NBER WORKING PAPER SERIES

GERMAN-JEWISH EMIGRES AND U.S. INVENTION

Petra Moser
Alessandra Voena
Fabian Waldinger

Working Paper 19962
http://www.nber.org/papers/w19962

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2014

We wish to thank seminar participants at the ASSA Meetings in Chicago, Bologna, Carnegie Mellon, Chicago, Collegio Carlo Alberto, Harvard, ISNIE, MIT, Naples, the NBER Group on the Development of the American Economy (DAE), and the NBER Group on High-Skilled Immigration, Northwestern, NYU, Paris, Sant'Anna Pisa, Stanford, St. Gallen, Torino, Tübingen, UBC, UCLA, and Warwick, as well as Claudia Goldin, Dietmar Harhoff, Eric Hilt, Naomi Lamoreaux, Maggie Levenstein, Neale Mahoney, Tom Nicholas, and Paul Rhode for helpful comments. We are especially indebted to Ryan Lampe for his invaluable help with collecting the patent data. Arthur Blouin, Mariela Dal Borgo, Nicola Bianchi, Anna Elisabeth Hage, Stephanie Lee, Marina Kutyavina, Michael Talbot, and Tilky Xu provided outstanding research assistance. Moser gratefully acknowledges financial support through NSF CAREER Grant 1,151,180 and the Center for Advanced Study in the Behavioral Sciences. The authors declare that they have no relevant or material financial interest that relate to the research described in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

German-Jewish Emigres and U.S. Invention
Petra Moser, Alessandra Voena, and Fabian Waldinger
NBER Working Paper No. 19962
March 2014
JEL No. J61,N12,O3

ABSTRACT

Historical accounts suggest that Jewish émigrés from Nazi Germany revolutionized U.S. science. To analyze the émigrés' effects on chemical innovation in the U.S. we compare changes in patenting by U.S. inventors in research fields of émigrés with fields of other German chemists. Patenting by U.S. inventors increased by 31 percent in émigré fields. Regressions that instrument for émigré fields with pre-1933 fields of dismissed German chemists confirm a substantial increase in U.S. invention. Inventor-level data indicate that émigrés encouraged innovation by attracting new researchers to their fields, rather than by increasing the productivity of incumbent inventors.

Petra Moser
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305-6072
and NBER
pmoser@stanford.edu

Alessandra Voena
Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
avoena@uchicago.edu

Fabian Waldinger
University of Warwick
f.waldinger@warwick.ac.uk

Historical accounts suggest that German Jewish scientists who fled from Nazi Germany revolutionized U.S. innovation. By 1944, more than 133,000 German Jewish émigrés found refuge in the United States. Most of them were urban white-collar workers; one fifth were university graduates. The National Refugee Service listed roughly 900 lawyers, 2,000 physicians, 1,500 writers, 1,500 musicians, and 2,400 academics (Sachar 1992, p. 495-496; Möller 1984, p. 1). In physics, émigrés such as Leo Szilard, Eugene Wigner, Edward Teller, John von Neumann, and Hans Bethe formed the core of the Manhattan project that developed the atomic bomb. In chemistry, émigrés such as Otto Meyerhof (Nobel Prize 1922), Otto Stern (Nobel Prize 1943), Otto Loewi (Nobel Prize 1936), Max Bergmann, Carl Neuberg, and Kasimir Fajans

> "soon effected hardly less than a revolution…their work on the structures of proteins and amino acids, on metabolic pathways and genetics, almost immediately propelled the United States to world leadership in the chemistry of life" (Sachar 1992, p. 749).

Alternative accounts, however, indicate that the émigrés' contributions may have been limited as a result of administrative hurdles and anti-Semitism. Jewish scientists met with a "Kafkaesque gridlock of seeking affidavits from relatives in America, visas from less-than-friendly United States consuls" (Sachar 1992, p. 495).[1] Once they were in the United States, a rising wave of anti-Semitism made it difficult for them to find employment; in "the hungry 1930s, antisemitism (sic) was a fact of life among American universities as in other sectors of the U.S. economy" (Sachar 1992, p. 498).[2]

This paper presents the first systematic empirical analysis of the effects of German Jewish émigrés on U.S. innovation. Analyses of present-day immigrants to the United States, which exploit geographic variation in the exposure to immigrants, yield ambiguous results. State-level variation of contemporary data indicates that college-educated immigrants may encourage patenting among natives (Hunt and Gauthier-Loiselle 2010). Analyses at the city-level, however, suggest no significant effect (Kerr and Lincoln 2010).

A significant challenge to analyses of geographic variation is that immigrants may choose to live in more innovative regions, so that estimates may overstate immigrants' effects

[1] With the outbreak of the war, refugees became subject to stringent affidavit requirements, including guarantees of substantial cash deposits in American banks. Barely 10 percent of Jews on waiting lists were able to qualify. In 1940, Washington further tightened its visa policy to avoid infiltration by "enemy agents" (Sachar 1992, p. 533).

[2] When émigré scholars eventually managed to find positions, their transition was not easy: "In the Germanic tradition, they often appeared aloof and condescending, a style unfamiliar to the more democratic atmosphere of American campus life" (Sachar 1992, p. 499).

on innovation. To address this problem, Kerr and Lincoln (2010) instrument for the number of immigrants per city by interacting variation in national grants of H1-B visas with city level demand for immigrant workers. In an alternative approach, Borjas and Doran (2012) examine effects of Soviet mathematicians on the research output of incumbent U.S. mathematicians by comparing changes in publications by U.S. mathematicians for fields in which Soviet émigrés were active with other fields. Their analysis suggests that incumbent U.S. mathematicians published less after Soviet mathematicians arrived in the United States, possibly because émigré and U.S. mathematicians competed for journal space and other resources, which were fixed in the short run.

Our analysis extends existing empirical tests by examining *total* changes in U.S. research output, as well as changes for incumbents (which are the focus of Borjas and Doran 2012) and entrants to the fields of émigrés. Taking advantage of the fact that patents are a good measure of innovation in chemistry, because chemical innovations are exceptionally suitable to patent protection (e.g., Cohen, Nelson, and Walsh 2002; Moser 2012), we focus on changes in chemical inventions. By comparison, the contributions of émigré physicists (including those who worked on the Manhattan Project) are difficult to capture empirically because they produced knowledge that was often classified and rarely patented.[3]

Difference-in-differences regressions compare changes in U.S. patenting by U.S. inventors in research fields of German Jewish émigrés with changes in U.S. patenting by U.S. inventors in fields of other German chemists. This approach allows us to control for a potential increase in U.S. invention in fields where German chemists, who had dominated chemical research in the early 20th century, were active inventors. Research fields are measured at the level of 166 United States Patent Office (USPTO) technology classes that include at least one patent by an academic chemist from Germany or Austria between 1920 and 1970. Baseline estimates indicate that the arrival of German Jewish émigrés led to a 31 percent increase in innovation after 1933 in the research fields of émigrés.

[3] Even for chemistry our analysis is limited to *patented* inventions, and many innovations, which benefitted from the arrival of the émigrés may not have been patented. Moser (2005, 2012) addresses this challenge by collecting data on innovations with and without patents from catalogues for international technology fairs between 1851 and 1915. These data indicate that the share of chemical innovations that occurred inside the patent system increased substantially in response to improvements in analytic methods, which reduced the effectiveness of secrecy as an alternative mechanism to protect intellectual property and made it easier to codify chemical inventions (Moser 2012). For the late 20th century, inventor surveys indicate that chemicals and pharmaceuticals are the only industries in which inventors consider patents to be the most effective mechanism to protect intellectual property (e.g. Cohen, Nelson, and Walsh 2002).

Baseline estimates may be biased if the United States attracted more productive scientists or if the émigrés were more likely to work in research fields in which U.S. inventors would become more productive.[4] Historical evidence, however, suggests that émigrés to the United States may have been negatively selected, because Britain, which was geographically and culturally closer to the German university system, was the first refuge for many émigrés (Ambrose 2001, p.215), and established universities, such as Oxford and Cambridge, were keen to offer employment to the most prominent dismissed German scientists.[5]

Historical accounts also suggest that selection into research fields may have been negative because anti-Semitism in the United States restricted access to the most promising fields. For example, the U.S. chemical firm Du Pont rejected the "father" of modern biochemistry Carl Neuberg, because he "looked" too Jewish (Sachar 1992, p. 495). According to Hounshell (1988, pp. 295-296) hiring practices in Du Pont's Chemical Department "were flawed in one important respect: A strong strain of anti-Semitism and sexism prevailed.…" More generally, Deichmann (1999, p. 3) explains that "biochemists and physical chemists were accepted at American universities, whereas organic chemists were not."

