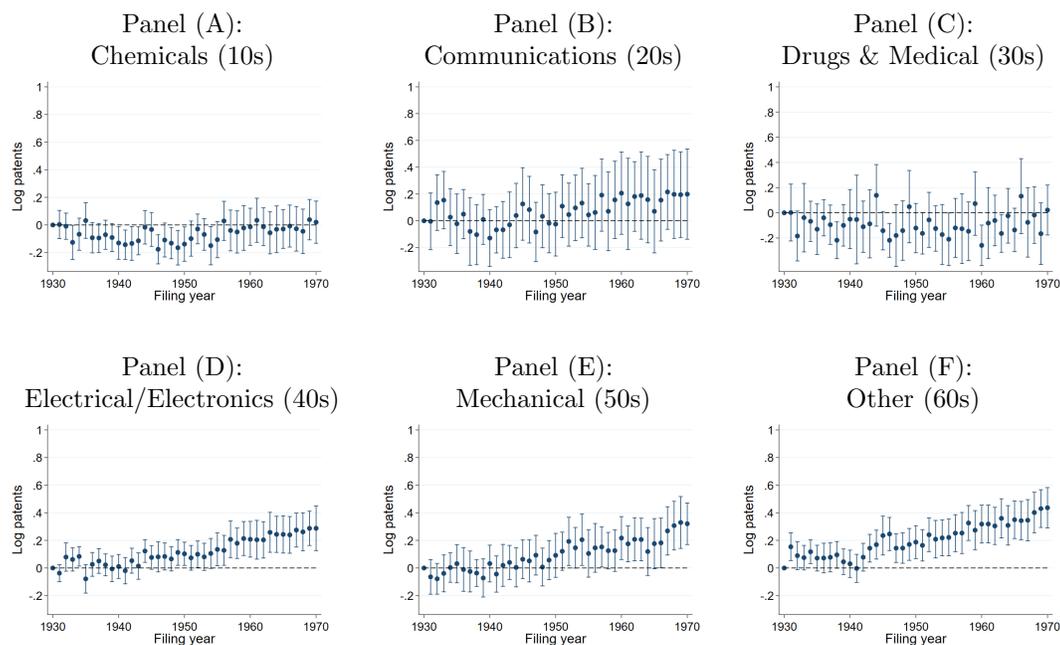
C.3 Additional results

Figure C.2: Effects of OSRD on citation-weighted patents and per-capita patents



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting on citation-weighted patents and per-capita patents. The independent variable measures the log fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

Figure C.3: Effects of OSRD on cluster patenting, 1930-1970, by technology area



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting, splitting the sample into high-level technology categories (Hall et al. 2001). The independent variable measures the log fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

Figure C.4: Effects of OSRD on cluster patenting, 1930-1970, omitting individual states



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting, omitting individual states from the estimation sample. The independent variable measures the log fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

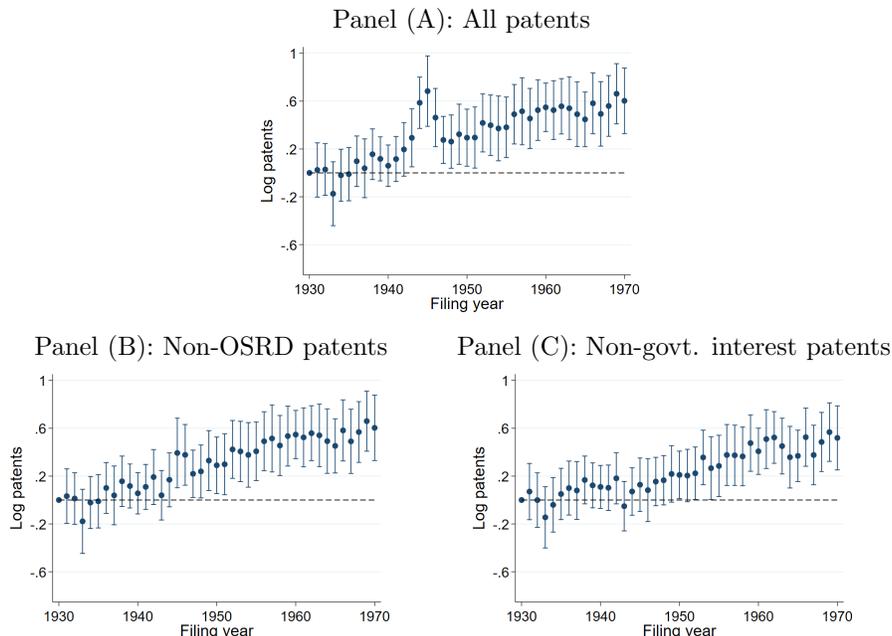
C.4 Alternative treatment measure: OSRD intensity quartiles

The following figures re-estimate our agglomeration results using a categorical treatment measure. We group clusters (county-categories) in four quartiles of treatment intensity (fraction of 1941-1948 patents OSRD-funded, conditional on any), and estimate changes in cluster patenting over time by treatment quartile, relative to a reference category of clusters without any OSRD-funded invention. Concretely, the specification we estimate here and elsewhere is:

$$\ln(\text{Patents})_{ict} = \sum_{q=1}^4 \sum_{t=1931}^{1970} \beta_{qt} \cdot \mathbb{1}(\text{Treatment quartile } q) \cdot \text{Year}_t + \alpha_{ic} + \delta_t + \varepsilon_{ict}$$

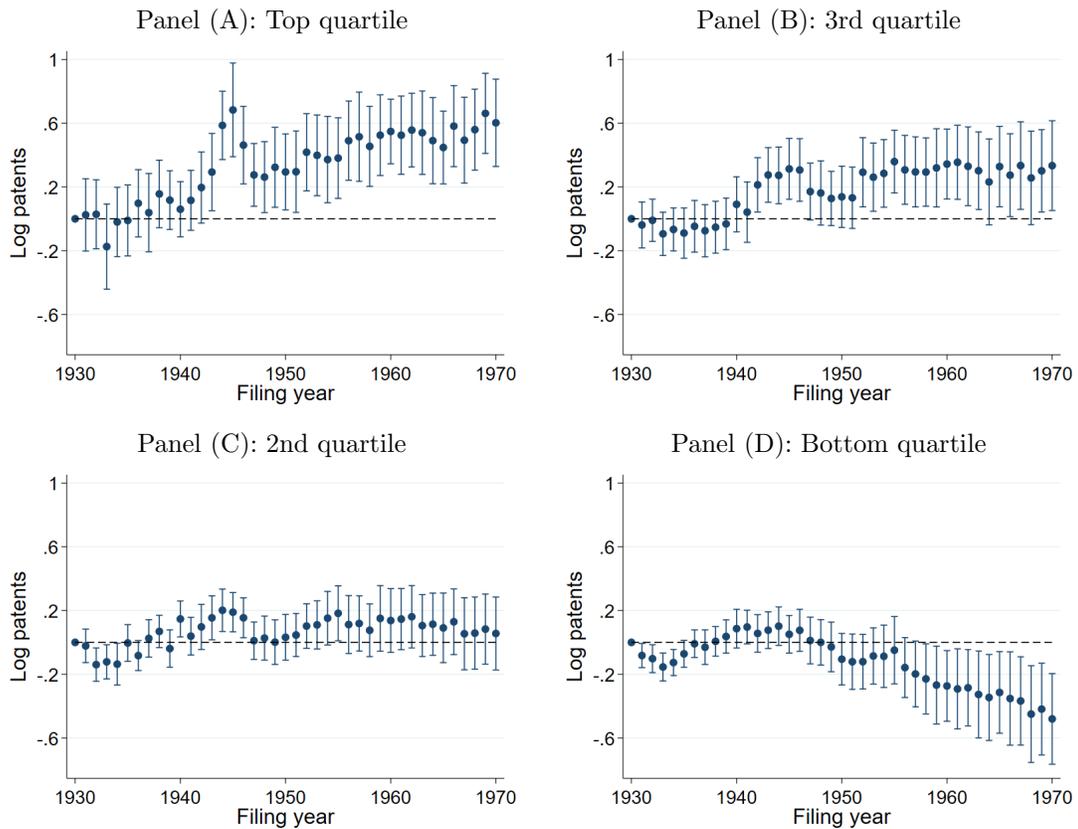
where i indexes counties, c indexes patent categories, and t indexes years, and the sample runs from 1930 to 1970, with standard errors clustered at the county level. Figure C.5 presents our key result for the top quartile of treated clusters, reproducing Figure 4 of the paper with this alternative treatment measure. Figure C.6 presents estimates for all four treatment quartiles, where the effect of the World War II shock can be seen to diminish in treatment intensity. Our preferred treatment measure for this paper is continuous, rather than categorical, because continuous measure preserve power (whereas categorical measures slice the sample into smaller cells, with wider confidence bands, as can be seen in these figures).

Figure C.5: Effects of OSRD on cluster patenting, 1930-1970, top treatment quartile



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting. The independent variable measures the log fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

Figure C.6: Effects of OSRD on cluster patenting, 1930-1970, by treatment intensity quartile



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting. The independent variable measures the quartile of treatment intensity, conditional on treatment (the fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded, conditional on any), with parameters across all four panels estimated jointly (in one regression) relative to a reference group of county-categories without any OSRD patents. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

C.5 Correlation of OSRD and postwar government-funded patenting

Table C.3 estimates the relationship of World War II OSRD patenting to postwar government-funded patenting across county-categories. The outcome variable is measured as the government-funded share of 1950-1969 patents, in levels (Column 1), logs (Columns 2 and 3), and relative to the median, conditional on non-zero (Columns 4 and 5). The explanatory variable is measured as the OSRD-funded share of 1941-1948 patents, in levels, logs, and quartiles. The table shows path dependence in the local intensity of government-funded invention.

