GERMAN JEWISH ÉMIGRÉS AND U.S. INVENTION

Petra Moser
Alessandra Voena
Fabian Waldinger

ONLINE APPENDIX

DATA APPENDIX

1. Inventor-level Data on Patenting

We implement a four-step process of data collection and cleaning to construct inventor-level data on changes in patenting for 166 classes of chemical inventions, using records from Google's *Patent Grant Optical Character Recognition (OCR) Text (1920-1979) Database*. First, we extract inventors from the OCR database using a Pearl script. We then clean the data by correcting common errors in Google's OCR and by removing substrings that do not contain the actual inventor. In the next step we create an algorithm that separates inventors using information on 3,439 common first names from the U.S. Censuses and Social Security records. Lastly, we assign unique identifiers based on Levenshtein (1966) distances. In this section, we describe each of these four steps in more detail.

Step 1: Pearl Script to Identify Inventors in Google's Patent Grant Optical Character Recognition (OCR) Text (1920-1979) Database

The inventor data come from Google's *Patent Grant Optical Character Recognition (OCR) Text (1920-1979)* database. We program a Pearl script to search for the inventor in the full text of each patent document.

To optimize the quality of our inventor data we adjust the Pearl script to reflect changes in the layout of the patent document. Until 1953 the inventor name appears in two sections of the patent document: near the title of the invention and at the end of the document. We collect both and use the string that has a higher probability of identifying the inventor for patents issued until 1953 (more details below).

The other major change of reporting inventors occurred in 1933 (after patent number 1,920,164) when USPTO switched from reporting inventors at the beginning of the patent document in upper-case letters (e.g. "ARNOLD WEISSBERGER") to lower-case (e.g. "Arnold Weissberger").

To obtain inventor names found near the title, the code searches for relevant substrings of the marker "United States Patent Office" (non-case sensitive). After identifying the marker we extract the next 10 lines of the OCR document; they usually contain the title of the invention and names of all inventors. As the title is usually spelled in capital letters we use regular expressions to cut any consecutive strings of capitalized letters. We then concatenate all remaining strings and use commas to delimit the resulting string that is usually formatted "inventor name, geographic location of inventor, assignee".

To obtain inventor names that appear at the end of the patent document, we look for the marker "BRS DOCUMENT" which indicates the end of a patent document within the OCR. We then take the 10 lines that precede this marker. To isolate inventors from substrings containing other information, our Pearl code removes lines that only contain spaces. It also removes lines that contain the strings "AISD" (assigned date), "CCOR", "CCXR" (classes), "ISY" (assigned year), and consecutive capitalized letters. As above, we then concatenate all remaining lines and separate inventors with commas.

Step 2: Cleaning Code

a. Remove substrings that do not contain the inventor

In many cases the output of step 1 contains just the correct inventor(s) for each patent. Sometimes, however, the output contains additional substrings that do not identify inventors. Additional substrings can be part of the output because markers such as "United States Patent Office" or "assignor" are often misspelled in Google's OCR data. To isolate the inventor from other text we therefore search for regular patterns that indicate inventors and discard other parts of the inventor string. The following list gives an overview of our cleaning:

1) If "United States Patent Office" is misspelled in Google's OCR data the inventor string contains misspelled versions of "United States Patent Office". We manually identify more than 1,100 substrings with misspelled versions of "United States Patent Office" and remove them.
2) Sometimes the OCR adds additional letters after "United States Patent Office" (which do not describe the inventors). The resulting string therefore contains individual letters at the beginning of the string followed by a large number of blanks before the actual inventors are listed. We therefore cut individual letters followed by large numbers of blanks from the inventor string.
3) If "assignor" is misspelled in Google's OCR data the inventor string contains misspelled versions of "assignor". We manually identify 48 substrings with misspelled versions of "assignor" and remove them.
4) In some cases the inventor string includes the beginning of the description of the invention. We therefore remove everything after "This invention" and 113 misspelled variations of "This invention".

5) Similarly we remove 27 misspelled versions of “application filed” from the inventor string.
6) In early versions of the patent layout the inventor’s place of residence is marked with “of” e.g. “Ernst Berl, of Darmstadt, Germany”. In some cases the inventor string contains the inventor’s place of residence and we therefore cut “of” (and 32 misspelled versions of “of”) plus the following word from the inventor string.
7) We remove substrings that include the name of a U.S. state, e.g. “California” and 83 misspelled versions of state names (exceptions: Virginia and Georgia that can also be inventor names).
8) We remove 339 substrings that include U.S. cities such as “Cleveland” and misspelled versions of cities e.g. “Clev6land”.
9) We remove substrings that include foreign countries such as “Germany” or “France”.
10) Patents with patent numbers higher than 1,920,164 use upper case spelling for the initial and lower-case spelling for the rest of the inventor (e.g. “Arnold Weissberger”). Substrings with consecutive upper case letters do thus not identify the inventor for patent numbers > 1,920,164. We therefore cut substrings containing only upper-case letters for patents with patent numbers > 1,920,164.
11) If the inventor string only includes lower case letters we set the inventor to missing as inventor names always contain upper case letters. We manually identify exceptions where the inventor string only contains lower case letters but still includes a large part of the inventor and keep them in the data.