To examine whether OLS regressions over- or under-estimate the émigrés' effects, we implement an instrumental variable analysis, which exploits the dismissal of Jewish scientists by the Nazi government. On April 7, 1933 only 67 days after the Nazis assumed power in Germany, the *Law for the Restoration of the Professional Civil Service* required that "Civil servants who are not of Aryan descent are to be placed in retirement" (*Gesetz* §3).

> "At a stroke, every Jew in Germany employed by the government or by state-sponsored local institutions was ordered to be dismissed from his or her post. From university professor to local postmistress, they all had to go…Prominence and reputation shielded no one, as over 1,200 Jewish academics were summarily dismissed" (Ambrose 2001, p. 20).

After the annexation of Austria in 1938, dismissals were extended to Austrian universities, so that the term "German scientists" in this paper includes chemists from both countries.

---

4 More generally, a Roy model of migration implies that more productive immigrants move to locations where returns to skills exceed returns in their home country (Borjas 1987).

5 Arnold Weissberger, for example, moved to Rochester only after he could not secure a university position in Britain and was deemed "unsuitable for industry." Another prominent scientist who worked with Weissberger at Kodak, Gertrud Kornfeld, had studied photochemistry and reaction kinetics as a postdoctoral fellow at the University of Berlin in 1933. Kornfeld first tried to find a position in England, and when this failed moved to Vienna on a fellowship of the American Association of University Women and from there to the United States (Deichman 2005, p. 585-586).

IV regressions use the pre-1933 fields of *dismissed* chemists as an instrument for the fields of émigrés to the United States. Pre-1933 research fields were determined before the Nazis' rise to power and did not depend on expectations about the types of research that would become productive in the United States after 1933. Consistent with historical accounts of negative selection, IV estimates imply a 71 percent increase in patenting, which implies that OLS estimates under-, rather than overestimate the true effects of the émigrés on U.S. invention.

Results are robust to a broad range of alternative specifications, including count data models, regressions with citation-weighted patents as a quality-adjusted measure of patenting, and alternative definitions of the post-period. The most significant decline in the estimated effects occurs when we control for class-specific linear pre-trends in patenting.

In the second part of the analysis, we investigate the mechanism by which the émigrés' arrival encouraged innovation in the United States, using a new data set on the patent histories of all U.S. inventors in the 166 classes of chemical invention.[6] This analysis indicates that the arrival of the émigrés encouraged U.S. invention by helping to attract domestic inventors to the research fields of émigrés, rather than by increasing the productivity of incumbent U.S. inventors. Moreover, data on the prior patent histories of entrants indicate that the majority of entrants to the fields of émigrés had never patented in the 166 classes in our data before, suggesting that the émigrés' arrival affected an overall increase in invention, rather than a shift across fields.

The data also indicate that the effects of the émigrés on U.S. invention may have been amplified and made more persistent through the networks of their co-inventors, which we identify from patent documents. Analyses of contemporary data indicate that researchers in the life sciences benefitted greatly from collaborations with prominent scientists (Azoulay, Graff Zivin, and Wang 2010). In the case of German Jewish émigrés, co-inventors of émigrés became active patentees in the fields of émigrés especially after 1940, and continued patenting through the 1950s. These patterns suggest that a natural delay in the transmission of knowledge from émigré professors to their U.S. collaborators influenced the timing of the increase in U.S. invention. In addition to co-inventors of the émigré professors, co-inventors of co-inventors of the émigrés also substantially increased their inventive activity in émigré

[6] This new data set covers inventors on U.S. patents between 1920 and 1970. For more recent U.S. patent issues, between 1975 and 2010, Lai et al. (2011) have created data on inventor identity and networks of co-inventors.

fields after 1933, and remained substantially more productive throughout the 1950s and 1960s.

Finally, in interpreting these results, it is important to keep in mind that we only observe a small, albeit exceptionally prominent segment of the total flow of German Jewish immigrants to the United States. As a first step towards investigating the effects of this broader flow, we document the research activities of a group of more junior German chemists, who had not yet become professors at German universities. Patent data indicate that these more junior scientists were active in the research fields of émigré professors, suggesting that the fields of émigré professors are a useful proxy for the fields of a broader movement of German Jewish émigrés.

I. THE DATA

To perform this analysis, we have collected new data sets to measure aggregate changes in U.S. patenting across research fields and to investigate changes in research output at the level of individual U.S. inventors. The first data set measures changes in U.S. patents per year across research fields that were differentially affected by the arrival of German Jewish chemists; these data include 1,365,689 U.S. patents by U.S inventors between 1920 and 1970. Research fields are measured at the level of 166 United States Patent Office (USPTO) technology classes; 60 of these classes include patented inventions by German Jewish émigrés to the United States. The second data set captures changes in patenting for individual U.S. inventors across research fields with varying levels of exposure to the arrival of the German Jewish émigrés; these data allow us to examine changes in the productivity of incumbent U.S. inventors and measure changes in entry across research fields.

A. Émigré and other chemistry professors at German and Austrian universities

To capture all 535 chemistry professors and postdoctoral fellows (*privatdozent*) at German and Austrian universities, we use data from faculty directories in the *Kalender der Deutschen Universitäten und Hochschulen* 1932/33, 1933, and *Kürschners Deutscher Gelehrtenkalender,* 1931. Names of dismissed professors were drawn from the *List of Displaced German Scholars* (1937), which the U.K.-based *Emergency Alliance of German*

Scholars Abroad created to help dismissed scientists find employment abroad.[7] The *List* includes German chemistry professors, such as

> BERL, Dr. Ernst, o. Professor; b. 77. (English.) 1916/19: Privatdozent, Technische Hochschule, Vienna; 1919.33: O. Prof. Technische Hochschule, Darmstadt; since 1934: Research Prof. Carnegie Institute of Technology, Pittsburgh. Spec.: Inorganic Chem.; Organic Chem.; Technology; Heavy Chemicals and Derivatives; Cellulose. Perm.

Additional data from Deichmann (2001), Strauss et al. (1983), and Kröner (1983) allow us to identify chemists who were dismissed from Austrian universities after the annexation of Austria in 1938, and chemists who had died before the *List* was published in 1937.

Overall, ninety-three chemists, 17.4 percent of all German and Austrian professors in chemistry, were dismissed between 1933 and 1941. Eighty-seven percent of dismissed chemists were Jewish (Deichmann 2001); most of the remaining dismissed had a Jewish spouse. A small number of scientists who "based on their previous political activities cannot guarantee that they have always unreservedly supported the national state," (*Gesetz* §4), were dismissed as well.[8]

To identify German Jewish émigrés to the United States, we have collected the employment histories for all dismissed scholars, as well as their birth and death years from the *International Biographical Dictionary of Central European Émigrés 1933 - 1945* (Strauss et. al. 1983), and from obituaries in the *New York Times*. We count any dismissed scholar who was professionally active in the United States as a German Jewish émigré to the United States; this yields a total of 26 émigrés.[9] Biographical information confirms anecdotal evidence that émigrés to the United States were younger than other dismissed scholars. In 1933, the average émigré chemist was 45.4 years old, compared to 49.2 years for other dismissed professors.

B. U.S. patents of émigré and other German professors (1920-1970)

To identify the research fields of all German chemistry professors, we collect the U.S.

---

[7] Waldinger (2010, 2012, 2013) has used these data to measure the effects of dismissals on German universities. Dismissals had negative effects on Ph.D. student outcomes (Waldinger, 2010). Departments with dismissals also experienced large and persistent declines in research output (Waldinger 2013). This decline was driven by a fall in the quality of hires and not by localized productivity spillovers (Waldinger, 2012).

[8] Jewish professors who had been civil servants since 1914, fought in World War I, or lost a father or son in the war, were exempt in 1933, but were dismissed after 1935.