Table C.3: Correlation of OSRD patenting and postwar govt-funded patenting across clusters

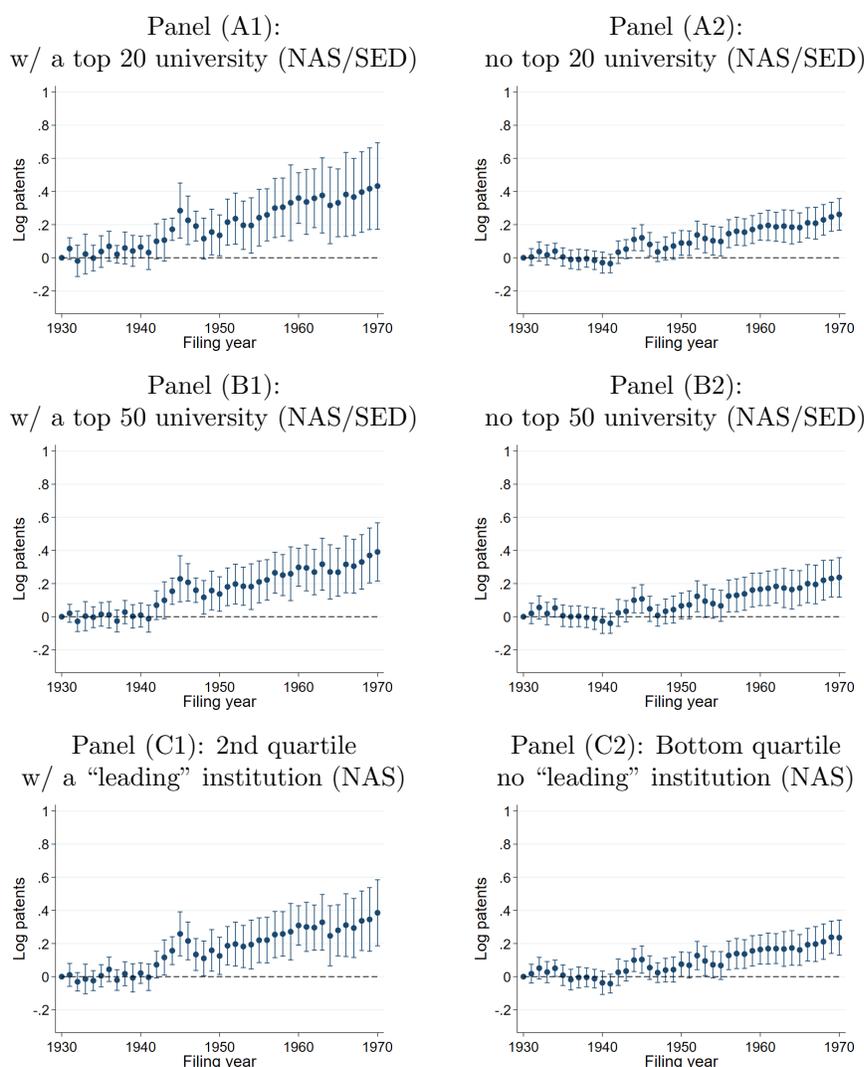
	USG rate, 1950-1969			Rel. to median	
	(1) Level	(2) Log+1	(3) Log	(4) Above	(5) Below
Wartime OSRD rate, 1941-1948	0.370*** (0.071)				
Ln(1 + Wartime OSRD rate)		0.415*** (0.065)			
Ln(Wartime OSRD rate)			0.463*** (0.040)		
Wartime OSRD rate in top quartile				0.417*** (0.045)	0.095*** (0.036)
Wartime OSRD rate in 3rd quartile				0.349*** (0.049)	0.262*** (0.043)
Wartime OSRD rate in 2nd quartile				0.251*** (0.045)	0.466*** (0.043)
Wartime OSRD rate in bot. quartile				0.063* (0.038)	0.672*** (0.039)
N	14521	14521	606	14521	14521
R^2	0.02	0.03	0.30	0.04	0.06
Y Mean	0.04	0.04	-2.48	0.10	0.17

Notes: Table relates county-category postwar government-funded patents to wartime OSRD patents. The OSRD rate measures the OSRD-funded share of county-category patents between 1941 and 1948, and the USG rate measures the government-funded share of county-category patents between 1950 and 1969. All columns include county-category and year fixed effects. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. SEs clustered by county in parentheses.

C.6 Differential effects in clusters with major universities

The following figures examine whether the effects of the OSRD shock vary in counties with vs. without a major university. As we explain in Section 4 of the paper, we measure major universities in two ways. First, we use a National Academy of Sciences report, and the NSF Survey of Earned Doctorates, to measure PhD graduates in the physical and biological sciences from 1950 to 1969 and identify the top 10, 20, and 50 universities by PhDs granted in these fields. Second, we borrow a measure of “40 leading institutions” from the National Academy report (which, according to the report, is ranked in terms of their total PhD production from 1920 to 1962). When we control for these measures, our results are unchanged. However, as Figure C.7 shows, we find effects of OSRD that are roughly twice as large in clusters with a major university.

Figure C.7: Effects of OSRD on cluster patenting, 1930-1970, by university presence



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patents in county-categories with and without a top university, by various measures. See text for definitions. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

C.7 Aggregate outcomes

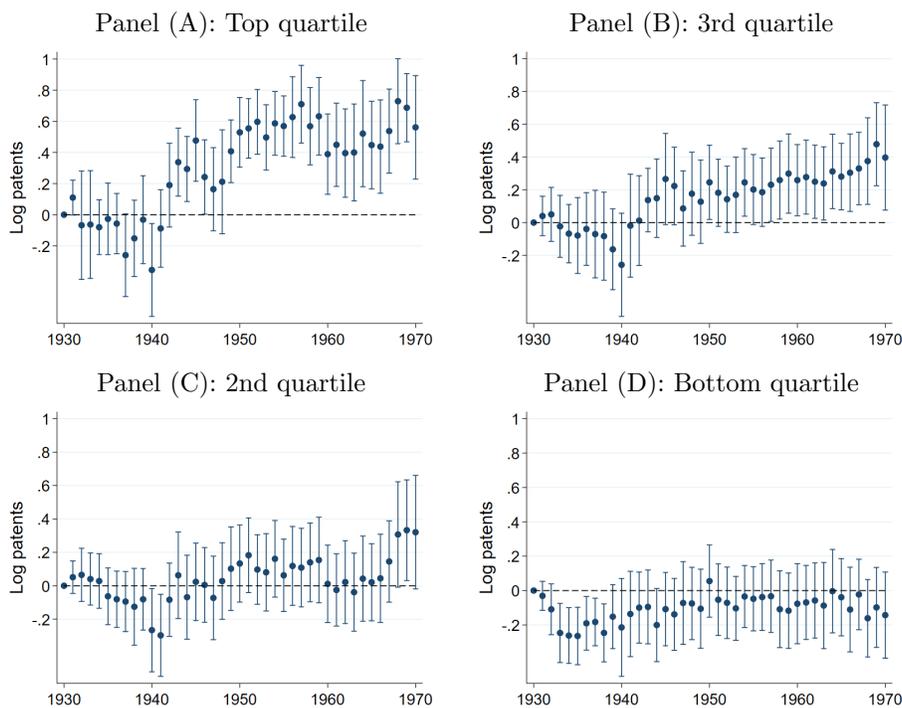
The following figures present annual estimates of the effects of World War II R&D on the direction of U.S. invention, disaggregating the simple difference-in-differences presented in Figure 9 in the body of the paper. We estimate the following specification:

$$\begin{aligned} \ln(\text{Patents})_{ict} = & \sum_{q=1}^4 \sum_{t=1931}^{1970} \beta_{qt} \cdot \mathbb{1}(\text{Country } i = \text{US}) \cdot \mathbb{1}(\text{Class } c \in \text{quartile } q) \cdot \text{Year}_t \\ & + \alpha_{ic} + \delta_{it} + \gamma_{ct} + \varepsilon_{ict} \end{aligned}$$

where i , c , and t index countries, technology classes, and years, the latter terms represent interacted fixed effects, and standard errors are clustered at the country-class level.

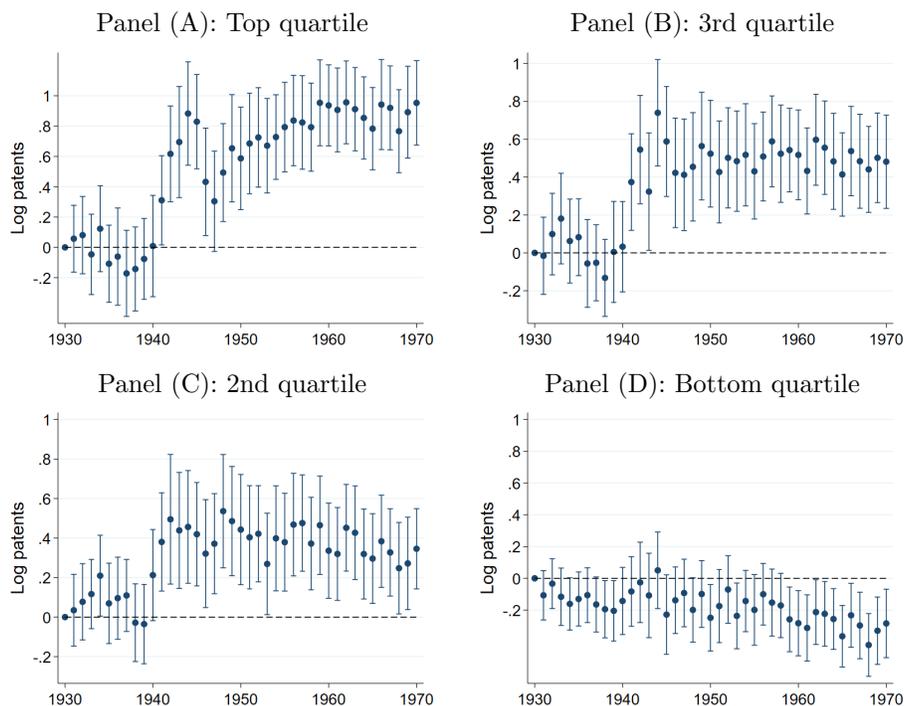
Figure C.8 presents the estimates for patenting at the USPTO vs. foreign patent offices. Figure C.9 does so for USPTO patents of U.S. versus foreign inventors. In both cases, the effect of the World War II shock is visible but diminishes in treatment intensity.

Figure C.8: Patenting at the USPTO versus UK/FR patent offices, by quartile of treated patent classes (relative to untreated classes), 1930-1970



Notes: Figure shows annual difference-in-differences estimates of the effects of the OSRD shock on patenting at USPTO versus UK and FR patent offices, in technology classes (IPC classes) in each quartile of OSRD treatment, as measured by the fraction of U.S. patents in those classes between 1941-1948 which were OSRD-funded, versus those without OSRD treatment. Error bars represent 95% confidence intervals, computed from SEs clustered at the country-class level.

Figure C.9: Patenting at the USPTO by domestic versus foreign inventors, by quartile of treated patent classes (relative to untreated classes), 1930-1970



Notes: Figure shows annual difference-in-differences estimates of the effects of the OSRD shock on USPTO patents with U.S. versus foreign inventors, in technology classes (USPCs) in each quartile of OSRD treatment, as measured by the fraction of U.S. patents in those classes between 1941-1948 which were OSRD-funded, versus those without OSRD treatment. Error bars represent 95% confidence intervals, computed from SEs clustered at the country-class level.

D Examining key specifications

D.1 Theoretical underpinnings

Our main specification is motivated by Romerian classes of endogenous growth models (e.g., Romer 1990), where the flow of innovation is proportional to the stock.

To see this more clearly, let A_t represent a cluster’s stock of varieties at the end of period t . Innovation, $\Delta A_t = A_t - A_{t-1}$, can be written $\Delta A_t = gA_{t-1}$, where $g > 0$ is the growth rate (which is, in turn, a function of R&D inputs). Iterating, and letting A_0 be the stock of varieties in a base period (such as 1940, just before the World War II R&D investment), we get $\Delta A_t = g(1+g)^{t-1}A_0$. The OSRD shock, in this case, can be interpreted as a windfall φ to the stock of local innovation or inventive capabilities in a particular field. Innovation in treated clusters will then follow a process $\Delta A'_t = g(1+g)^{t-1}(A_0 + \varphi)$, but in untreated clusters it follows ΔA_t .