b. Correct Common Misspellings

Our cleaning code also corrects common misspellings that originate from the OCR process. The following list gives an overview of the most important corrections:

1) T) → D exceptions manually corrected
2) I) → D exceptions manually corrected
3) !-I → H
4) I-I → H
5) I-1 → H
6) II → H exceptions manually corrected
7) IT → H exceptions manually corrected
8) :-I → H
9) 1-1 → H

10) A/I → M
11) IYI → M
12) lYl → M
13) lYI → M
14) 1VI → M exceptions manually corrected
15) TYI → M exceptions manually corrected
16) 1VL → M
17) lV[→ M
18) I\' → N
19) 1\T → N
20) I\T → N
21) !\' → N
22) 0. → O.
23) P. → R if not a middle initial, exceptions manually corrected
24) P, → R exceptions manually corrected
25) lt → R if inventor should be upper case (if patent number <= 1920164)
26) .T → J at the beginning of the inventor string
27) ,T → J at the beginning of the inventor string
28) VV → W
29) NV → W if inventor should be lower case (if patent number > 1920164)
30) 13 → B if patent number < 1920165
31) 33: → H if patent number < 1920165
32) 33[→ H if patent number < 1920165
33) 33, → R if patent number < 1920165
34) 331 → H if patent number < 1920165
35) 33 → B if patent number < 1920165
36) 3) → D if patent number < 1920165
37) !Q → D if patent number < 1920165
38) XANN → MANN
39) XOND → MOND

c. Correct misspelled first names

We also correct a total of 1,530 misspelled versions (e.g. “Jos@ph” instead of “Joseph”) for the following first names: Abraham, Adolf, Adolph, Alan, Albert, Alexander, Alexis, Alfonso, Alfred, Allen, Andre, Andrew, Antony, Archibald, Arnold, Arthur, August, Barbara, Barney, Benjamin, Bernhart, Bertolo, Bestor, Bob, Brentano, Bruce, Carl, Carlo, Carlton, Carroll, Cecil, Charles, Clarence, Claude, Conrad, Craig, Daniel, David, Dayton, Delbert, Donald, Douglas, Earl, Earle, Edgar, Edmund, Edvard, Edward, Edwin, Elisabeth, Emma, Emil, Ernest, Ernst, Erwin, Esther, Eugene, Everett, Felix, Fernand, Fernando, Forrest, Francis, Frank, Franklin, Franz, Fred, Frederick, Fredrich, Fremont, Friedrich, Fritz, Garry, Gebhard, Geoffrey, George, Gilbert, Granville, Gustave, Hamilton, Hans, Harold, Harries, Harrison, Harry, Harvey, Helmut, Henri, Henrietta, Henry, Herbert, Herman, Hermann, Hildegard, Horace, Howard, Hubertus, Hugo, Jacob, Jagan, James, Jesse, Johan, Johannes, John, Jose, Josef, Joseph, Joshua, Judson, Julius, Karl, Karl-Heinz, Karoly, Kazimer, Larry, Lawrence, Lee, Lemuel, Leon, Leonard, Lewis, Louis, Ludwig, Major, Marc, Margaret, Marie, Marion, Mark, Marshall, Marta, Martin, Marvin, Matthew, Matthias, Maurice, Max, Maximilian, Melville, Melvin, Michael, Michele, Mildred, Milton, Nathaniel, Nelson, Nils, Noel, Norman, Oliver, Oswald, Patrick, Paul, Peter, Peyton, Philip, Pierre, Ralph, Ray, Raymond, Reginald, Rene, Reynold, Richard, Robert, Roland, Royce, Rudolf, Rudolph, Russell, Ryan, Samuel, Seth, Shirl, Sidney, Simon, Solomon, Spencer, Stanley, Starry, Stephen, Stewart, Taylor, Theodore, Thomas, Vernon, Victor, Viktor, Vincent, Wallace, Walter, Werner, Wilford, Wilfred, Wilhelmus, William, Willem.

d. Choose Between Inventors If Available Between Different Parts of the Patent Document

As mentioned above, until 1953 the patent document lists the inventor in two different places: near the beginning of the document and at the end. After patent number 2,672,389 the inventor can only be easily identified near the beginning of the document. If the inventor is listed in two places in the document the information at the beginning of the document is usually of higher quality because the entry at the end sometimes contains witnesses or patent examiners.