[9] Of the remaining dismissed German chemists, 26 became professionally active in the United Kingdom, 6 in Latin America, 5, each in Palestine and Turkey, 4, each in Scandinavia and Switzerland, 3, each in France and Canada, and 2 in Belgium and the Netherlands.

patents that were issued to each of the 535 German chemistry professors between 1920 and 1970 by searching USPTO patent documents through Google Patents (www.patents.google.com). For example, a search for "Arnold Weissberger" yields

> USPTO 2,350,127, issued on May 30, 1944, application filed September 26, 1940, Inventors: Henry Dudley Porter and Arnold Weissberger, Rochester N.Y., Assignors to Eastman Kodak Company, Rochester N.Y., for a "method of forming sulphonic acid chlorides of couplers groups."

For each patent we compare the description of the invention, the date of the patent application and the location of the patentee with the employment histories and the life span of the German chemist to ensure that the patent is a match.[10]

This process yields a total of 946 U.S. patents between 1920 and 1970, including 282 patents by 43 dismissed German chemists and 157 patents by 13 German Jewish émigrés to the United States. Until 1932, émigrés patented few inventions in the United States, with an average of 0.46 patents per year between 1920 and 1932 (Figure 1). After 1933, émigrés to the United States began to patent more in the United States. U.S. patents of émigrés increase from less than 5 per year until 1940, to roughly 10 patents per year until the early 1950s; in terms of application years, this implies an increase in patenting around 1937. Émigrés began to patent less in the mid 1950s, when the average émigré was approaching retirement. By comparison, U.S. patents of other (non-émigré) German chemists began to increase in the 1920s, reaching more than 40 patents per year in 1934. U.S. patents by other German chemists declined after the United States entered World War II on December 11, 1941, and remained low in the immediate aftermath of the war, but recovered in the late 1950s.

[FIGURE 1 approximately here]

U.S. patents by dismissed German chemists increased from 7 per year between the mid 1920s and 1942 to 10 and above in the 1940s; similar to U.S. patents of émigré chemists, U.S. patents by dismissed chemists began to decline in the mid 1950s, when dismissed professors were roughly 70 years old (Figure 1).

---

[10] A search for common names like Hermann Fischer (a lecturer at the University of Berlin in 1933) yields patents by other inventors, which we eliminate by examining each patent. Hermann was the 6th most popular first name when Fischer was born and Fischer is the 4th most common last name in Germany today (Duden, 2000 and www.beliebte-vornamen.de). Only eight dismissed professors have both a first and last name that is among the top 50 most common German names.

C. Matching patents with USPTO classes

To measure the effects of the immigrant chemists across fields of U.S. invention, we use the U.S. patents of German chemists to identify their research fields, measured at the level of main classes within the USPTO system of classifying inventions. For example, Ernst Berl's patent 2,000,815 on May 7, 1935, was assigned to class 205

> "Electrolysis: Processes, Compositions Used Therein, and Methods of Preparing the Compositions"[11]

The U.S. patents of German chemists span 166 USPTO classes, including 60 classes that include at least one patent by an émigré and 106 control classes that include patents by other German chemists, but not the émigrés. Forty-nine USPTO classes include pre-1933 patents by at least one dismissed chemist; we use these classes to instrument for the 60 classes that include patents by at least one émigré (Table 1).

D. U.S. patents by U.S. inventors per class and year

To measure changes in U.S invention across research fields that were differentially affected by the arrival of German Jewish émigrés, we collect all U.S. patents in the 166 classes with patents by German chemists between 1920 and 1970 from the USPTO database *U.S. Patent Master Classification File*.[12] To separate U.S. inventors from foreign inventors, we develop an algorithm to search for the inventors' country of origin in the full text of all U.S. patents that were issued between 1920 and 1970; we access text files of these patents through *Patent Grant Optical Character Recognition (OCR) Text (1920-1979)*.[13]

The dependent variable measures the number of U.S. patents that are issued to U.S. inventors in a given class and year. To measure the émigrés' effect on U.S. inventors net of changes in the émigrés' own patenting activity we exclude patents by émigrés from counts of domestic U.S. patents. Issue dates are available directly from the USPTO.[14] For the 946 U.S. patents of German and Austrian chemists, we also examine the full text of each patent

---

[11] Class 205 is the primary class for this patent; 51 percent of patents are also assigned to a cross-reference class. We include both types of classes, but results are robust to limiting the sample to primary classes. See Lampe and Moser (2012) for additional detail on patenting in cross-reference subclasses as a measure of innovation.
[12] Available at https://eipweb.uspto.gov/2010/MasterClassPatentGrant/mcfpat.zip.
[13] To assess measurement error as a result of OCR, we compare our search results with nationality data in the NBER patent data for years between 1963 and 1970, which are covered by both data sets (the NBER patent data is available at http://elsa.berkeley.edu/~bhhall). This comparison suggests that measurement error is relatively small. For example, 98 percent of patents that we assign to U.K. inventors are U.K. inventors in the NBER data.
[14] At www.uspto.gov, accessed in June 2011.

document to collect both the application and issue year. These data indicate that on average patents are issued 3.3 years after the application, with a standard deviation of 2.0 years.

E. Individual level patent histories for U.S. patentees

To examine the mechanism by which the arrival of émigré scientists may have increased U.S invention, we collect a new data set to measure changes in the number of active U.S. inventors across fields and over time, and document their patent histories in the fields of German chemists. These data include unique identifiers for 964,526 inventors that are listed on the 1,365,689 U.S. patents issued in the research fields of German chemists between 1920 and 1970, as well as information on the timing of their entry into patenting. This section presents a brief summary of the data collection; the data appendix includes a more detailed description.

First, we develop an algorithm to extract strings of data, which contain the names of the patentees, for all 1,365,689 U.S. patents issued between 1920 and 1970 in the 166 classes with patents by German academic chemists. This algorithm uses regular expressions to identify strings that are more likely to contain the inventor's name.[15]

Then we clean the inventor data by correcting more than 3,300 common OCR errors and removing more than 1,100 sub-strings that do not contain inventor names. For example, a common mistake in current OCR software is to misread letters, such as "H" as "I-I," or to misspell names, such as "William" as "Williax." We correct these misspellings by comparing the original images of patent documents with the information that is listed in Google's OCR data, and create an algorithm that corrects these mistakes; this algorithm removes more than 3,300 common mistakes. We then append the algorithm to remove phrases (sub-strings) in Google's OCR data that the algorithm mistakenly assigns to names. For example, misspellings of the term "United States Patent Office" may be counted as part of a name by mistake. We examine the records for such misspellings and append the algorithm to remove 1,100 common errors of this type. We create another algorithm to separate co-inventors that are listed together on a patent document. This algorithm uses first names as an indicator for the beginning of the name of a separate inventor. It performs an automatic search for 3,439 common first names as listed in the U.S. Census of 1920 and in the U.S. Social Security

[15] The full text of patents is available in Google's *Patent Grant Optical Character Recognition (OCR) Text 1920-1979*). Regular expressions are a mechanism to automatically identify strings of text, using patterns of characters and words. See Aho (1990) for a detailed discussion of regular expressions.

Records between 1900 and 1999 (see data appendix).

Finally, we create unique inventor identifiers to track the patenting history of U.S. inventors, using Levenshtein distances to define when two names are different enough to be counted as separate inventors.[16] Levenshtein distances measure the minimum number of insertions, deletions, or substitutions that are necessary to make two strings of characters identical. This allows us to address minor remaining spelling errors, such as writing "Arnold Weissberger" with a missing r as "Arnold Weissberge." To allow for the fact that more letters can be misspelled in longer names, we calculate a normalized Levenshtein distance by dividing the number of necessary changes by the total number of letters in an inventor name. For example, the absolute Levensthtein distance for the two spellings of Weissberger is one - because one character has to be inserted to create a complete match - and the normalized measure is 1/18, one letter has to be changed relative to 18 letters in the first name plus the last name, plus one space. A match is defined as a character with a normalized Levensthein measure below 0.2. We will use these data in section III, to investigate the mechanism by which the German Jewish émigrés may have influenced U.S. innovation. In section II, immediately below, we investigate whether the émigrés caused a significant increase in U.S. innovation.