We seek to compare innovation in treated and untreated clusters, which we will henceforth equate with patenting. Observe that the ratio of treated and untreated clusters’ patenting at time t is $\Delta A'_t/\Delta A_t = (1 + \varphi/A_0)$. Taking logs, we have that $\ln(\Delta A'_t) - \ln(\Delta A_t) = \ln(1 + \varphi/A_0)$, which indicates a constant difference at each post-treatment period t in log patents across treated and control clusters. This difference, in turn, is a function of OSRD-funded patents normalized by the size of the cluster (φ/A_0). We can turn this into a specification for a regression by attaching a β to $\ln(1 + \varphi/A_0)$. If untreated clusters are the omitted category against which treated clusters are compared, we can sweep $\ln(\Delta A_t)$ into a fixed effect. If we then want to estimate β across a panel of clusters, we can add cluster ic fixed effects (where i and c index counties and technology categories, respectively). Our model is then $\ln(\Delta A'_t) = \beta \cdot \ln(1 + \varphi/A_0) + \alpha_{ic} + \delta_t$.

This model can thus be translated into an estimating equation which can be used to test whether OSRD had a long-run effect—and more generally, whether the data are consistent with endogenous growth. If $\hat{\beta} > 0$, it suggests the OSRD shock had a Romerian effect, and if $\hat{\beta} \approx 1$, the data match the assumptions of Romer (1990) exactly. When we estimate this model as specified, we get $\hat{\beta} \approx 1$, reinforcing a Romerian interpretation of the data. We can also estimate time-varying parameters β_t , to evaluate whether the effects of this windfall change over time.¹⁶

The specification we estimate in the paper approximates the model above, but also includes a few differences. We measure φ/A_0 as the OSRD share of a cluster’s 1941 to 1948 patents, rather than normalizing by pre-war patents.¹⁷ We also replace $\ln(1+X)$ with $\ln(X)$, which allows estimated

¹⁶Time-varying parameters can be motivated in a variety of ways. For example, recombinant innovation is one reason why log patenting in differently-endowed clusters may diverge over time (Weitzman 1998). Complementarities are another. To motivate time-varying effects of endowment differences in a simple, mechanism-agnostic way, we can augment this model by defining $A_t = (1+g)^t A_0^{\beta_t}$, where β_t is a time-varying scaling parameter on initial stocks (e.g., $\beta_t = t\beta$). This leads to an estimating equation with $\beta_t \cdot \ln(1 + \varphi/A_0)$.

¹⁷Our choice to use cluster patents from 1941 to 1948 as the denominator allows the fraction to be interpreted as the intensity of a cluster’s mid-1940s engagement in the war effort, relative to its overall size at that time. We have performed the analysis with both 1930s patents and 1941-1948 patents in the denominator, and the choice is not material to the results since the two measures are very correlated ($\rho=0.91$).

parameters to be interpreted as elasticities. Though these choices slightly decouple the specification from the underlying theory, they also simplify the empirical interpretation. As we show in the next subsection, this does not materially affect our main takeaways.

Formally, our estimating equation is thus as follows:

$$\text{Ln}(\text{Patents})_{ict} = \sum_{t=1931}^{1970} \beta_t \cdot \text{Ln}(\text{OSRD rate})_{ic} \cdot \text{Year}_t + \alpha_{ic} + \delta_t + \varepsilon_{ict} \quad (\text{D.1})$$

where ic and t index clusters and years, respectively.

D.2 Alternative approaches

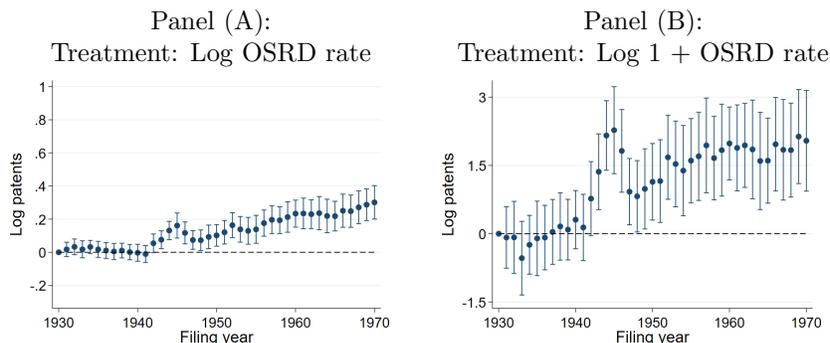
D.2.1 Treatment: Log 1 plus OSRD rate

Following the derivation in Section D.1 above, we could consider replacing $\text{Ln}(\text{OSRD rate})_{ic}$ in our main estimating equation with $\text{Ln}(1 + \text{OSRD rate})_{ic}$:

$$\text{Ln}(\text{Patents})_{ict} = \sum_{t=1931}^{1970} \beta_t \cdot \text{Ln}(1 + \text{OSRD rate})_{ic} \cdot \text{Year}_t + \alpha_{ic} + \delta_t + \varepsilon_{ict} \quad (\text{D.2})$$

This would have the added benefit of eliminating zeros in the argument, such that this object is defined for many more cluster years, but it would challenge elasticity interpretations, since natural logarithms are not scale independent (Mullahy and Norton 2022, Chen and Roth 2023). Figure D.1 presents the results when we do so. In the left panel we present the β_t parameters from Equation (D.1), as a reference point, and in the right panel the β_{1t} parameters from Equation (D.2). The results show a similar pattern, but larger magnitudes.

Figure D.1: Effects of OSRD on cluster patenting, 1930-1970



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting. The independent variable in Panel (A) measures the log fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. The independent variable in Panel (B) measures the log of 1 + the fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

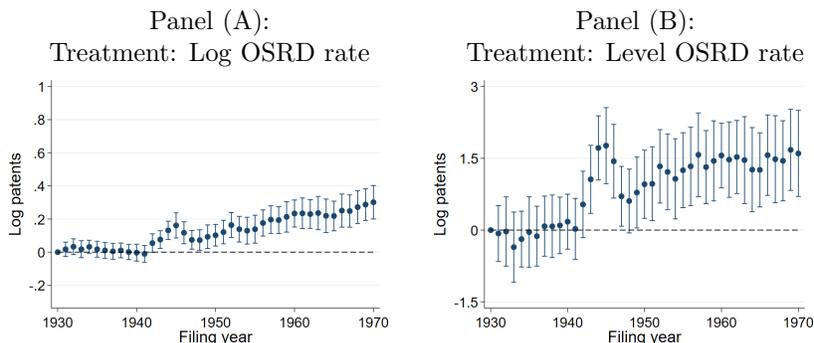
D.2.2 Treatment: Level OSRD rate

We could also consider replacing $\text{Ln}(\text{OSRD rate})_{ic}$ with $(\text{OSRD rate})_{ic}$ in levels:

$$\text{Ln}(\text{Patents})_{ict} = \sum_{t=1931}^{1970} \beta_t \cdot (\text{OSRD rate})_{ic} \cdot \text{Year}_t + \alpha_{ic} + \delta_t + \varepsilon_{ict} \quad (\text{D.3})$$

In this case the treatment measure will have more skew (which natural logs previously compressed), but on the other hand may permit more straightforward interpretation of the parameter (as semi-elasticities; e.g., a 10 percentage point increase in the OSRD rate yields a $0.1\beta_t$ increase in patenting). Figure D.2 presents the results when we do so. In the left panel we present (again) the β_t parameters from Equation (D.1), and in the right panel the β_{1t} parameters from Equation (D.3). The results show (again) a similar pattern, but larger magnitudes.

Figure D.2: Effects of OSRD on cluster patenting, 1930-1970



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting. The independent variable in Panel (A) measures the log fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. The independent variable in Panel (B) measures the level of the fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

D.2.3 Separating OSRD and total 1941-1948 patents

An alternative approach is to measure the treatment as a cluster's OSRD-funded patents. The interpretation of marginal effects would in this case answer a somewhat different—and arguably slightly narrower—question, but a valid one nonetheless. Moreover, in doing so, one would want to control for total cluster patenting during the years OSRD was active, so as to not confound OSRD clusters with generally-inventive clusters, which may correlate.

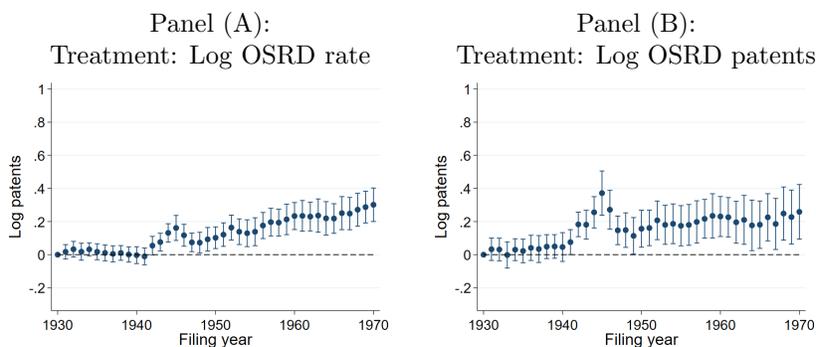
In this appendix, we examine an alternative specification, where we continue to measure logged cluster-year patenting as the outcome variable, but we replace clusters' log OSRD rate with their log OSRD patents as a measure of the shock, as well as log total 1941-1948 patents as a control.

The specification we estimate is accordingly the following:

$$\begin{aligned} \ln(\text{Patents})_{ict} = & \sum_{t=1931}^{1970} [\beta_{1t} \cdot \ln(\text{OSRD patents})_{ic} \cdot \text{Year}_t \\ & + \beta_{2t} \cdot \ln(\text{1941-48 patents})_{ic} \cdot \text{Year}_t] + \alpha_{ic} + \delta_t + \varepsilon_{ict} \end{aligned} \quad (\text{D.4})$$

These two constructions are mechanically very nearly the same, since $\ln(\text{OSRD rate}) = \ln(\text{OSRD patents}) - \ln(\text{1941-48 patents})$, but this new specification differs in that it permits the parameters on the two objects to vary (β_{1t} and β_{2t} , versus β_t). That the specifications are similar does not imply they must yield the same results, and indeed a parameter restriction like this (in Equation 1) could have a large impact on its estimated value. Figure D.3 presents the results from each, where in the left panel we present (again) the β_t parameters from Equation (D.1), and in the right panel the β_{1t} parameters from Equation (D.2). The results are statistically and quantitatively similar across specifications, reflecting in the end their close relation.

Figure D.3: Effects of OSRD on cluster patenting, 1930-1970



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting. The independent variable in Panel (A) measures the log fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. The independent variable in Panel (B) measures the log number of U.S. patents in each county-category between 1941-1948 which were OSRD-funded, controlling for log total patents in 1941-1948. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

E Alternative interpretations: OSRD or war demand?

One potential threat to this paper’s conclusions is the question of whether the OSRD effect we estimate may be attributable to other causes.¹⁸ The absence of pre-trends reassures us that OSRD investments did not coincide with local technological or commercial possibility, which was unlikely to spontaneously arise. But it does not rule out the possibility of contemporaneous shocks in the same technologies and locations, especially due to broader war-related causes. This can be a threat to identification if, for example, OSRD R&D coincided with other applied R&D investments during the war, or an attribution error if OSRD displaced war-induced innovation in the same technologies, performed in the same places, that would have taken place by other means. For example, the same firms involved in OSRD research may have made these investments on their own, in the course of fulfilling military supply contracts. Or these firms might have contracted with the same universities OSRD worked with. Or a different agency (especially the military branches) might have funded R&D in the same technologies and regions if OSRD had not.

To a casual reader these differences may not seem problematic. We measure the World War II R&D shock through OSRD records, but a shock is a shock, and maybe OSRD is just an instrumentality: we could view it as a proxy for whatever R&D investments took place in World War II, regardless of the means. We do not find this view satisfactory, as it’s not sufficiently discerning of mechanisms. In particular, government demand spilling in upstream R&D (demand-pull)—which one might view as either a potential confounder or a counterfactual to the OSRD-led research effort—is fundamentally different than the technology-push mechanism we attribute our results to, and understanding which was at play is essential to the generalizability of our findings.