For each patent we therefore choose the inventor as follows:

1) We first use the inventor listed at the beginning of the patent document.
2) If the inventor from the beginning of the document is missing, we use the inventor from the end of the document if the patent number is smaller than 2,672,389 (after

this patent number the string from the end of the document does not include the correct inventor).

3) If the inventor from the beginning of the document contains numbers or the characters ‘:’ ‘@’ ‘=’ ‘&’ ‘)’ ‘!’, we use the inventor from the end of the document if the patent number is smaller than 2,672,389 and if the inventor from the end of the document is a string longer than 5 characters and includes at least two words.
4) If the inventor from the beginning of the document is a string with less than 7 characters, we use the inventor from the end of the document if the patent number is smaller than 2,672,389 and if the inventor from the end of the document is a string longer than 5 characters and includes at least two words.
5) If the inventor from the beginning of the document does not contain spaces, we use the inventor from the end of the document if the patent number is smaller than 2,672,389 and if the inventor from the end of the document is a string longer than 5 characters and includes at least two words.
6) If the inventor from the beginning of the document contains lower case characters before the inventor is reported in lower case (i.e. patent number < 1,920,164), we use the inventor from the end of the document if the patent number is smaller than 2,672,389 and if the inventor from the end of the document is a string longer than 5 characters and includes at least two words.

Step 3: Separating Inventors

In our third step we separate inventors. This addresses the following issues:

1) After the previous cleaning steps the data contain all inventors in one string, even if a patent was filed by multiple inventors.
2) Even after extensive cleaning in step 2, the inventor string may still include substrings that do not identify inventors.

The following procedure addresses both of these issues.

Separate inventors if they are separated by “and”

We first separate inventors that are separated by “and”: e.g. “Ernst Zerner and Marcel Gradsten”. We also identify 86 misspelled versions of “and” and separate inventors accordingly.

Identify inventors that do not need separation

Whenever the inventor string contains only two words, we treat such words as first name and last name of an inventor, and we do not proceed with further separation. Similarly, if the string contains one word followed by one or two initials and by another word, we do not proceed with any further separation. All other strings are examined through a process described below.

Separate other inventors using data on 3,439 first names

Strings that contain more than one inventor are separated with an algorithm that uses 3,439 common female and male names from U.S. Censuses (1920, Ruggles et. al. 1997) and Social Security records (1900-1999, Shackleford 2000) to isolate individual inventors.[1] The algorithm proceeds as follows: we first search for a common first name starting at the beginning of a string. If the first name is not separated from the following middle name or family name, we introduce a space after the first name to isolate it from the rest of string.

We then identify first and middle names in the inventor string. This algorithm compares each word in the string to each of the 3,439 first names in our list. We classify a word as a first or middle name if it matches one of the common names with a Levenshtein distance that is less or equal to 25%. We identify individual inventors as substrings with the pattern first name, middle name, plus one unmatched word, or with the pattern first name plus one unmatched word.[2] This process yields 842,068 unique inventor names. Some of these names may be misspelled (such as "Arnold Weissberge" instead of "Arnold Weissberger"). In the next stage described below we address these misspellings.

Step 4: Generate unique inventor identifiers

In the last stage, we use Levenshtein distances to construct a unique inventor identifier, which allows for misspellings of the inventors' name. As our algorithm processes about a million inventor strings we assign inventor identifiers in two steps. We first group inventors by first names and then use Levenshtein distances to assign unique inventor identifiers within those groups.

First, inventors are grouped by their first names. We use the list of common first names described above and find all inventors that share the same first name. To allow for

[1] First names available at http://www.galbithink.org/names/us200.htm.
[2] As our list of common first names also includes initials we also identify inventors who report a first name, middle initial, and a last name.

remaining misspellings of the first name the first name groups are based on a maximum Levenshtein distance of 25%, i.e. Arnol is in the same group as Arnold (the normalized Levenshtein distance of the two strings is: 1/7 = 14.3 percent).

We then use the STATA *strgroup* command within each first name group of inventors to generate unique identifiers for strings that have a Levenshtein distance of 20% or lower.[3] E.g. in the group of all inventors with the first name Arnold (or Arnol, or other similar first names) we generate a unique identifier if the Levenshtein distance between two strings is less than 20% ("Arnold Weissberger" will be assigned the same identifier as "Arnold Weissberge").