II. EFFECTS OF ÉMIGRÉS ON DOMESTIC INVENTION IN THE UNITED STATES

In the first step of the analysis we compare changes in patenting by U.S. inventors in research fields of German Jewish émigrés with changes in patenting in fields of other German chemists. Summary statistics suggest a significant increase in U.S patenting in fields that include at least one patent by an émigré. In USPTO classes with émigré patents, patents by U.S. inventors nearly double after 1933, from 149.3 to 287.3 per class and year (Table 1, column 2). By comparison, in USPTO classes with patents by other German chemists, patents by U.S. inventors increase substantially less, from 218.4 to 248.6 per class and year (Table 1, column 3).

[TABLE 1 approximately here]

Data on U.S. patents per field and year indicate a disproportionate increase in U.S. invention

[16] We are grateful to Julian Reif, who developed a matching algorithm to implement the Levenshtein distance matching measure and made it available at http://ideas.repec.org/c/boc/bocode/s457151.html.

after 1933 in research fields of émigrés compared with fields of other German chemists (Figure 2A). Lower patent counts in émigré fields before 1933 are consistent with historical accounts, which suggest that U.S. universities were more likely to accept German Jewish émigrés in fields where U.S invention was weak (Deichman 1999, p. 3). Separating fields of émigrés according to the *number* of émigré patents shows that fields with more émigré patents experienced a larger increase in U.S. invention after 1933 (Figure 2B). The following paragraphs present OLS and IV regressions to systematically investigate these changes.

[FIGURES 2A and 2B approximately here]

A. OLS estimates of changes in patents by U.S. inventors

Baseline OLS regressions estimate

$$(1)\ \textit{Patents by U.S. inventors}_{c,t} = \alpha_0 + \beta\ \textit{émigré class}_c \cdot \textit{post}_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$$

where the dependent variable counts U.S. patents by domestic inventors in technology class *c* and year *t* between 1920 and 1970. The indicator variable $\textit{émigré class}_c$ equals 1 if technology class *c* includes at least one patent between 1920 and 1970 by a German Jewish émigré to the United States; the indicator variable $\textit{post}_t$ equals 1 starting with the year when dismissals first occurred in Germany (1933) and in Austria (1938).[17] USPTO technology classes that include patents by other Germany chemists but not the émigrés form the control group.

The vector $X_{c,t}$ includes three controls for variation in patenting at the level of research fields and years. First, the variable *# of foreign patents* measures the total number of U.S. patents in class *c* and year *t* by foreign inventors from countries that did not receive any dismissed chemists. This helps control for unobservable factors, such as scientific breakthroughs, that may have increased patenting by U.S. inventors independently of the arrival of the émigrés. Second, the variable *class age* measures the number of years that have passed since the first patent was issued in technology class *c* and its square; this helps control for variation in the speed of invention across the life cycle of a technology. Third, the indicator variable *patent pools* distinguishes technology classes in which competing firms

[17] As discussed above, Jewish professors were dismissed from Austrian universities after the annexation of Austria in 1938. Thus, the indicator variable *post* equals 1 for years after 1932 for classes with patents by émigrés from Germany and after 1937 for classes with patents by émigrés from Austria (but not Germany).

agreed to pool their patents; it controls for a potential decline in innovation as a result of the formation of a patent pool (Lampe and Moser 2012).[18] Year fixed effects δ control for unobservable variation in patenting over time that is common across technologies, and class fixed effects f control for unobservable variation in patenting across technologies that is constant over time.[19]

OLS estimates imply that the arrival of émigré chemists increased U.S patenting by a minimum of 31 percent. In classes that include at least one émigré patent, domestic inventors produced 105.2 additional patents per year after 1933, compared with classes that include at least one patent by another German chemist (Table 2, column 1, significant at 1 percent). Controlling for the *# of foreign patents* reduces the estimated effect to 91.7 additional patents per year; controlling for *class age* reduces the estimate to 84.8, and controlling for *patent pools* further reduces the estimate to 75.4 (Table 2, columns 2-4, significant at 1 percent). Compared with a mean of 240.9 patents per class and year in classes with patents by other German chemists, the most conservative estimate of 75.4 implies a 31 percent increase in domestic patenting.

[TABLE 2 approximately here]

Additional specifications use variation in the count of émigré patents across USPTO classes to measure the intensity of exposure to the émigrés:

(2) $\textit{Patents by U.S. inventors}_{c,t} = \alpha_0 + \beta \; \textit{\# émigré patents}_c \cdot \textit{post}_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$

[18] New Deal policies, such as the *National Industrial Recovery Act* (1933-35), which exempted the majority of U.S. industries from antitrust regulation, created a favorable environment for pools and other types of cooperative agreements in the 1930s. Patent data for 20 industries that formed pools between 1930 and 1938 suggest that the creation of a pool led to a decline in innovation, which was particularly pronounced if the pool combined firms that had competed to improve substitute technologies before the pool had formed (Lampe and Moser 2012).

[19] Results are robust to additional controls for research fields in which domestic invention benefitted from the ability to access foreign-owned invention as a result of the *Trading-with-the-Enemy Act* (TWEA). After World War I, domestic invention (measured by the number of U.S. patents by domestic inventors) increased by 20 percent in USPTO subclasses of chemical inventions in which the TWEA allowed U.S. firms to produce enemy-owned inventions (Moser and Voena 2012).

where # *émigré patents*$_c$ measures the number of émigré patents between 1920 and 1970 in class *c*. Estimates of these regressions imply an increase in U.S invention by 4 patents per year for each additional émigré patent (Table 2, column 8, significant at 5 percent).[20]

Specifications that separately estimate effects according to the number of émigré patents confirm that émigré fields with more patents by émigrés experienced a larger increase in U.S. invention after 1933.[21] In classes with one patent by an émigré, U.S. inventors patented 16.6 additional inventions per class and year after 1933 compared with fields by other German chemists, but the effect is not statistically significant (Appendix Table A1, column 2). In classes with two patents by émigrés, U.S. inventors patented 95.4 additional inventions (Appendix Table A1, column 2, significant at 1 percent). In classes with three or more patents by émigrés, U.S. inventors patented 129.6 additional inventions (Appendix Table A1, column 2, significant at 1 percent).

B. Annual coefficients for years before and after 1933

To investigate the timing of the increase in U.S. invention, we estimate the difference-in-differences coefficient $ß_t$ separately for each year, allowing it to be different from zero *before* 1933.

$$(3)\ \textit{Patents by U.S. inventors}_{c,t} = \alpha_0 + \sum_{t=1920}^{1970} ß_t \textit{émigré class}_c \cdot \textit{year}_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$$

where the variable $year_t$ represents an indicator variable for each year between 1920 and 1970, and 1932 is the excluded category.

Estimates of annual coefficients indicate that the observed increase in patenting cannot be explained by differential pre-trends. Annual coefficients are close to zero before 1933 and increase to the highest level in the 1950s and early 1960s (Figure 3).[22]

[FIGURE 3 approximately here]

[20] Results are robust to alternative definitions of the post-period, including specifications that define *post* to begin in 1936 (reported below).

[21] Among 60 émigré classes, 24 classes include 1 émigré patent, 10 classes include 2 émigré patents, and 26 classes include three or more émigré patents.

[22] Figure 3 is the regression analog of Figure 2A, which plots the difference between average patents per year in classes with and without émigré patents (Figure 2A). Differences between the two figures are driven primarily by the inclusion of class fixed effects.

These results, which are consistent with a protracted adjustment process (Sachar 1992), indicate that unobservable factors that preceded the arrival of the émigrés are unlikely to have been the driving force behind the increase in U.S. patenting. An additional set of regressions controls for class-specific linear pre-trends in patenting:

$$(4)\quad \textit{Patents by U.S. inventors}_{c,t} = \alpha_0 + \sum_{\tau=1933}^{1970} ß_\tau \textit{émigré class}_c \cdot \textit{year}_\tau + \eta_c \cdot t + \gamma' X_{c,t} + \delta_t + f_c + \upsilon_{c,t}$$

where we allow time trends t to differ for each of the 166 classes η_c (by including the interaction term $\eta_c \cdot t$), and the variable $year_\tau$ represents an indicator variable for each year between 1933 and 1970) Controlling for linear pre-trends leaves the point estimates substantially unchanged but makes them less precise over time, so that many of the annual coefficients are no longer statistically significant (Figure A1). An F-test statistic of 3.26, however, rejects the joint hypothesis that all annual coefficients are equal to zero with a p-value below 0.0001.