In this appendix, we set out to more fully evaluate alternative-to-OSRD explanations. To reiterate the concern, it is that even without OSRD funding, the technologies and places OSRD funded may have seen R&D surge during the war. This could be because:

1. War supply contractors would have funded this R&D, had it not been funded by OSRD (e.g., in the course of serving a procurement contract, or in bidding for future ones).
 - (a) The war equipment suppliers may have performed this R&D themselves—and in the same technologies and locations that OSRD funded.
 - If war equipment suppliers were also the main OSRD R&D performers for a given technology, then their spill-in R&D investments in this counterfactual may have naturally taken the same geographic distribution OSRD R&D did.
 - If war equipment suppliers were not the principal OSRD R&D performers, but war production was co-located with OSRD R&D, that implies these firms had facilities in ‘OSRD places’ where this counterfactual R&D could occur, which might for this reason also take the same geographic distribution as OSRD R&D did.

¹⁸We thank an anonymous reviewer for posing questions which prompted the analysis in this appendix.

- If war equipment suppliers were not the principal OSRD R&D performers, and war production was not co-located with OSRD R&D, these firms could have still in principle set up R&D operations for the same technologies in the same places where OSRD did—even if their existing R&D facilities were elsewhere (or nonexistent—though we would argue it’s unlikely that firms with no R&D experience would be able to scale up R&D operations during and for the war effort).
- (b) The war equipment suppliers may have subcontracted to firms and universities in the same technologies and locations that OSRD funded.
2. Another agency (especially procuring agencies, like the War and Navy Departments) may have funded the requisite R&D—maybe contracting for R&D in the same technology areas, with the same firms and institutions, and in the same locations.

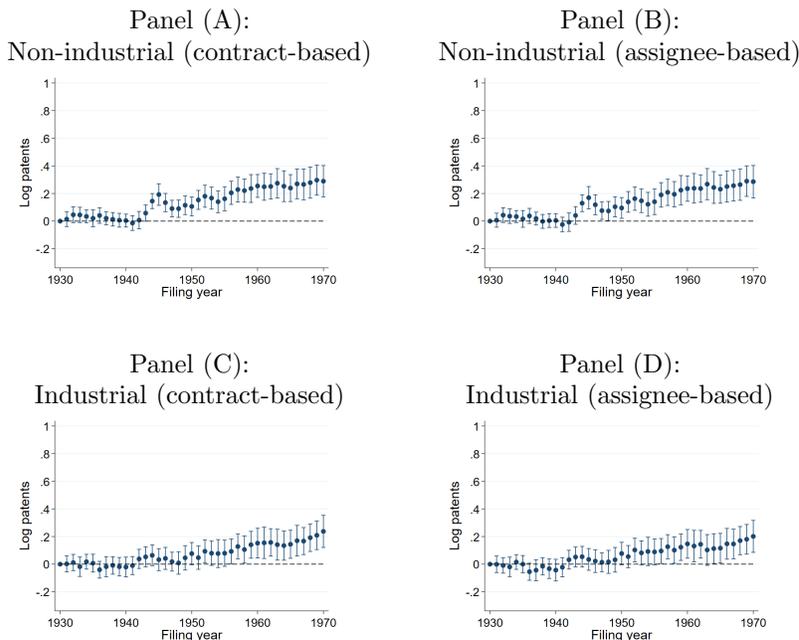
The remainder of this appendix will follow this sequence. First we evaluate whether war equipment suppliers were likely candidates to replace OSRD R&D. Supposing they were, and that they would perform this R&D themselves, we will then explore how likely it is that the geographic pattern would match that of OSRD R&D. We also discuss the prospects that in this counterfactual, war equipment suppliers might contract out the R&D in a pattern similar to what OSRD did. Finally, supposing that firms would not be R&D sponsors, we evaluate the likelihood that other agencies could replace OSRD as a similar-impact R&D funder. This analysis will use a mix of old and new data sources and combine empirical and qualitative evidence.

Were war equipment suppliers were likely candidates to replace OSRD R&D?

Across a broad set of indicators, this seems unlikely. The first piece of evidence is that only 25% of OSRD obligations, and 45% of OSRD-funded patents, were spent with or produced by firms. What really powered OSRD’s work was a collection of new wartime central laboratories which led major projects, like the MIT Rad Lab (radar), Johns Hopkins Applied Physics Laboratory (proximity fuses), CalTech Jet Propulsion Laboratory (rockets), and others at other institutions like Harvard, Columbia, GWU, and so on—not to mention Manhattan Project research, development, and enrichment sites. We think this is work that firms were unlikely to replace on their own in the absence of a coordinated, government-led wartime R&D effort.

Some additional robustness checks reinforce the importance of non-industrial R&D to our results. In Figure E.1, we re-estimate our main specification (Equation 1 in the paper) using only non-industrial OSRD-funded patents in our treatment measure. We measure non-industrial OSRD patents in two ways: those produced under non-firm contracts, and those without a firm assignee. For comparison, we also estimate Equation (1) with a firm-only treatment measure. Panels (A) and (B) estimate the effects of non-industrial OSRD patent intensity, and Panels (C) and (D) the effects of industrial OSRD patent intensity. The effects are present for all treatment measures, but larger in magnitude for the non-industrial treatment measures.

Figure E.1: Variation in estimated effects by treatment sector



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting. The independent variable measures the log fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded, and produced in the given sector. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

One might still conjecture that non-industrial and industrial OSRD-funded R&D were co-located, and that the latter is what really mattered—such that when the treatment is measured by non-industrial R&D only, Equation 1 is confounded by co-occurring industrial R&D that then enters the error term. We will return to the question of geography momentarily. Beyond this, one might also conjecture that if universities and university-based, government-funded central labs had not performed this R&D, firms may have, driven by war demand.

To evaluate this question more fully, we contacted the authors of Li and Koustas (2019), who collected microdata on major war supply contracts (i.e., those with value $> \$50,000$ in 1945 dollars) from postwar publications of the Civilian Production Administration. These data report the vast majority of World War II military equipment purchases at the individual contract level, and include the firm, product description, value, contract date, and more.

The war supply data offer an interesting and useful comparison to the OSRD R&D contract data. The first thing we note is that their portfolios were quite different. It is well known that war supply was concentrated in heavy equipment like aircraft, warships, tanks, and cargo vehicles and vessels (trucks, oil tankers, etc.)—whereas OSRD research was more heavily allocated to specific warfighting technology which was novel and high-impact, like radar, rocketry, atomic fission, proximity fuses, and such—which were relatively low cost compared to tanks and fighter planes. Accordingly, the major war supply contractors were broadly quite different from OSRD contractors. We show

this in Tables E.1 and E.2, which list the top 15 war supply firms, OSRD firms, and war supply categories, by their share of war supply and OSRD contracts.

Though the firms in the left and right halves of Table E.1 are generally different, a few firms show up in both sets, like General Electric (GE), Western Electric, Westinghouse, and Douglas. Douglas had no OSRD patents (that we know of), but we can take a deeper dive into the three others (GE, Western Electric, Westinghouse), however, which were the largest electrical firms of this era and important OSRD contractors. To get a better understanding of these firms, Table E.3 lists their top products supplied under war production contracts, and Table E.4 their top production locations. Even for these firms, the focus of war production is (mostly) different than the technology their OSRD-funded R&D focused on, and the locations of their war production are likewise (mostly) different than the locations of their OSRD patents. This is especially true for Western Electric and Westinghouse (whereas GE performed most of its R&D and production in the same, Schenectady location). These locations were generally not important OSRD hubs.

Table E.1: Top 15 firms by war supply and OSRD contract value

War supply contracts				OSRD R&D contracts			
Rank	Firm	Percent	Cum. Pct.	Rank	Firm	Percent	Cum. Pct.
1	Ford Motor Co.	2.6%	2.6%	1	Western Electric Co.	3.4%	3.4%
2	Douglas Aircraft Co.	2.0%	4.6%	2	General Electric Co.	1.7%	5.1%
3	Wright Aeronautical Corp.	1.9%	6.5%	3	Radio Corp. of America	1.3%	6.4%
4	Chrysler Corp.	1.9%	8.4%	4	E. I. Dupont De Nemours & Co.	1.2%	7.6%
5	Cons. Vultee Aircraft Corp.	1.8%	10.2%	5	Monsanto Chemical Co.	1.0%	8.6%
6	General Electric Co.	1.7%	11.9%	6	Eastman Kodak Co.	1.0%	9.6%
7	Curtiss Wright Corp.	1.6%	13.6%	7	Zenith Radio Corp.	0.9%	10.5%
8	Western Electric Co.	1.6%	15.1%	8	Westinghouse Elect. & Mfg. Co.	0.9%	11.4%
9	Lockheed Aircraft Corp.	1.4%	16.6%	9	Remington Rand, Inc.	0.8%	12.2%
10	Bethlehem Steel Co.	1.4%	18.0%	10	Sylvania Electric Products, Inc.	0.7%	12.9%
11	United Aircraft Corp.	1.3%	19.3%	11	Standard Oil Dev. Co.	0.7%	13.5%
12	General Motors Corp.	1.1%	20.4%	12	Erwood Sound Co.	0.6%	14.1%
13	Packard Motor Car Co.	0.9%	21.3%	13	Douglas Aircraft Co.	0.5%	14.7%
14	Glenn L. Martin Co.	0.9%	22.2%	14	Budd Wheel Co.	0.4%	15.1%
15	Westinghouse Elec. & Mfg. Corp.	0.8%	23.0%	15	Gulf Research & Dev. Co.	0.3%	15.4%

Notes: Table lists the top 15 firms with war supply contracts and OSRD R&D contracts, alongside their share of total contracts. War supply data from Li and Koustas (2019).

Table E.2: Top 15 products by war supply contract value

Rank	Product	Percent	Cum. Pct.
1	Bomber Airplanes	7.5%	7.5%
2	Airplane Engines	6.3%	13.7%
3	Fighter Airplanes	2.1%	15.8%
4	Trucks	1.7%	17.5%
5	Communication Equip.	1.4%	18.9%
6	Cargo Vessels	1.4%	20.3%
7	Airplane Engines	1.3%	21.6%
8	Aviation Gasoline	1.3%	22.9%
9	Destroyers Ships	1.3%	24.2%
10	Cargo Ships	1.2%	25.4%
11	Medium Tanks	1.1%	26.5%
12	Pursuit Airplanes	1.1%	27.6%
13	Ordnance Equipment	1.1%	28.7%
14	Airplane Prop Assemblies	1.0%	29.7%
15	Airplane Parts	0.9%	30.6%

Notes: Table lists the top 15 product categories for war supply contracts. Data from Li and Koustas (2019).