APPENDIX REFERENCES

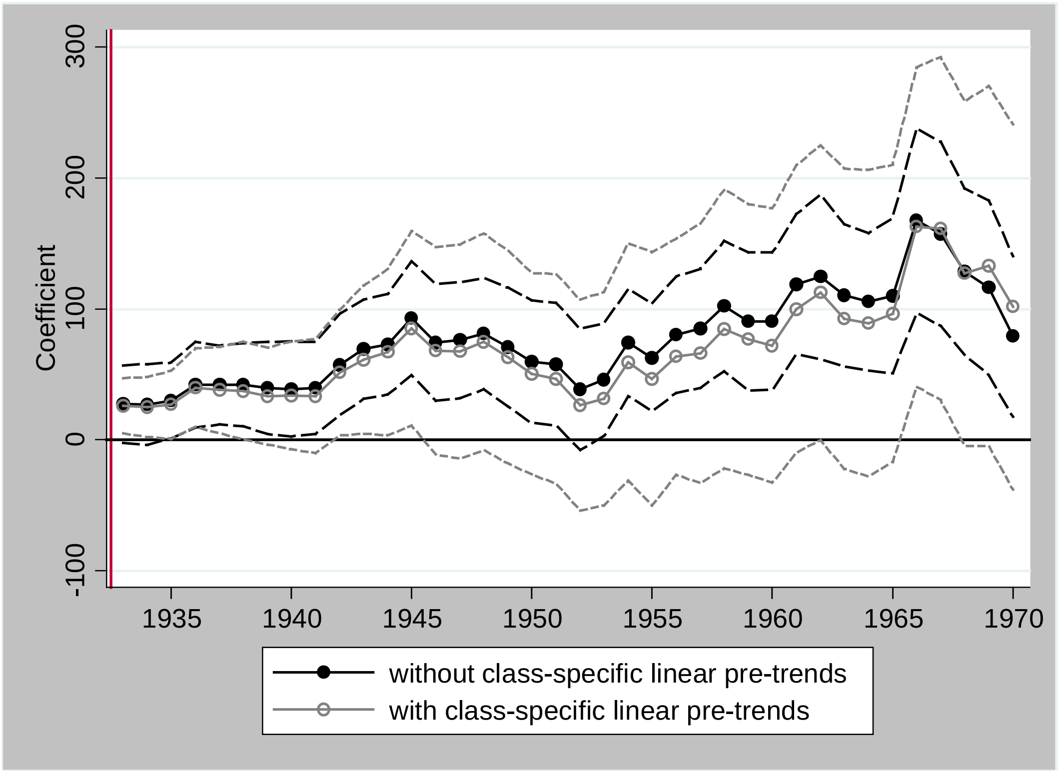
Levenshtein Vladimir. 1966. "Binary codes capable of correcting deletions, insertions, and reversals". Soviet Physics Doklady 10: 707–10.

Ruggles, Steven, and Matthew Sobek et. al., Integrated Public Use Microdata Series: Version 2.0 (Minneapolis: Historical Census Projects, University of Minnesota, 1997).

Shackleford, Michael W, A.S.A., "Name Distributions in the Social Security Area," Social Security Administration, Office of the Chief Actuary, *Actuarial Note Number 139*, originally published June 1998 (updated Oct. 2000).

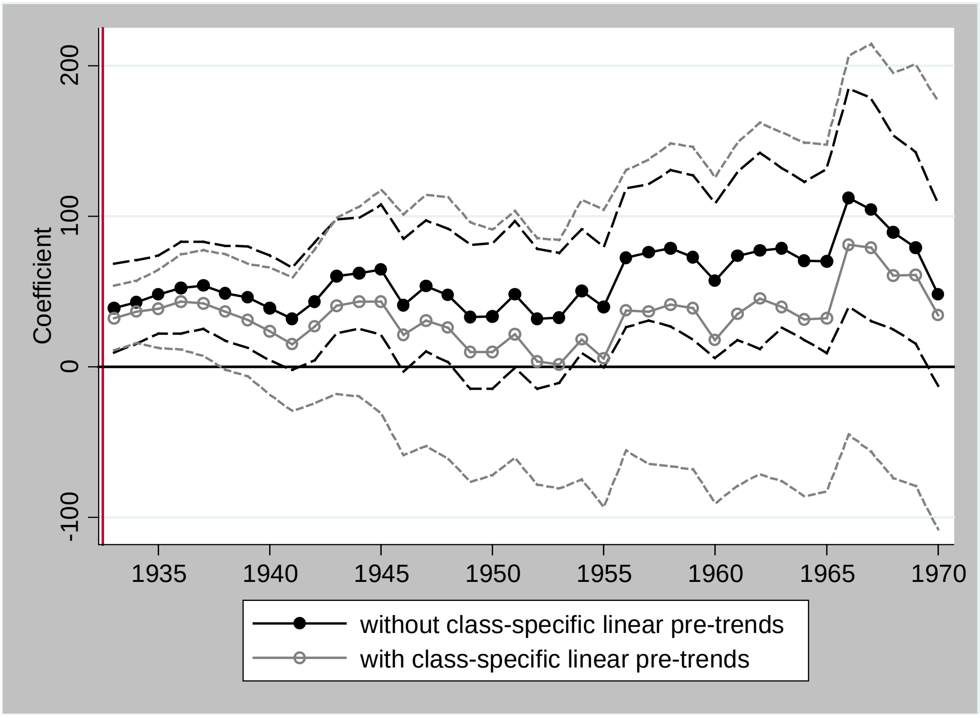
[3] The STATA strgroup command by Julian Reif is available at http://ideas.repec.org/c/boc/bocode/s457151.html.