C. Pre-1933 fields of dismissed as an instrument – First stage

Baseline OLS estimates may, however, be biased, if the United States attracted the most productive émigrés, or if émigré scientists were attracted to more productive fields once they had arrived in the United States. In fact, patent data indicate that USPTO classes with émigré patents were on average four years younger than classes without émigré patents. In 1932, 84.6 years had passed since the first patent grant in the average émigré class, compared with 88.7 years for other classes. A test for the equality of means rejects equality with a p-value of 0.085 (Table 1, columns 2 and 3). Invention in younger research fields may have increased independently of the émigrés.

To address endogeneity, we use the pre-1933 patents of *dismissed* chemists to instrument for the 1920-1970 patents of émigrés to the United States. This approach exploits the fact that the research decisions of German Jewish chemists prior to their dismissal are unlikely to have depended on their expectations about the types of research that would become more productive in the United States after 1933.

To examine whether the pre-1933 patents of dismissed chemists are a valid instrument, we compare pre-1933 characteristics of classes with and without pre-1933 patents of dismissed chemists. First, dismissed chemists may have worked in younger fields that experienced a more rapid increase in patenting after 1933. The data, however, reveal no statistically significant differences for classes with and without pre-1933 patents of dismissed chemists (at an average age of 87.4 years compared with 87.3 in 1932, with a p-value of 0.929 for the equality of means test, Table 1, columns 4 and 5). A related concern is that dismissed chemists may have worked in more productive fields before 1933. To investigate this issue, we compare counts of U.S. patents by *foreign* inventors in classes with and without pre-1932 patents of dismissed chemists. This comparison also reveals no significant differences. If anything, classes with pre-1933 patents of dismissed chemists attracted slightly fewer foreign patentees until 1933, but this difference is not statistically significant (with 0.70 versus 1.01 U.S. patents by foreign inventors and a p-value of 0.216).

First-stage regressions estimate:

(5) $\textit{Émigré class}_c \cdot \textit{post}_t = \zeta_0 + \phi\, \textit{pre-1933 dismissed class}_c \cdot \textit{post}_t + \theta' X_{c,t} + \lambda_t + \mu_c + \upsilon_{c,t}$

A coefficient of 0.339 for the variable $\textit{pre-1933 dismissed class}_c \cdot \textit{post}_t$ and an F-statistic on the excluded instrument of 18.25 (Table 3, column 2) confirms that pre-1933 fields of dismissed chemists are a strong predictor for fields of émigrés. An analogous first-stage regression uses the number of pre-1933 patents by dismissed chemists in class *c* as an instrument for the number of patents by émigrés in class *c*. For this regression, the coefficient is 1.303, and the F-statistic on the instrument is 8.99 (Table 3, column 4).

[TABLE 3 approximately here]

D. Reduced form estimates for pre-1933 fields of dismissed chemists

Similar to data for patents per year in émigré fields, data for fields with pre-1933 patents by *dismissed* chemists also indicate a disproportionate increase after 1933 in U.S. invention (Figure 4); by the mid 1950s, U.S. inventors produced more patents in fields with pre-1933 patents by dismissed German Jewish chemists.

[FIGURE 4 approximately here]

To analyze whether patenting by U.S. inventors in pre-1933 fields of dismissed chemists increased after 1933 compared with fields of other German chemists we estimate the reduced form:

$$\text{(6) } \textit{Patents by U.S. inventors}_{c,t} = \alpha_0 + \beta \textit{ pre-1933 dismissed class}_c \cdot \textit{post}_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$$

where the indicator variable *pre-1933 dismissed class*$_c$ equals 1 for technology classes *c* that include at least one pre-1933 patent by a dismissed German chemist.

In USPTO technology classes that include at least one pre-1933 patent by a dismissed chemist, U.S inventors produce 57.8 additional patents per year after 1933 (Table 3, column 6, significant at 1 percent). Compared with an average of 240.9 patents per class and year between 1920 and 1970 in fields of other (non-émigré) German chemists, this implies a 24 percent increase in domestic patenting. Analogous reduced form estimates imply that U.S inventors produced 22.3 additional patents per class and year for each additional patent by dismissed German chemists (Table 3, column 8, significant at 1 percent).

Specifications that separately estimate effects according to the number of pre-1933 patents by dismissed chemists confirm that fields with more pre-1933 patents by dismissed chemists experienced a larger increase in U.S. invention after 1933.[23] In classes with one pre-1933 patent by a dismissed chemist, U.S. inventors patented an additional 28.6 inventions per class and year after 1933 compared with fields by other German chemists, but the effect is not statistically significant (Appendix Table A1, column 4). In classes with two pre-1933 patents by dismissed chemists, U.S. inventors patented an additional 97.3 inventions (Appendix Table A1, column 4, significant at 1 percent). In classes with three or more pre-1933 patents by dismissed chemists, U.S. inventors patented an additional 98.1 inventions (Appendix Table A1, column 4, significant at 1 percent).

To investigate the sensitivity of the reduced form results to differential pre-trends we estimate an additional set of regressions that control for linear class-specific pre-trends:

$$\text{(7) } \textit{Patents by U.S. inventors}_{c,t} = \alpha_0 + \sum_{\tau=1933}^{1970} ß_\tau \textit{pre} - 1933 \textit{ dismissed class}_c \cdot \textit{year}_\tau + \eta_c t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$$

Time-varying estimates with linear pre-trends track estimates without pre-trends albeit at a lower level and with standard errors that increase as we move away from the pre-period (Figure A2), suggesting that the baseline estimates may overestimate the true effects of

[23] Among 48 classes with pre-1933 patents by dismissed chemists, 27 classes include 1 pre-1933 patent by a dismissed chemist, 9 classes include 2 pre-1933 patents by a dismissed chemist, and 12 classes include three or more pre-1933 patents by a dismissed chemist.

immigration. An F-test statistic of 2.37 rejects the joint hypothesis that all coefficients are equal to zero with a p-value equal to 0.0001.

E. Instrumental variables estimates

IV regressions that use *pre-1933 dismissed class*$_c$ as an instrument for *émigré class*$_c$ imply that U.S. inventors produce 170.1 additional patents per class and year in fields of émigrés compared with fields of other German chemists (Table 4, column 2, significant at 1 percent). Compared with a mean of 240.9 patents per class and year between 1920 and 1970 in fields of other German chemists, this implies an increase in U.S. patenting of 71 percent.

[TABLE 4 approximately here]

IV regressions proxy for the effects of knowledge that dismissed German chemists had acquired in Germany and brought to the United States. More precisely, the local average treatment effect of the IV regressions (LATE, Imbens and Angrist 1994), estimates the increase in patenting by U.S. inventors for classes in which *émigrés* to the United States patented because *dismissed* chemists had patented in the same classes before 1933. In addition to the fact that the IV estimates a LATE, some of the difference between the OLS and IV estimates may reflect measurement error, which attenuates the OLS estimates. The large difference between OLS and IV estimates is also consistent with historical accounts of negative selection at the level of individual scientists and fields (e.g., Deichmann 1999).

Regressions that use the number of pre-1933 patents by dismissed chemists as an instrument for the number of émigré patents indicate that U.S. inventors produced 17.1 additional patents per year for each additional émigré patent (Table 4, column 4, significant at 5 percent).

F. Robustness checks

Results are robust to a broad range of alternative specifications, including count data models, regressions with citation-weighted patents as a quality-adjusted measure of patenting, and alternative definitions of the post-period.

The first robustness check estimates the main specifications as Poisson regressions with conditional fixed effects to address the count data characteristic of patents. They yield comparable or larger estimates than OLS. Poisson estimates for the difference-in-differences estimator *émigré class*$_c$ · *post*$_t$ imply a 44 percent increase in U.S. patenting in fields of émigrés (Appendix Table A2, column 1, significant at 1 percent), compared with 31 percent in OLS. For each additional émigré patent, U.S. patenting increased by 6 percent (Appendix Table A2, column 2, not statistically significant).

Poisson estimates for the reduced form imply a 49 percent increase in U.S. patenting in pre-1933 research fields of dismissed chemists (Appendix Table A2, column 3, significant at 1 percent). For each additional pre-1933 patent of a dismissed chemist, domestic patenting increased by 39 percent (Appendix Table A2, column 4, significant at 1 percent).