Table E.3: Supplied products of principal electrical equipment firms

Rank	General Electric					Western Electric					Westinghouse				
	Product	Percent	Cum. Pct.	Rank	Product	Percent	Cum. Pct.	Rank	Product	Percent	Cum. Pct.	Rank	Product	Percent	Cum. Pct.
1	Communication Equip.	8.9%	8.9%	1	Communication Equip.	28.2%	28.2%	1	Communication Equip.	13.2%	13.2%	1	Communication Equip.	13.2%	13.2%
2	Turbosuperchargers	8.3%	17.1%	2	Radio Sets	9.6%	37.8%	2	Machine Guns	8.5%	21.8%	2	Machine Guns	8.5%	21.8%
3	Fire Control Systems	8.1%	25.3%	3	Radar Equipment	7.0%	44.7%	3	Ordnance Material	6.9%	28.6%	3	Ordnance Material	6.9%	28.6%
4	Radio Equipment	4.4%	29.7%	4	Radio Equipment	6.2%	51.0%	4	Torpedoes	4.9%	33.6%	4	Torpedoes	4.9%	33.6%
5	Internal Comb. Eng.	4.0%	33.7%	5	Radio Parts	4.1%	55.1%	5	Radio Equipment	3.8%	37.3%	5	Radio Equipment	3.8%	37.3%

Notes: Table lists the top 5 products supplied under war production contracts by principal electrical equipment firms. Data from Li and Koustas (2019).

Table E.4: Manufacturing locations vs. patent locations of principal electrical equipment firms

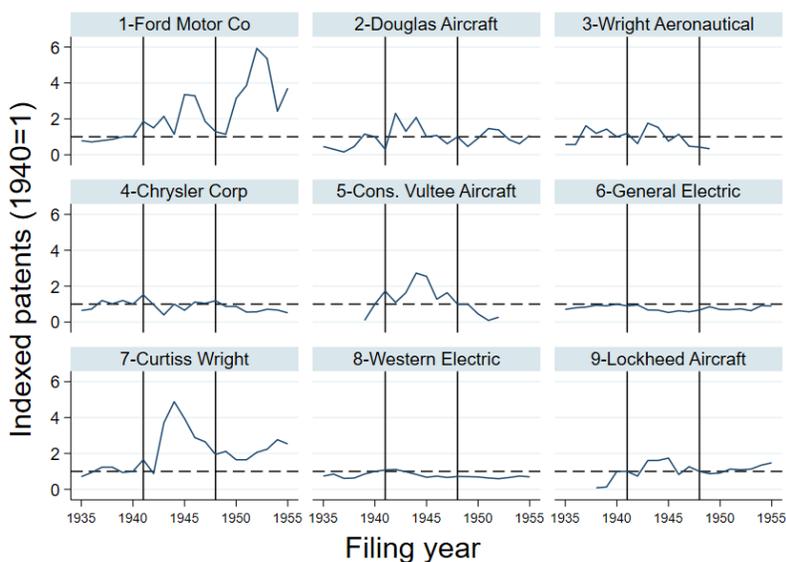
Rank	War supply manufacturing locations					Westinghouse					
	Location	Percent	Cum. Pct.	Rank	Location	Percent	Cum. Pct.	Rank	Location	Percent	Cum. Pct.
1	Schenectady, NY	42.8%	42.8%	1	Hudson, NJ	48.6%	48.6%	1	Baltimore, MD	25.8%	25.8%
2	Onondaga, NY	11.7%	54.6%	2	Cook, IL	34.6%	83.2%	2	Delaware, PA	18.3%	44.0%
3	Essex, MA	9.1%	63.7%	3	Baltimore, MD	7.4%	90.7%	3	Macomb, MI	8.5%	52.6%
4	Fairfield, CT	7.4%	71.1%	4	New York, NY	7.4%	98.1%	4	Allegheny, PA	8.3%	60.9%
5	Allen, IN	7.4%	78.4%	5	Passaic, NJ	0.7%	98.8%	5	Hampden, MA	8.0%	68.9%

Rank	OSRD patent locations					Westinghouse					
	Location	Percent	Cum. Pct.	Rank	Location	Percent	Cum. Pct.	Rank	Location	Percent	Cum. Pct.
1	Schenectady, NY	75.5%	75.5%	1	Essex, NJ	19.1%	19.1%	1	Essex, NJ	57.1%	57.1%
2	Fairfield, CT	5.7%	81.1%	2	Morris, NJ	18.7%	37.7%	2	Allegheny, PA	20.6%	77.8%
4	Cuyahoga, OH	3.8%	84.9%	3	Union, NJ	16.0%	53.7%	4	Baltimore, MD	4.8%	82.5%
4	Rensselaer, NY	3.8%	88.7%	4	New York, NY	8.9%	62.6%	4	Montgomery, MD	4.8%	87.3%
				5	Westchester, NY	8.6%	71.2%	5	Mercer, PA	3.2%	90.5%

Notes: Table lists the top 5 manufacturing locations for war production by principal electrical equipment firms in Panel (A), and top 5 OSRD patent locations for these firms in Panel (B). War supply data from Li and Koustas (2019).

Figure E.2 provides one last piece of evidence: we link top war suppliers to the patent data to measure their patenting around the war years. The time series in this figure are indexed to 1940, which is normalized to one. Despite the massive military demand these firms served, their patenting patterns over this period were very mixed. A couple of the major aircraft manufacturers had elevated patenting during the war (especially Curtiss Wright, and to a lesser degree Consolidated Vultee/Convair), but not all (e.g., more modest increases for Douglas, Lockheed). The electrical firms had materially lower patenting for the duration of the war. As we work further down the supplier list we continue to see that besides aircraft manufacturers, major war suppliers saw little to negative changes in patent filings—including most automakers, shipbuilders, steel producers, Du Pont (also an OSRD contractor), and others. One takeaway of this evidence is that war supply contracts did not necessarily spill in R&D by the same firms.

Figure E.2: Patents filed 1935 to 1955 for top war supply contractors (indexed to 1940)



Notes: Figure shows annual patenting between 1935 and 1955 for top war supply firms, indexed to 1940. The dashed horizontal line marks parity (index=1). Solid vertical lines mark 1941 and 1948, to visually demarcate the war years.

Broadly, we think this evidence collectively suggests against counterfactual (a), that war equipment suppliers were likely candidates to replace OSRD R&D.

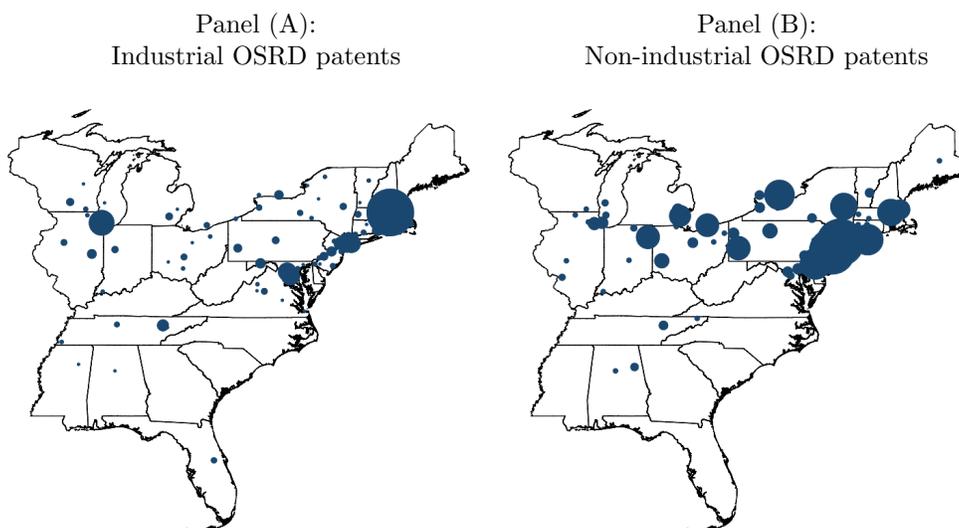
Suppose they were. Would in-house R&D have followed the same geographic pattern?

In any case, let us suppose that firms may have replaced OSRD R&D (in the same clusters) in its absence. We might then ask whether this firm R&D was likely to follow the same geographic distribution. Here, again, the evidence we will show suggests it would not. The first reason is that the major war equipment suppliers and OSRD industrial R&D performers were generally different (Table E.1). However, that does not rule out that the OSRD industrial R&D performers could

have picked up the slack that OSRD would have left behind.

A second piece of evidence is that industrial and non-industrial OSRD-funded patents were generally in different places. Figure E.3 illustrates this, mapping industrial and non-industrial OSRD patents based on the contractor’s sector (Panels A and B). For visualization purposes we restrict these maps to states east of the Mississippi River, to make them easier to compare side-by-side. The difference in their geographic distribution is—to us—visually apparent.

Figure E.3: Geography of industrial and non-industrial OSRD patents (contract-based definitions; zooming into eastern states for easier viewing)



Notes: Figure maps counties with industrial and non-industrial OSRD-funded patents, using contract-based definitions. Bubble sizes proportional to each county’s total number of OSRD patents. Maps restricted to eastern states (where a large majority of OSRD-funded invention was produced) for ease of visualization.

Regression evidence indicates that these differences are present at the cluster (county x technology) level as well—even conditional on subject matter. Table E.5 regresses clusters’ (level and log) count of industrial OSRD patents on non-industrial OSRD patents, conditional on technology area (NBER category) fixed effects. We also do so for a given cluster’s rank in industrial and non-industrial OSRD patents. The specification we estimate is the following:

$$\text{Industrial OSRD patents}_{ic} = \beta \cdot \text{Nonindustrial OSRD patents}_{ic} + \alpha_c + \varepsilon_{ic}$$

where i indexes counties, c indexes patent categories, and α_c are NBER category FEs. Industrial and non-industrial OSRD patenting appears to negatively correlate in space. This is a natural result, in our view, because fundamentally different work was done by non-industrial and industrial contractors: whereas the former were more involved in basic and applied research, design, and early prototyping, the latter were more heavily engaged in late-stage development and engineering. Oral histories describe how this happened in practice: for example, in one, a Rad Lab researcher

explained this division of labor and the handoffs between them, where “People from General Electric and Westinghouse would come up to the Rad Lab and get the designs ... and then they would go back and [make] modifications” (quotes from oral history with Theodore Saad conducted in 1991 by Andrew Goldstein, available at the IEEE History Center).¹⁹

Table E.5: Clusters’ industrial vs. non-industrial OSRD patents

	OSRD firm patents (def. 1)			OSRD firm patents (def. 2)		
	(1) Count	(2) Log count	(3) Rank	(4) Count	(5) Log count	(6) Rank
Nonfirm OSRD patents	-0.020*** (0.007)			-0.020*** (0.006)		
Log nonfirm OSRD patents		0.053 (0.097)			0.002 (0.100)	
Nonfirm OSRD patents rank			-0.503*** (0.029)			-0.392*** (0.024)
N	690	115	690	690	128	690
R ²	0.07	0.21	0.36	0.07	0.13	0.31

Notes: Table estimates the relationship of industrial and non-industrial OSRD patents at the county level, using contract-based (columns 1-2) and assignee-based (columns 3-4) definitions. In columns (1) and (3) we measure these in levels; in columns (2) and (4) we rank clusters in their number of industrial and non-industrial patents and correlate ranks. All columns include tech. category fixed effects. *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Robust SEs in parentheses.

The evidence we have presented thus far indicates that industrial and non-industrial OSRD-funded R&D was performed in difference places. What about OSRD-funded R&D and war production? If the R&D were physically co-located with production, war equipment suppliers conceivably may have taken up some of this R&D in the same places. To evaluate this question we borrow data from ICPSR 2896 (Haines 2010), which provides county-level war supply contract value, in total and for combat equipment specifically (which comprised $\approx 2/3$ of supply contracts), and which also provides county-level war facilities spending. Because there is no straightforward way to crosswalk product categories in the war supply microdata (from Li and Koustas 2019) to technology areas in the patent data, we will restrict our attention here to county-level analysis (rather than using Li and Koustas (2019)’s data for county-category level analysis).