FIGURE A1 – YEAR-SPECIFIC OLS ESTIMATES
CONTROLLING FOR CLASS-SPECIFIC LINEAR PRE-TRENDS
U.S. PATENTS PER YEAR IN RESEARCH FIELDS OF ÉMIGRÉS



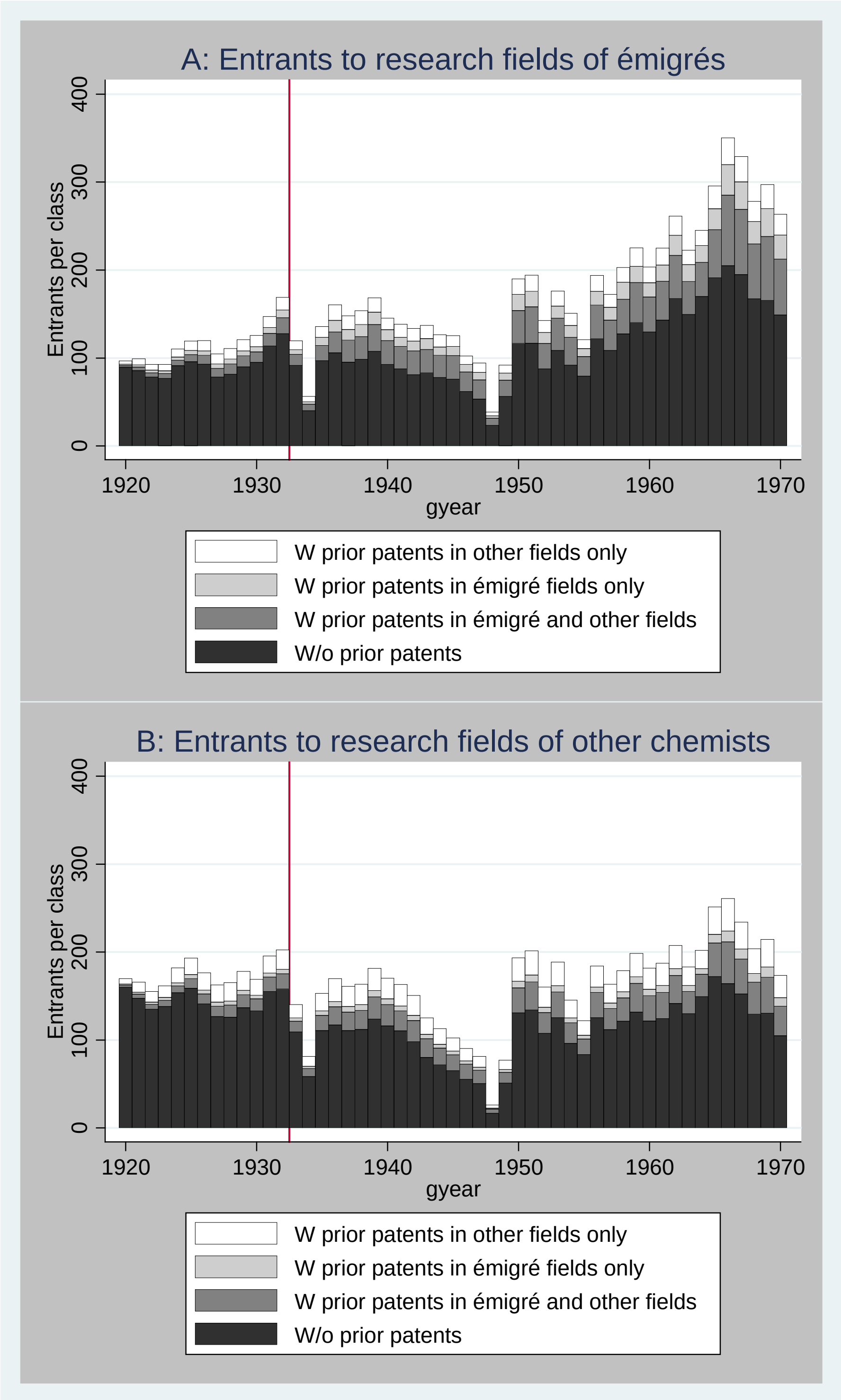
Notes: Time-varying estimates without class-specific linear pre-trends estimate $ß_t$ in the regression *Patents by U.S. inventors*$_{c,t}$ = $\alpha_0 + \sum_{t=1933}^{1970} ß_t \, émigré \; class_c \cdot \; year_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$ where $year_t$ is a set of dummies for every year between 1933 and 1970. Time-varying estimates with class-specific linear pre-trends report coefficients $ß_\tau$ in the regression *Patents by U.S. inventors*$_{c,t}$ = $\alpha_0 + \sum_{\tau=1933}^{1970} ß_\tau \, émigré \; class_c \cdot \; year_\tau \; + \eta_c \cdot t + \delta_t + f_c + \upsilon_{c,t}$. In both specifications, the dependent variable measures U.S. patents issued to U.S. inventors per class and year. Patents by émigré chemists are excluded from the counts of U.S. inventors. The variable émigré $class_c$ equals 1 for research fields of émigrés, defined at the level of 60 classes that include at least one patent between 1920 and 1970 by a German or Austrian émigré to the United States. The control group consists of research fields of other German chemists, defined at the level of 106 USPTO classes that include at least one patent between 1920 and 1970 by another German chemist but include no patents by émigrés. Years between 1920 and 1932 are excluded to estimate pre-trends. Standard errors are clustered at the level of research fields (166 classes).