An additional test accounts for differences in the quality of patents using data from Lampe and Moser (2012) on counts of later patents that cite each patent as relevant prior art.[24] In this test, the dependent variable *citation-weighted patents by U.S. inventors*$_{c,t}$ measures the number of times a patent issued in year *t* and class *c* was cited in patents issued between 1921 and 1979.

$$(8)\ \textit{Citation-weighted patents by U.S. inventors}_{c,t} = \alpha_0 + \beta\ \textit{Émigré class}_c \cdot \textit{post}_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$$

OLS estimates imply an increase of 211.8 citation-weighted patents per class and year after 1933 in research fields of émigrés (Appendix Table A2, column 5, significant at 1 percent). Compared with a mean of 616.2 citation-weighted patents per class and year in the control, this implies a 34 percent increase, slightly above the baseline estimate of 31 percent for raw patents. For each additional patent by an émigré, U.S. inventors produce 12.7 additional citation-weighted patents after 1933 (Appendix Table A2, column 6, significant at 1 percent).

Instrumental variable regressions indicate that U.S. inventors produced an additional 412.2 citation-weighted patents per year after 1933 in classes with émigré patents (Appendix Table A2, column 7, significant at 10 percent). Compared with a mean of 616.2 citation-weighted patents per class and year in the control group, this implies a 67 percent increase. Analogous regressions, which measure the number of émigré patents, indicate that U.S.

[24] Citations are the standard approach to control for the quality of patented inventions. For example, Trajtenberg (1990) documented that citations are correlated with the estimated social surplus that 456 improvements in CAT scanners created over time. Hall, Jaffe, and Trajtenberg (2000) show that citation-weighted patent stocks are more highly correlated with market value (measured by Tobin's q) than patent stocks. Moser, Ohmstedt, and Rhode (2013) find that citations are positively correlated with the size of patented improvements in hybrid corn.

inventors produced 50.5 additional citation-weighted patents after 1933 (Appendix Table A2, column 8, significant at 10 percent).

In the baseline, we define the *post* period to begin in 1933 to exploit the exogenous timing of dismissals. Émigrés, however, may have become active in the United States with some delay; to address this issue, we check that the estimates are not driven by an increase in U.S. patenting that occurs too early to reflect an effect of the émigrés. To perform this test, we re-estimate the main specifications with alternative definitions of the *post* period, beginning in 1936 and 1940.[25]

OLS estimates, in which the *post* period begins in 1936, indicate that U.S. inventors produced 74.9 additional patents per year after 1936 in fields of émigrés compared with fields of other German chemists (Appendix Table A3, column 1, significant at 1 percent). Analogous IV estimates imply that U.S. inventors produced 152.2 additional patents per year after 1936 (Appendix Table A3, column 5, significant at 5 percent). Thus, both OLS and IV estimates are similar to the main estimates, suggesting that the results are not driven by the definition of the *post* period. Equivalent analyses in which *post* begins in 1940 confirm these findings.[26]

III. Investigating the Mechanism Using Inventor-Level Data

To investigate the mechanism by which the arrival of German Jewish émigrés increased U.S. innovation, we perform additional tests using a new inventor-level data set of changes in U.S. patenting. Specifically, we examine changes in the productivity of incumbent U.S. inventors, as well as changes in entry by new patentees across fields of chemistry. We also investigate networks of co-inventors, which may have amplified the effects of German Jewish émigrés, and document the arrival of other German chemists, which indicates that the emigration of German chemistry professors was part of a broader movement of scientists to the United States.

A. Effects on incumbent U.S. inventors

[25] For classes treated by Austrian émigrés only, the *post* period begins with the annexation of Austria in 1938, and in 1940 for the second robustness check.

[26] Defining the *post* period to begin in 1940, the OLS coefficient on *émigré class* $\cdot$ $post_t$ is 73.160 with a standard error of 18.908 (and p-value<0.001). The IV coefficient is 131.836 with a standard error of 57.652 (p-value=0.023). The OLS coefficient on # *émigré patents* $\cdot$ $post_t$ is 3.991 with a standard error of 1.956 (p-value=0.043). The IV coefficient is equal to 17.136 with a standard error equal to 6.909 (p-value=0.014).

To investigate the émigrés' effects on incumbent U.S. inventors we examine changes in patenting for 210,410 U.S. inventors that had patented at least one invention before 1933 in a research field of German chemists.

Summary statistics indicate a decline in patenting for incumbent inventors regardless of their exposure to the arrival of the émigrés. Since 75 percent of incumbent inventors only had one patent, the probability of patenting drops mechanically after 1933, but there is no significant difference for incumbents that were more or less exposed to the émigrés. Incumbent inventors who patented the majority of their inventions in émigré fields patented at least one invention per year with a probability of 0.015 after 1933 compared with 0.097 before 1933 (Table 5, column 4). By comparison incumbent inventors who patented mostly in fields of *other* German chemists patented at least one invention per year with a probability of 0.013 after 1933 compared with 0.098 before 1933 (Table 5, column 2).

[TABLE 5 approximately here]

OLS and IV regressions estimate the differential effects of the émigrés on incumbent inventors, depending on the share of the incumbent's patents in research fields of émigrés:

$$(9)\ Patenting_{i,t} = \alpha + \beta\ share\ of\ patents\ in\ émigré\ classes_i \cdot Post_t + \gamma' Z_{i,t} + \delta_t + f_i + \varepsilon_{i,t}$$

where the dependent variable equals 1 if the incumbent U.S. inventor i patented at least one invention in year t, and 0 otherwise. The coefficient β measures the change in the probability of patenting after 1933 for inventors who have a higher share of their patents in fields of émigrés. The variable $Z_{i,t}$ controls for variation in productivity over the life cycle of an inventor; specifically, we control for changes in productivity relative to the year of an inventor's first patent, by measuring how many years the inventor is still away from his first patent, and how many years have passed since the inventors' first patent. Both variables enter linearly and as a quadratic. The variable f_i represents a full set of fixed effects for each of the 210,410 incumbent U.S. inventors to control for characteristics of the inventors (e.g. their inherent ability) that do not vary over time. Year fixed effects δ_t control for changes in the probability of patenting over time (e.g. as a result of changes in patent policies or industry-level productivity shocks) that influence all inventors. Standard errors are clustered at the level of the class that includes the majority of the incumbent inventor's patents.

OLS estimates indicate that incumbent inventors who had a 10 percent larger share of their patents in émigré classes became 0.07 percentage points *less* likely to patent an invention after 1932 (Table 6, column 2, significant at 1 percent). Regressions without controls for productivity across the inventor's patenting career imply an increase of 0.02 percentage points (Table 6, column 1, significant at 5 percent).[27]

[TABLE 6 approximately here]

Instrumental variable regressions use the share of an inventor's pre-1933 patents in fields with pre-1933 patents of dismissed chemists (interacted with a post-dismissal dummy) as an instrument for the share of the inventor's overall patents in research fields of émigrés (interacted with a post-dismissal dummy). Thus, first stage regressions estimate:

(10) $\textit{Share in émigré classes}_i \cdot \textit{post}_t = \phi\, \textit{pre-1933 share in classes with pre-1933 patents of dismissed}_i \cdot \textit{post}_t + \theta' Z_{i,t} + \lambda_t + \mu_i + \upsilon_{i,t}.$

A coefficient of 0.402 for the variable *pre-1933 share in classes with pre-1933 patents of* $\textit{dismissed}_i \cdot \textit{post}_t$ and an F-statistic on the excluded instrument of 21.65 in the first stage regression (Table 7, column 2, significant at 1 percent) confirm that an inventor's pre-1933 share in pre-1933 classes of dismissed chemists is a good predictor for the inventor's share in émigré classes.

[TABLE 7 approximately here]

Reduced form estimates indicate that researchers who have an additional 10 percent of their pre-1933 patents in pre-1933 fields of dismissed chemists were 0.09 percentage points less likely to patent after 1933 (Table 7, column 4, significant at 1 percent). Instrumental variable estimates imply that chemists who had an additional 10 percent of their patents in fields of émigrés were 0.22 percentage points less likely to patent after 1933 (Table 6, column

[27] Only 0.5 percent of inventors receive more than 1 patent in a given year, 3.0 percent receive 1 patent, and 96.5 percent receive no patents. Reflecting this data structure, estimates of the intensive margin are similar to estimates of the extensive margin (Appendix Table A4).