In Table E.6, we correlate (logged) war supply contracts with county-level measures of OSRD treatment. Columns (1) and (2) estimate total war supply contracts as the outcome; Columns (3) and (4), combat equipment contracts; and Columns (5) and (6), war facilities spending. Odd-numbered columns measure treatment as the log OSRD rate, and even-numbered columns as the log number of OSRD patents, conditional on log total 1941-1948 patents (a related but slightly different specification, which we discuss further in Appendix D).

¹⁹See <https://ethw.org/Oral-History:TheodoreSaad>.

Table E.6: County-level war supply contracts and OSRD treatment intensity

	War supply contracts		Combat equipment		Facilities spending	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(OSRD rate, 1941-48)	-0.921*** (0.138)		-0.828*** (0.159)		-0.524*** (0.116)	
Ln(OSRD patents, 1941-48)		-0.235** (0.118)		-0.213* (0.120)		-0.175 (0.108)
Ln(Patents, 1941-48)		1.315*** (0.102)		1.501*** (0.147)		0.926*** (0.129)
N	160	160	144	144	135	135
R ²	0.27	0.64	0.15	0.55	0.12	0.37

Notes: Table estimates the relationship of county-level log war supply contracts (from ICPSR 2896) and the county-level log OSRD rate (columns 1 and 3) and county-level log OSRD patents (columns 2 and 4). *, **, *** represent significance at the 0.1, 0.05, and 0.01 levels, respectively. Robust SEs in parentheses.

We find that war supply contracts and OSRD activity correlated negatively in space.²⁰ In other words, “war supply places” and “OSRD places” (to borrow the language of Kantor and Whalley 2022) were different. These results are consistent with broader understanding, such as the observation by Garin and Rothbaum (2022) that “Concerns about supply chain security and production bottlenecks led the military to insist that ... plants built to supply the war effort be built outside of established manufacturing hubs in less-congested inland locales”—a departure from the geographic distribution of OSRD activity. The U.S. R&D and manufacturing base were more generally not fully coincident in the late 1930s, contributing to this divide.

Between industrial and non-industrial OSRD R&D having different geographic distributions, and war supply contracts and OSRD R&D also having different geographies, we think this evidence largely rules out counterfactual case (a)(i)—that firms might have replaced OSRD R&D following the same geographic pattern—though not fully, since we have not yet ruled out that in the counterfactual, firms would have set up R&D operations where non-firm OSRD R&D took place. Though difficult to evaluate, this counterfactual also seems unlikely. The MIT Rad Lab, for example, was staffed by researchers who migrated in from around the country, as we show in concurrent research (Gross and Roche 2023). This feature of OSRD—which we have seen for other large OSRD research projects as well—would seem difficult for a firm to reproduce.

Would these firms have instead contracted out R&D, as OSRD did?

This is another alternative we can consider, though it is a pretty intricate counterfactual which would require a lot of pieces to come together: firms would need to not only finance the R&D, but also find partners in the same places, negotiate contracts, coordinate efforts, and more. Based on our view of OSRD through recent work (e.g., Gross and Sampat 2022c), we do not think this is a likely counterfactual. Mobilizing U.S. science and technology for war was a complex policy and organizational problem, and for many reasons (including reasons related to OSRD’s breadth, complexity, and centralized administration) not something that atomistic firms would be able to

²⁰We find similar, albeit statistically weaker, patterns for facilities spending.

reproduce on their own—not least without the weight of the federal government, let alone the relationships and administrative diplomacy of Vannevar Bush.

On these grounds, we would rule out counterfactual case (a)(ii).

Would another agency have funded this R&D?

The last alternative we can consider for a no-OSRD counterfactual is that other government agencies (like the War Department or the Navy) might have served the same functions. For the purposes of this paper, this would largely be the same story, just by other means. But even this counterfactual, in our view, is unlikely. The War and Navy Departments were not set up to do what OSRD did—at least not as well. OSRD enjoyed a different and much more flexible institutional arrangement, operating somewhat outside of the scope of traditional government, with relatively little attachment to existing structures, little internal politics or bureaucracy, and little external red tape. Most importantly, OSRD conducted no intramural research—whereas the military had active intramural research programs (e.g., at the Army Signal Corps and Naval Research Laboratory), which were the locus of its R&D activity during the war, and their wartime jockeying for influence suggests would have continued to be even if given resources with which to scale. OSRD’s distinctive position allowed it to engage a considerably different (and then-novel) administrative model in the way of R&D policy (see Gross and Sampat 2022c). Moreover, OSRD was not itself distracted by the operation of the war effort, but rather mobilized alongside it.

We thus think history suggests against counterfactual case (b).

The evidence in review

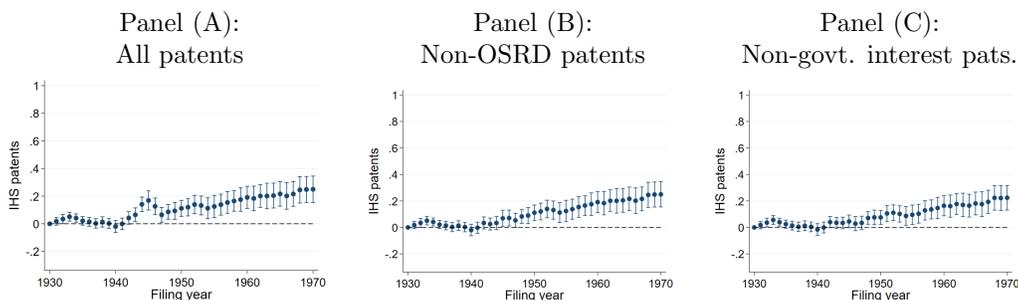
Collectively, we think this evidence suggests against the hypothesis that OSRD’s R&D activity and impacts would have been reproduced by other means in a no-OSRD counterfactual. This includes a demand-driven version the World War II shock, where military demand spills in R&D in the same technologies and regions. This is not to say that all of OSRD’s work or outputs would have been absent in its absence, but rather that the scale, scope, geography, and impacts of any alternative arrangement would almost necessarily have been different.

If nothing else, we can observe that military orders followed the technology, rather than vice versa. This R&D was ex-ante uncertain, and the form or extent of its military applications often not yet known. OSRD’s work was underway long before orders were placed or ideas proven to work—and well before the U.S. entered the war in December 1941.

F Alternative outcome measures

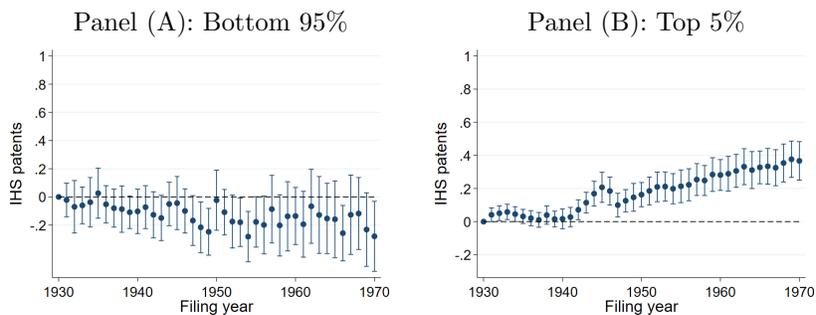
This appendix reproduces our main results on the postwar agglomeration of U.S. invention, replacing logged outcome measures with the inverse hyperbolic sine (IHS) transformation. Figures F.1 to F.5 reproduce Figures 4 to 8 from the body of the paper.

Figure F.1: Effects of OSRD on cluster patenting, 1930-1970



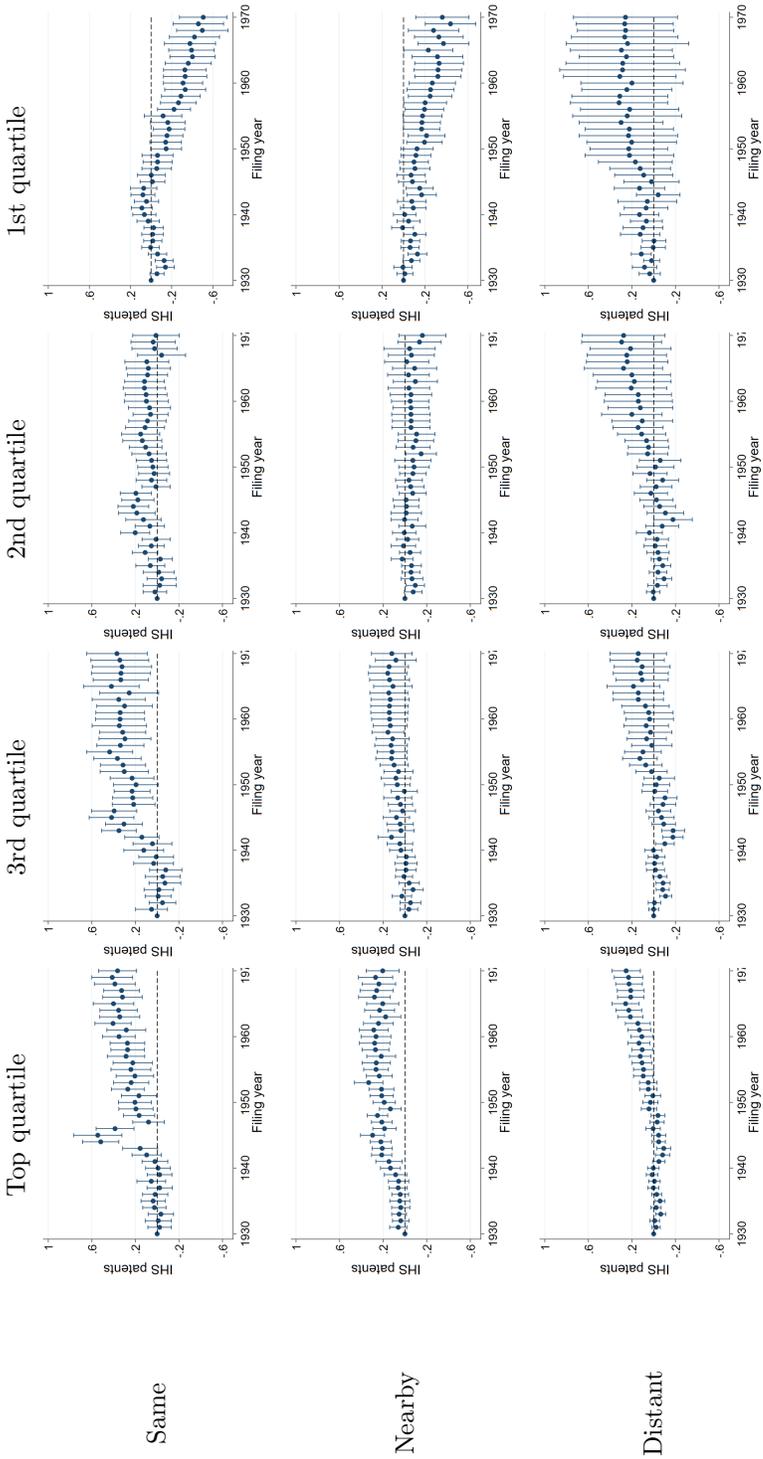
Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting. The independent variable measures the IHS fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