FIGURE A2 – YEAR-SPECIFIC ESTIMATES OF DIFFERENTIAL CHANGES IN PATENTING FOR RESEARCH FIELDS WITH PRE-1933 PATENTS BY DISMISSED GERMAN CHEMISTS CONTROLLING FOR CLASS-SPECIFIC LINEAR TRENDS (REDUCED FORM)



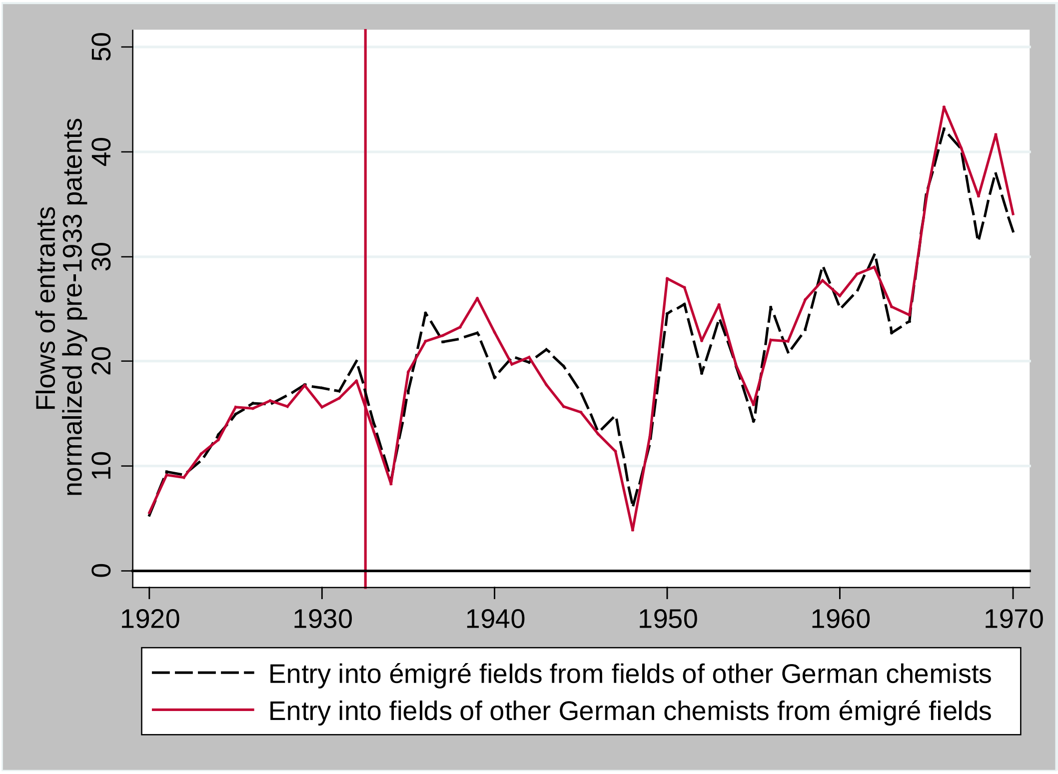
Notes: Time-varying estimates without class-specific linear pre-trends estimate $ß_t$ in the regression *Patents by U.S. inventors*$_{c,t}$ = $\alpha_0 + \sum_{t=1933}^{1970} ß_t\, pre-1933\ dismissed\ class_c \cdot year_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$ where $year_t$ is a set of dummies for each year between 1933 and 1970. Time-varying estimates with class-specific linear pre-trends report coefficients $ß_\tau$ in the regression *Patents by U.S. inventors*$_{c,t}$ = $\alpha_0 + \sum_{\tau=1933}^{1970} ß_\tau pre-1933\ dismissed\ class_c \cdot year_\tau + \eta_c t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$. In both specifications, the dependent variable measures U.S. patents issued to U.S. inventors per class and year. Patents by émigré chemists are excluded from the counts of U.S. inventors. The variable *pre-1933 dismissed class*$_c$ equals 1 for pre-dismissal research fields of dismissed chemists, defined at the level of 48 classes in which a dismissed chemist was issued a U.S. patent between 1920 and 1932. The control group consists of the research fields of other German chemists (defined at the level of 118 USPTO technology classes that include at least one patent by another German chemist, but include no pre-1932 patents by a dismissed chemist). Years between 1920 and 1932 are excluded to estimate the pre-trends. Standard errors are clustered at level of research fields (166 classes).