4, significant at 1 percent), confirming that effects on incumbent inventors cannot explain the observed overall increase in patenting.

We also examine raw data on changes in inventive output after 1933 for three groups of inventors that were more or less exposed to the arrival of the émigrés (Figure 5).[28] Since incumbent inventors are defined as inventors who have produced at least 1 patent before 1933, and 75 percent of incumbents only have 1 patent, patent counts drop mechanically after 1933. Comparing the probability of patenting for incumbents who were differentially exposed to the arrival of émigrés, however, indicates no differential change in patenting. There is no noticeable difference in the probability of patenting after 1933 for incumbents with more than half of their patents in fields of émigrés compared with incumbents with fewer than half of their patents in fields of émigrés (Figure 5). Equivalent comparisons for incumbents with different shares of their pre-1933 patents in pre-1933 fields of *dismissed* German chemists (Figure 6) also indicate no differential change.[29]

[FIGURES 5 and 6 approximately here]

In sum, the data indicate that knowledge spillovers from the émigrés to incumbent inventors are unlikely to have been the driving factor behind the substantial increase in U.S patenting after 1933 in research fields of émigrés. These results are consistent with evidence from publications data, which suggest that incumbent U.S. mathematicians did not benefit from the arrival of Soviet émigrés (Borjas and Doran 2012).[30]

B. Effects on entry into research fields of émigrés

An alternative mechanism, by which the arrival of highly skilled émigrés may have encouraged innovation, is by encouraging U.S scientists to switch into fields of émigrés or by

[28] As a group, incumbent inventors with 50 percent of their patents in émigré fields are more productive, by construction, than inventors with either fewer or more than 50 percent of their patents in émigré fields, because the group of inventors with 50 percent of their patents is restricted to inventors with at least two patents.

[29] Analogous comparisons for alternative divisions of the sample (e.g., 25% in émigré fields versus 75% in émigré fields) confirm these results.

[30] Borjas and Doran (2012) find that the arrival of Soviet mathematicians who emigrated to the United States after the collapse of the Soviet Union crowded out publications in top journals by incumbent U.S. mathematicians. For chemistry, physics, and mathematics, Waldinger (2012) shows that there was no significant effect of the dismissals of Jewish professors on publications by other German professors who stayed in Germany, even though the dismissals had significant negative effects on Ph.D. students in mathematics (Waldinger 2010).

attracting a *new* group of U.S scientists to the fields of émigrés.[31] To investigate this mechanism, we use a researcher's first patent in a USPTO class to measure the researcher's year of entry into a new field, and compare changes in the rate of entry after 1933 for fields of émigrés and fields of other German chemists. To distinguish entry by new inventors from entry by inventors who had already been active in other fields of chemistry, we also separate entrants with and without prior patents in the 166 research fields in our data.

Summary statistics indicate a substantial increase in entry by domestic U.S. scientists to fields of émigrés after 1932. Until 1932, 116.1 U.S. researchers per class and year entered the fields of émigrés, compared with 175.1 U.S. researchers in fields of other German chemists. After 1933, 179.3 U.S. researchers per class and year entered the fields of émigrés, compared with 162.8 in fields of other German chemists (Table 8, columns 2 and 3, Panel A and Figure 7). Similarly, the data indicate a substantial increase in entry by U.S. scientists who had never patented in any of the 166 classes before. Until 1932, 92.0 new U.S. researchers per class and year entered the fields of émigrés, compared with 143.8 new researchers in fields of other German chemists. After 1932, 112.1 new researchers per class and year entered the fields of émigrés, compared with 109.0 in fields of other German chemists (Table 8, columns 2 and 3, panel B).

[FIGURE 7 approximately here]
[TABLE 8 approximately here]

To investigate changes in entry by U.S. patentees, OLS regressions estimate

$$(11)\ Entry_{c,t} = \alpha_0 + \beta\ émigré\ class_c \cdot post_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$$

where the dependent variable counts new researchers per class and year, measured by a researcher's first patent in class *c*. As above, $émigré\ class_c \cdot post_t$ equals 1 after the dismissals for class *c* if it includes at least one patent by an émigré; the vector $X_{c,t}$ includes controls for variation in patenting at the level of classes and years, as defined for equation (1); δ_t are year fixed effects and f_c are class fixed effects.

OLS estimates indicate that an additional 58.2 U.S. researchers entered the fields of émigrés per class and year after the dismissals (Table 9, column 2, significant at 1 percent).

[31] Borjas and Doran (2013) document that U.S. mathematicians switched away from the research fields of Soviet mathematicians to avoid direct competition.

Compared with an average of 165.9 entrants to fields of other German chemists, this implies 35 percent additional entrants for fields of émigrés.

[TABLE 9 approximately here]

Separating entry of new inventors from entry of inventors who had already been active in other fields of chemistry, we find that new inventors accounted for three quarters of additional entrants into émigré fields after the dismissals. Estimates for the dependent variable *entrants into patenting* indicate that the number of new patentees in émigré classes - without prior patents in any of the 166 classes - increased by 44.0 entrants per class and year (Table 9, column 4, significant at 1 percent).

We perform a more detailed analysis which separates entrants into research fields of émigrés who had previously patented in other fields into three groups: inventors with prior patents in other émigré classes only, inventors with prior patents in non-émigré classes only, and inventors with prior patents in both other émigré classes and non-émigré classes. The majority of entrants who had previously patented in other fields had patented in both non-émigré classes and other émigré classes before they began to patent in an émigré class (Appendix Figure A3, Panel A). Relatively few entrants had either patented exclusively in other émigré classes or in non-émigré classes, suggesting that non-émigré classes are an appropriate control.[32]

To further examine whether classes with patents by non-émigré German chemists are a good control, we compare patterns of switching between émigré and non-émigré classes. Controlling for the total number of pre-1933 patents, nearly the same numbers of patentees switched from émigré into non-émigré classes and from non-émigré into émigré classes. Most importantly, there is no evidence for a differential change after 1933 (Appendix Figure A4).

To address the potential concern that entry into research fields of émigrés may be endogenous, we use the pre-1933 research fields of dismissed chemists as an instrument for the fields of émigrés. By construction, first stage regressions for this specification are identical to first stage regression for the baseline, and confirm that the pre-1933 fields of

[32] The corresponding analysis for entrants into fields of other German chemists similarly indicates that most entrants with previous patents (in any field) had patented in both émigré and other non-émigré classes before they began to patent in a specific non-émigré class (Appendix Figure A3, Panel B).

dismissed chemists are a good predictor of the fields of émigrés, with an F-statistic on the excluded instrument of 18.3 (Table 3, column 2). Summary statistics indicate that pre-1933 fields of dismissed chemists attracted fewer entrants before 1933. After 1933, entry into pre-1933 fields of dismissed chemists increased relative to other fields (Table 8 and Figure 8).

[Figure 8 approximately here]

Instrumental variable estimates indicate that entry into the fields of émigrés increased by 142.1 researchers per class and year after the dismissals (Table 9, column 6, significant at 1 percent); entry by patentees without prior patents in the 166 classes increased by 109.5 patentees per class and year (Table 9, column 8, significant at 1 percent). These results imply that about three quarters of the new researchers who entered the fields of émigrés had no prior patents in the 166 classes. Thus, entry data indicate that the émigrés' effect on U.S. patenting was driven primarily by their ability to attract a new group of domestic inventors to their fields.

C. Co-Inventors and co-inventors of co-inventors

To further investigate the mechanism by which émigrés encouraged U.S. innovation, we collect data on all co-inventors of the émigrés from joint U.S. patents. Specifically, the impact of the émigrés may have been amplified and made more persistent through their collaborators. Overall, 47 co-inventors were granted at least one patent with one of the émigrés. Between 1920 and 1970, co-inventors patented 576 inventions in the 166 classes; 134 of them were joint patents with émigrés.