Figure F.2: Effects of OSRD on cluster patenting, for clusters in counties in the bottom 95% versus top 5% of 1930s patenting



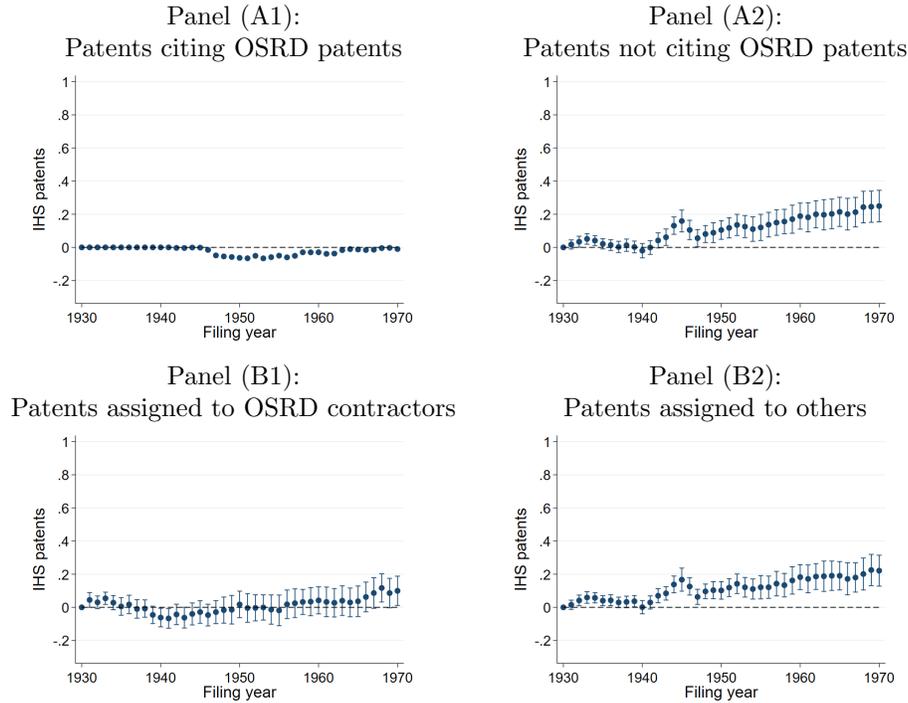
Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting, for counties in the bottom 95% and top 5% of 1930s patenting (i.e., existing technology clusters). Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

Figure F.3: Effects of OSRD on cluster patenting, 1930-1970, cross-technology area spillovers
 horserace regression of treatment in (i) same technology area, (ii) nearby technology areas, (iii) more distant technology areas



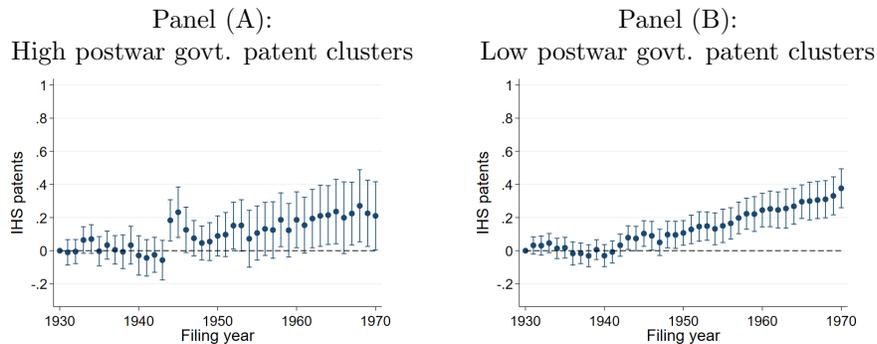
Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patenting. The independent variable measures the quartile of treatment intensity, conditional on treatment (the fraction of U.S. patents in each county-category between 1941-1948 which were OSRD-funded, conditional on any), of three types: (i) in the given county-category (top row); (ii) in the same county and proximate technology categories (same 1-digit NBER category, per Hall et al. (2001); middle row); and (iii) in the same county and more distant technology categories (other 1-digit NBER categories; bottom row). Parameters across all panels are estimated jointly (in one regression) relative to a reference group of county-categories without any OSRD patents. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

Figure F.4: Effects not explained by OSRD's direct impacts on local invention



Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category (i) patents citing versus not citing OSRD patents, and (ii) patents assigned to OSRD contractors versus others, as an exploration of the direct impacts of OSRD on postwar invention in the treated clusters. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

Figure F.5: Effects not explained by sustained government investment in local invention

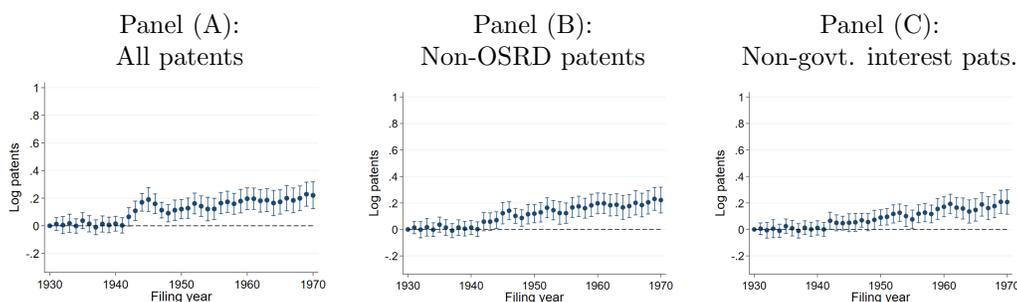


Notes: Figure shows annual estimates of the effects of the OSRD shock on county-category patents in county-categories with above and below median postwar (1950-1969) government-funded patent rates, as an exploration of the role of sustained public R&D investment as an explanation for persistence. Error bars represent 95% confidence intervals, computed from SEs clustered at the county level.

G CBSA-level analysis

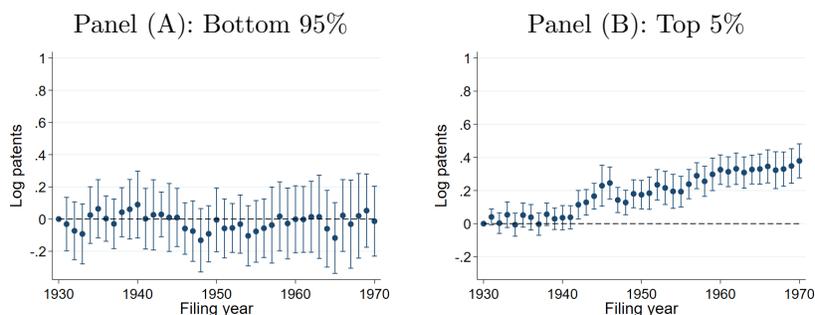
This appendix reproduces our main results on the postwar agglomeration of U.S. invention, aggregating counties up to core-based statistical areas (CBSAs), geographic units consisting of one or more counties around an urban center tied together by modern commuting patterns. Though we believe counties to be a preferred unit of analysis for the period we study—most of which preceded the postwar development of regional transportation systems—CBSA-based robustness checks can test whether the results are sensitive to the choice of geographic units. Figures G.1 to G.5 reproduce Figures 4 to 8 from the body of the paper at the CBSA-category level.

Figure G.1: Effects of OSRD on cluster patenting, 1930-1970



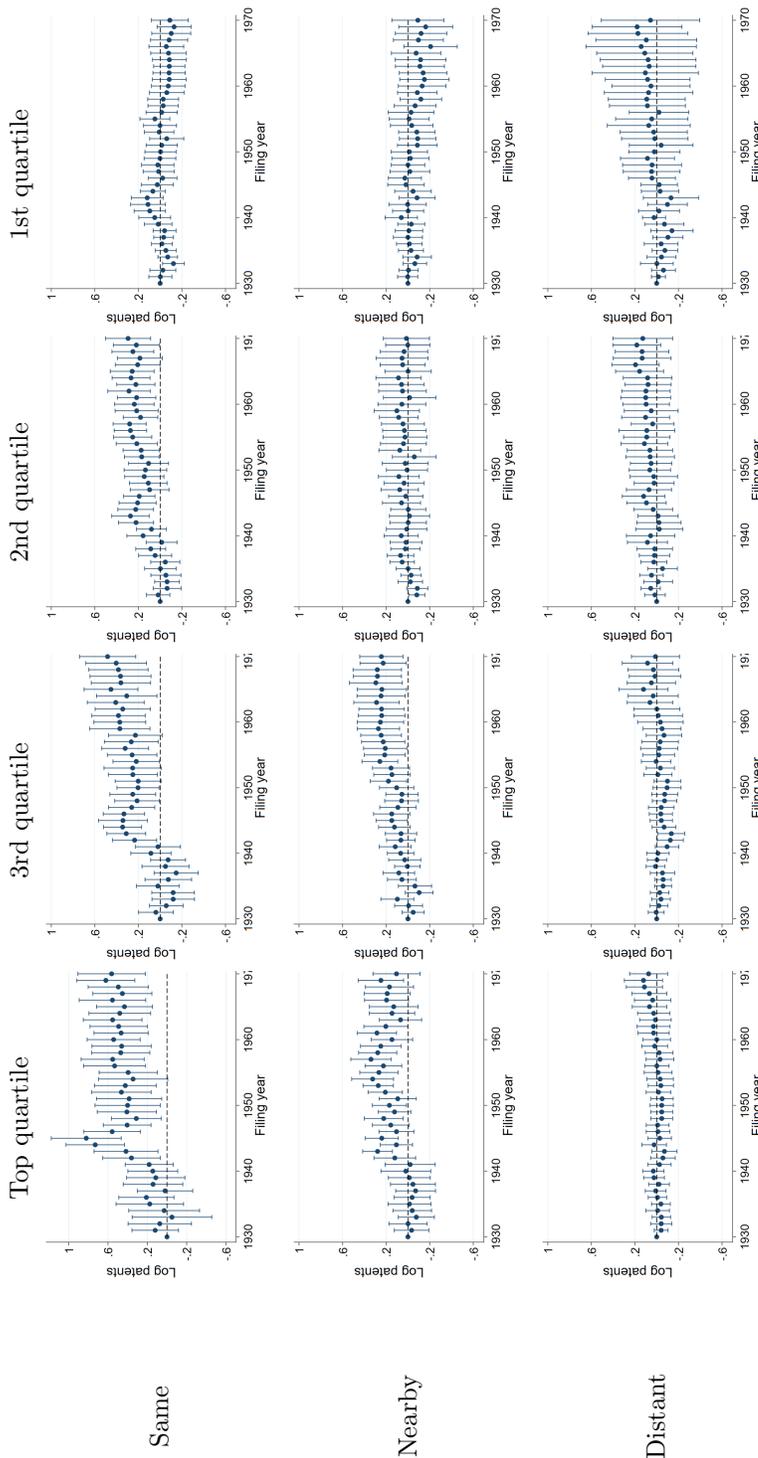
Notes: Figure shows annual estimates of the effects of the OSRD shock on CBSA-category patenting. The independent variable measures the log fraction of U.S. patents in each CBSA-category between 1941-1948 which were OSRD-funded. Error bars represent 95% confidence intervals, computed from SEs clustered at the CBSA level.