FIGURE A3 – ENTRY INTO FIELDS OF ÉMIGRÉ AND OTHER CLASSES BY PRIOR PATENT HISTORY



Notes: Panel A separates entrants to research fields of émigrés according to their prior patenting activity in other classes. Entrants *with prior patents in other fields only* measures entrants who had exclusively patented in classes with patents of other German chemists before they patented their first invention in a specific émigré class. Entrants *with prior patents in émigré fields only* measures entrants who had patented in other émigré classes but not in classes with patents of other German chemists before they patented their first invention in a specific émigré class. Entrants *with prior patents in émigré and other fields* measures entrants who had patented in other émigré classes and classes with patents by other German chemists before they patented their first invention in a specific émigré class. Entrants *w/o prior patents* measures entrants who had not patented in any of the 166 classes of our sample before they patented their first invention in a specific émigré class. Panel B performs the corresponding decomposition for entraints into classes of other German chemists.

FIGURE A4 – SWITCHING BETWEEN FIELDS OF ÉMIGRÉS AND OTHER GERMAN CHEMISTS



Notes: *Entry into émigré fields from fields of other German chemists* measure the normalized number of entrants into émigré fields who had prior patents in fields of other German chemists only. *Entry into fields of other German chemists from émigré fields* measure the normalized number of entrants into fields of other German chemists who have prior patents in fields of émigrés, only. Our data include 106 fields with patents by other German chemists (with an average of 218.4 patents until 1932) and 60 fields with patents by émigré chemists (with an average of 149.3 patents until 1932); as a result patentees are more likely to move from fields with patents of other German chemists to fields with patents by émigrés. To account for this mechanical difference, we normalize the number of entrants by the share of pre-1933 patents in each set of fields (i.e. we multiply the number of entrants from fields of other German chemists to fields of émigrés with (106*218.4)/(106*218.4 + 60*149.3), analogously, we multiply the number of entrants from fields of émigrés to fields of other German chemists with (60*140.3)/(106*218.4 + 60*149.3).

APPENDIX TABLES

TABLE A1 - ORDINARY LEAST SQUARES REGRESSIONS
DEPENDENT VARIABLE IS PATENTS PER CLASS AND YEAR BY U.S. INVENTORS

	(1)	(2)	(3)	(4)
1 émigré patent * Post	30.130	16.624		
	(30.557)	(27.411)		
2 émigré patents * Post	107.287**	95.360***		
	(41.700)	(34.845)		
3 or more émigré patents * Post	178.851***	129.608***		
	(23.229)	(20.841)		
1 dismissed patents * Post			30.022	28.559
			(29.316)	(24.421)
2 dismissed patents * Post			136.181***	97.289***
			(34.676)	(34.655)
3 or more dismissed patents * Post			156.390***	98.137***
			(27.761)	(24.670)
# foreign patents	No	Yes	No	Yes
Quadratic class age	No	Yes	No	Yes
Patent pools	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Class fixed effects	Yes	Yes	Yes	Yes
P-value (1 émigré patent * Post= 3 or more émigré patents * Post)	0.0000	0.0002		
P-value (1 dismissed patent * Post= 3 or more dismissed patents * Post)			0.0006	0.0254
Observations	8,466	8,466	8,466	8,466
R-squared	0.790	0.851	0.783	0.850

Standard errors clustered at the class level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is patents by U.S. inventors per USPTO class and year, excluding U.S. patents by émigrés. *n émigré patents* equals 1 when the number of U.S. patents by émigrés in class *c* is equal to *n*. Classes without émigré patents form the control. The dummy variable *Post* equals 1 for years after the dismissals. *n dismissed patents* equals 1 when the number of pre-1933 U.S. patents by dismissed chemists in class *c* is equal to *n*. *# of foreign patents* counts U.S. patents by foreign nationals in class *c* and year *t*. *Quadratic class age* is a second-degree polynomial for years since the first patent in class *c*. The indicator variable *patent pools* equals 1 for classes that were affected by a patent pool.