Figure G.2: Effects of OSRD on cluster patenting, for clusters in CBSAs in the bottom 95% versus top 5% of 1930s patenting



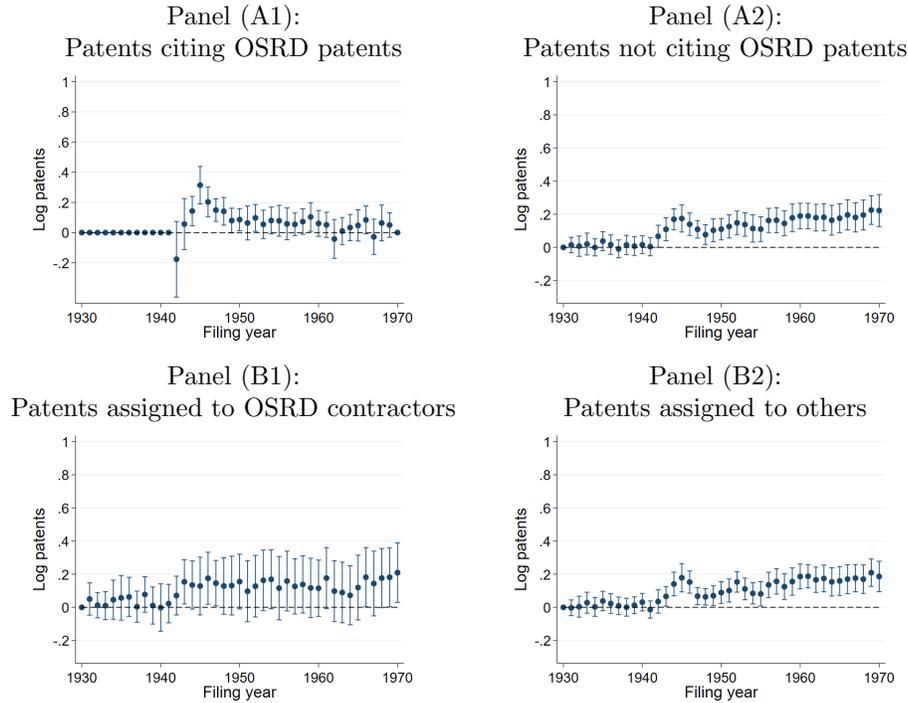
Notes: Figure shows annual estimates of the effects of the OSRD shock on CBSA-category patenting, for CBSAs in the bottom 95% and top 5% of 1930s patenting (i.e., existing technology clusters). Error bars represent 95% confidence intervals, computed from SEs clustered at the CBSA level.

Figure G.3: Effects of OSRD on cluster patenting, 1930-1970, cross-technology area spillovers
 horsrace regression of treatment in (i) same technology area, (ii) nearby technology areas, (iii) more distant technology areas



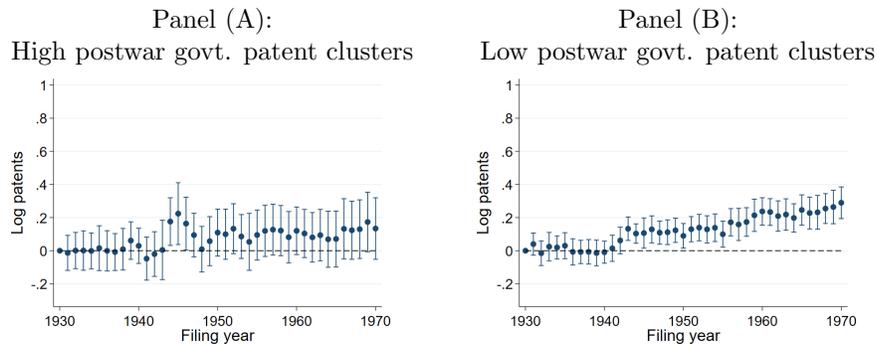
Notes: Figure shows annual estimates of the effects of the OSRD shock on CBSA-category patenting. The independent variable measures the quartile of treatment intensity, conditional on treatment (the fraction of U.S. patents in each CBSA-category between 1941-1948 which were OSRD-funded, conditional on any), of three types: (i) in the given CBSA-category (top row); (ii) in the same CBSA and proximate technology categories (same 1-digit NBER category, per Hall et al. (2001); middle row); and (iii) in the same CBSA and more distant technology categories (other 1-digit NBER categories; bottom row). Parameters across all panels are estimated jointly (in one regression) relative to a reference group of CBSA-categories without any OSRD patents. Error bars represent 95% confidence intervals, computed from SEs clustered at the CBSA level.

Figure G.4: Effects not explained by OSRD's direct impacts on local invention



Notes: Figure shows annual estimates of the effects of the OSRD shock on CBSA-category (i) patents citing versus not citing OSRD patents, and (ii) patents assigned to OSRD contractors versus others, as an exploration of the direct impacts of OSRD on postwar invention in the treated clusters. Error bars represent 95% confidence intervals, computed from SEs clustered at the CBSA level.

Figure G.5: Effects not explained by sustained government investment in local invention



Notes: Figure shows annual estimates of the effects of the OSRD shock on CBSA-category patents in CBSA-categories with above and below median postwar (1950-1969) government-funded patent rates, as an exploration of the role of sustained public R&D investment as an explanation for persistence. Error bars represent 95% confidence intervals, computed from SEs clustered at the CBSA level.

Appendix references

- Andrews, Michael. 2019. *Comparing historical patent datasets*. Working paper.
- Bergeaud, Antonin, Cyril Verluise et al. 2022. *A new dataset to study a century of innovation in Europe and in the U.S.*. CEP Discussion Paper No. 1850.
- Bergeaud, Antonin and Cyril Verluise. 2022. *PatentCity: A dataset to study the location of patents since the 19th century*. DOI: 10.7910/DVN/PG6THV Harvard Dataverse [publisher].
- Berkes, Enrico. 2018. *Comprehensive Universe of U.S. Patents (CUSP): Data and facts*. Working paper.
- Chen, Jiafeng and Jonathan Roth. 2023. *Log-like? Identified ATEs defined with zero-valued outcomes are (arbitrarily) scale-dependent*. Working paper.
- Derwent Innovation. 2018. *Patent database*.
- Dun & Bradstreet. 1980. *Private company listings*. Obtained through Mergent Data Explr.
- Eckert, Fabian, Teresa C Fort, Peter K Schott, and Natalie J Yang. 2021. *Imputing missing values in the U.S. Census Bureau’s County Business Patterns*. NBER Working Paper No. 26632.
- EPO. 2017. *PATSTAT database*. Version 5.10: Autumn 2017 edition.
- Fleming, Lee, Hillary Greene, G Li, Matt Marx, and Dennis Yao. 2019. “Government-funded research increasingly fuels innovation,” *Science*, Vol. 364, No. 6446, pp. 1139–1141.
- Forstall, Richard L. 1996. *Population of states and counties of the United States: 1790 to 1990*. U.S. Bureau of the Census, Washington: Government Printing Office.
- Garin, Andy and Jonathan Rothbaum. 2022. *The long-run impacts of public industrial investment on regional development and economic mobility: Evidence from World War II*. Working paper.
- Gross, Daniel P. 2023. “The hidden costs of securing innovation: The manifold impacts of compulsory invention secrecy,” *Management Science*, Vol. 69, pp. 2318–2338.
- Gross, Daniel P and Maria P Roche. 2023. *Coordinated R&D programs and the creation of new industries*. Working paper.
- Gross, Daniel P and Bhaven N Sampat. 2022. *The World War II crisis innovation model: What was it, and where does it apply?*. Forthcoming at Research Policy.
- Haines, Michael R. 2010. *Historical, demographic, economic, and social data: The United States, 1790-2002*. Inter-university Consortium for Political and Social Research [distributor], DOI: 10.3886/ICPSR02896.v3.
- Hall, Bronwyn H, Adam B Jaffe, and Manuel Trajtenberg. 2001. *The NBER patent citation data file: Lessons, insights and methodological tools*. NBER Working Paper No. 8498.
- Kantor, Shawn and Alexander Whalley. 2022. *Moonshot: Public R&D and growth*. Working paper.
- Li, Zhimin and Dmitri Koustas. 2019. “The long-run effects of government spending on structural change: Evidence from Second World War defense contracts,” *Economics Letters*, Vol. 178, pp. 66–69.
- Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2022. *IPUMS National Historical Geographic Information System*. IPUMS, DOI: 10.18128/D050.V17.0.
- Marco, Alan C, Michael Carley, Steven Jackson, and Amanda Myers. 2015a. *The USPTO Historical Patent Data Files: Two centuries of innovation*. SSRN working paper, available at <https://ssrn.com/abstract=2616724>.
- Marco, Alan C, Amanda Myers, Stuart Graham, Paul D’Agostino, and Kirsten Apple. 2015b. *The USPTO Patent Assignment Dataset: Descriptions and analysis*. SSRN working paper, available at <https://ssrn.com/abstract=2849634>.
- Mullahy, John and Edward C Norton. 2022. *Why transform Y? A critical assessment of dependent-variable transformations in regression models for skewed and sometimes-zero outcomes*. NBER Working Paper No. 30735.
- NAS. 1963. *Doctorate production in United States universities: 1920-1962*. National Academy of Sciences–National Research Council Publication No. 1142.
- NSF. 2022. *Survey of Earned Doctorates*. Data obtained from the National Center for Science and Engineering Statistics (NCSES) Table Builder at <https://ncesdata.nsf.gov/builder/sed>.
- Petralia, Sergio, Pierre-Alexandre Balland, and David Rigby. 2016a. *HistPat dataset*. Harvard Dataverse [publisher], DOI: 10.7910/DVN/BPC15W.
- Petralia, Sergio, Pierre-Alexandre Balland, and David L Rigby. 2016b. “Unveiling the geography of historical patents in the United States from 1836 to 1975,” *Scientific data*, Vol. 3, No. 1, pp. 1–14.

- Rai, Arti K and Bhaven N Sampat. 2012. "Accountability in patenting of federally funded research," *Nature Biotechnology*, Vol. 30, No. 10, pp. 953–956.
- Romer, Paul M. 1990. "Endogenous technological change," *Journal of Political Economy*, Vol. 98, No. 5, Part 2, pp. S71–S102.
- Seamans, Robert, Eunhee Sohn, and Daniel Sands. 2018. *Technological opportunity and the locus of innovation: Airmail, aircraft, and local capabilities*. Working Paper.
- Stewart, Irvin. 1948. *Organizing scientific research for war: The administrative history of the Office of Scientific Research and Development*. Boston: Little, Brown, and Company.
- U.S. Census Bureau. 1980. *County Business Patterns (CBP)*. Data obtained from the National Archives and Records Administration, available at <https://catalog.archives.gov/id/873800>.
- USPTO. 1977. *Technology assessment & forecast*.
- . 2007. *Concordance between the U.S. Patent Classification System and product fields based on the 1972 Standard Industrial Classification*. Available at https://www.uspto.gov/web/offices/ac/ido/oeip/taf/data/sic_conc/2007_diskette/.
- Watson, Donald Stevenson and Mary A Holman. 1964. "Patents from government-financed research and development," *Patent, Trademark, and Copyright Journal of Research and Education*, Vol. 8, p. 199.
- Weitzman, Martin L. 1998. "Recombinant growth," *Quarterly Journal of Economics*, Vol. 113, No. 2, pp. 331–360.