TABLE A2 – SPECIFICATION CHECKS
DEPENDENT VARIABLE IS PATENTS PER CLASS AND YEAR BY U.S. INVENTORS (COLS 1-4)
AND CITATION-WEIGHTED PATENTS PER CLASS AND YEAR (COLS 5-8)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Poisson			OLS		IV
						Citations-weighted dependent variable		
Émigré class * post	1.435***				211.849***		412.176*	
	(0.154)				(74.036)		(219.853)	
# émigré patents * post		1.061				12.707***		50.456*
		(0.039)				(3.217)		(26.796)
Dismissed class * post			1.493***					
			(0.133)					
# dismissed patents * post				1.386***				
				(0.139)				
# foreign patents	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic class age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Patent pools	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Class fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,466	8,466	8,466	8,466	8,466	8,466	8,466	8,466

Standard errors clustered at the class level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: Columns 1-4: Odds ratios from Poisson regressions. The dependent variable is patents by U.S. inventors per USPTO class and year, excluding patents by émigrés. Columns 5-8: The dependent variable is citation-weighted patents by U.S. inventors per USPTO class and year, excluding patents by émigrés. Citations-weighted patents are calculated by adding the number of times that a patent is cited in patent issues between 1921 and 2002 (from Lampe and Moser 2012) to each patent. Other variables are defined as above.

TABLE A3- ROBUSTNESS CHECK, TREATMENT BEGINS IN 1936
DEPENDENT VARIABLE IS PATENTS PER CLASS AND YEAR BY U.S. INVENTORS

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS		Reduced Form		IV	
Émigré class * Post	74.931***				152.184**	
	(19.143)				(58.403)	
# émigré patents * Post		3.859**				15.853**
		(1.913)				(6.842)
Dismissed class * Post			51.241**			
			(19.709)			
# dismissed patents * Post				20.623***		
				(6.591)		
# foreign patents	Yes	Yes	Yes	Yes	Yes	Yes
Quadratic class age	Yes	Yes	Yes	Yes	Yes	Yes
Patent pools	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Class fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,466	8,466	8,466	8,466	8,466	8,466
R-squared	0.851	0.849	0.848	0.849	0.846	0.826

Standard errors clustered at the class level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is patents by U.S. inventors per USPTO class and year, excluding U.S. patents by émigrés. *Émigré class* equals 1 for classes that include at least one U.S. patent by an émigré. *# émigré patents* measures the number of U.S. patents by émigrés in class *c*. Classes without émigré patents form the control. The dummy variable *Post* equals 1 for years after the dismissals. Instruments are *Dismissed class * Post* (columns 1 and 2) and *# dismissed patents * Post* (columns 3 and 4). *Dismissed class* equals 1 for classes that include at least one pre-1933 U.S. patent by a dismissed chemist. *# dismissed patents* indicates the number of pre-1933 U.S. patents by dismissed chemists in each class. *# of foreign patents* counts U.S. patents by foreign nationals in class *c* and year *t*. *Quadratic class age* is a second-degree polynomial for years since the first patent in class *c*. The indicator variable *patent pools* equals 1 for classes that were affected by a patent pool.

TABLE A4 – ORDINARY LEAST SQUARES, REDUCED FORM AND INSTRUMENTAL VARIABLES REGRESSIONS
INTENSIVE MARGIN: DEPENDENT VARIABLE IS THE NUMBER OF PATENTS BY DOMESTIC U.S. INVENTORS THAT WERE ACTIVE PATENTEES BEFORE 1933

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS		Reduced form		IV	
Share of patents in émigré classes * Post	0.002	-0.008***			-0.003	-0.027***
	(0.001)	(0.003)			(0.004)	(0.008)
Share of pre-1933 patents in dismissed classes * Post			-0.001	-0.011***		
			(0.002)	(0.002)		
Quadratic time to first patent	No	Yes	No	Yes	No	Yes
Quadratic time since first patent	No	Yes	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Inventor fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10,987,389	10,987,389	10,987,389	10,987,389	10,987,389	10,987,389
R-squared	0.011	0.036	0.011	0.036	-	-

Standard errors clustered at the level of an inventor's main class of patenting
*** p<0.01, ** p<0.05, * p<0.1

Notes: The dependent variable is the number of patents obtained by incumbent inventor *i* in year *t*. The sample includes all domestic U.S. patentees with at least one patent between 1920 and 1932. *Share of patents in émigré classes* measures a domestic U.S. inventor's combined share of patents across the 60 research fields of émigrés. The dummy variable *Post* equals 1 for years after the dismissals. *# of foreign patents* counts U.S. patents by foreign nationals in class *c* and year *t*. *Quadratic in time to first patent* is a second-degree polynomial for years until an inventor patents for the first time in any of our 166 classes. *Quadratic time since first patent* is a second-degree polynomial for years after an inventor patents for the first time in any of our 166 classes.