Scientists who became co-inventors of émigrés after the dismissal became disproportionately more likely to patent in émigré fields, not only in joint patents but also in their independent work. Before 1933, inventors who later became co-inventors of émigrés, patented 8 inventions. These patents were equally distributed across fields with and without émigré patents; 4 patents were exclusively assigned to émigré fields, and 4 patents were exclusively assigned to other fields. After 1933, co-inventors patented a total of 568 inventions, including 469 patents (83 percent) that were exclusively assigned to émigré fields (Table 10, Panel A), 24 patents (4 percent) that were exclusively assigned to other fields, and 75 patents (13 percent) that were assigned to both.

[TABLE 10 approximately here]

Confirming the time patterns of the main estimates (Figure 3), co-inventors patenting activity in émigré fields increased most dramatically after 1940, from less than 10 to more than 20 patents per year, and remained high until the second half of the 1950s (Figure 9). Even in the 1960s, the number of patents in émigré fields remained above 10 in the early part of the decade and increased to 18 patents in 1967. Co-inventors' patents that were assigned to both émigré and other fields began to increase in 1940, albeit at lower levels, and continued to increase until the late 1960s.

[FIGURE 9 approximately here]

We also identify the *co-inventors of co-inventors* of the émigrés. Overall, 154 co-inventors of co-inventors patented at least one invention jointly with a co-inventor of an émigré. Between 1920 and 1970, co-inventors of co-inventors patented 1,660 inventions in the 166 classes; 177 inventions were jointly patented with co-inventors of émigrés. Similar to first-degree co-inventors, co-inventors of co-inventors became disproportionately more likely to patent in émigré fields. Before 1933, co-inventors of co-inventors patented 131 inventions, including 48 patents (37 percent) that were exclusively assigned to émigré fields, (Table 10, Panel B), 59 patents (45 percent) that were exclusively assigned to other fields and 24 patents (18 percent) that were assigned to both. After 1933, co-inventors of co-inventors patented a total of 1,529 inventions, including 1,103 patents (72 percent) that were exclusively assigned to émigré fields (Table 10, Panel B), 162 patents (11 percent) that were exclusively assigned to other fields, and 264 patents (17 percent) that were assigned to both.

These data suggest that the émigrés' effect on their collaborators may have been a significant channel by which the arrival of émigré chemists increased U.S. invention. Collaborators of émigrés switched into research fields of émigrés after 1933, and continued to patent at higher levels throughout the 1950s. These patterns are even more pronounced when we consider networks of collaboration more broadly by including co-inventors of co-inventors.

D. Other, more junior German émigré chemists

While our main tests are limited to examining the effects of émigré *professors* on U.S. innovation, émigré professors may have been only the 'tip of the iceberg' of a broader movement of scientists, which also included junior, and less prominent German chemists. As a first step towards investigating this phenomenon, we collect data on younger German chemists who emigrated from Nazi Germany. Straus (1983) reports the names of 62 German chemists who were at least 18 years old in 1933 - but did not hold a faculty position at the time of the dismissals. These individuals included university students, and research assistants, as well as a small number of young industrial chemists who had worked at companies such as Hoffmann-La Roche, Hoechst, and Schering. Thirty-four of them moved to the United States after 1933. The average age of the junior émigrés was 30 in 1933, compared with an average age of 45 years for professors.

Patent data indicate that these junior chemists were active inventors in the same fields as émigré professors. Junior émigrés patented 175 inventions in the United States between 1920 and 1970 in the 166 classes of invention in our data; nearly all of these patents, 169 of 175 patents, were issued after 1933. 113 of the junior émigrés post-1933 patents (67 percent) were issued in classes with patents by senior émigrés; 34 patents (20 percent) were assigned to both émigré classes and classes with patents by other German chemists. Only 22 patents (13 percent) were assigned to classes that include only patents by other German chemists but not by émigré professors (Table 11). These statistics suggest that the research fields of prominent émigré professors, which we can capture with existing records, may be a proxy for the research fields of a broader, largely unobservable flow of German Jewish scientists, who may have contributed to the observed increase in U.S. invention.

[TABLE 11 approximately here]

IV. CONCLUSIONS

Historical accounts suggest that German Jewish émigrés revolutionized U.S. science and innovation, but empirical evidence has been scarce. This paper presents the first systematic analysis of the émigrés' effects on U.S. innovation. Baseline estimates compare changes in patenting by U.S. inventors after 1933 in chemistry for research fields of German émigrés with fields of other German chemists. This analysis indicates that U.S. invention increased by 31 percent after 1933 in fields of U.S. émigrés. A potential threat to the empirical approach is that émigrés may have chosen to work in fields, in which U.S. invention became more productive after 1933, after they had moved to the United States. To address this issue, we use the pre-1933 fields of dismissed German chemists as an instrument for the fields of émigrés to the United States. Consistent with historical accounts that émigrés to the United States may have been negatively selected, and that they were more likely to work in *less* productive research fields in the United States, estimates from instrumental variable regressions exceed estimates from OLS.

To investigate the mechanism by which the arrival of German Jewish émigré scientists encouraged U.S. innovation, we have collected a new inventor-level data set of changes in U.S. patenting. These data indicate that the arrival of German Jewish émigrés increased U.S. invention by attracting a *new* group of domestic U.S. inventors to the fields of émigrés, rather than by increasing the productivity of incumbent U.S. scientists. Our findings of limited positive effects on incumbents are consistent with results from publications data for mathematics (Borjas and Doran 2012), which suggest that the arrival of a new group of highly skilled scientists may crowd out publications by incumbents. Analyzing patents instead of publications, however, allows us to investigate effects on incumbents in a setting that is less affected by capacity constraints, and estimate the *overall* effects of high-skilled immigrants on innovation.

The data also indicate that networks of co-inventors may have helped to amplify the émigrés' effects on U.S. innovation. U.S. inventors who collaborated with émigré professors began to patent at substantially higher levels in the 1940s and continued to be exceptionally productive in the 1950s. These patterns suggest that émigré professors helped to increase U.S. invention in the long run, by training a new group of younger U.S. scientists, who then continued to train other scientists.

Importantly, our analysis is limited to investigating changes in U.S. invention in the research fields of a small, albeit prominent group of German Jewish émigré professors. Comparisons with patent data for a younger group of less prominent German Jewish scientists indicate that the fields of émigré professors may be a good proxy for the fields of a broader flow of German Jewish émigrés, which caused the observed increase in U.S. invention.

REFERENCES

Aho, Alfred V. 1990. "Algorithms for finding patterns in strings." In van Leeuwen, Jan. *Handbook of Theoretical Computer Science*, volume A: Algorithms and Complexity. The MIT Press. pp. 255–300.

Ambrose, Tom, *Hitler's Loss. What Britain and America Gained from Europe's Cultural Exiles* (London, UK: Peter Owen's Publishers, 2001).

Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang. 2010. "Superstar Extinction." *Quarterly*

Journal of Economics, vol. 125, no. 2, pp. 549-589.

Borjas, George, J. “Self-Selection and the Earnings of Immigrants,” *American Economic Review*, 77, No. 4 (1987), 531-553.

Borjas, George J., and Kirk D. Doran, “The Collapse of the Soviet Union and the Productivity of American Mathematicians,” *The Quarterly Journal of Economics,* 127, 3, (2012), 1143-1203.

Borjas, George J., and Kirk D. Doran, “Cognitive Mobility: Native Responses to Supply Shocks in the Space of Ideas,” *Journal of Labor Economics*, forthcoming.

Cohen, Wesley M., Richard R. Nelson, and John P. Walsh “Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not),” NBER Working Paper No. 7752, 2000.

Deichmann, Ute, “The Expulsion of Jewish Chemists and Biochemists from Academia in Nazi Germany,” *Perspectives on Science*, 7, No.1 (1999), 1-86.

Deichmann, Ute, *Flüchten, Mitmachen, Vergessen—Chemiker und Biochemiker in der NS-Zeit* (Weinheim: Wiley-VCH Verlag, 2001).

Deichmann, Ute, “Émigré Influence in Industry,” in *Germany and the Americas: Culture, Politics, and History: A Multidisciplinary Encyclopedia*, ed. Thomas Adam (Santa Barbara, CA: ABC-CLIO, 2005).

Duden. 2000. Familiennamen, cited after http://www.peterdoerling.de/Geneal/Nachnamen_100.htm.

Emergency Committee in Aid of Displaced German Scholars**,** *List of Displaced German Scholars* (London: Notgemeinschaft Deutscher Wissenschaftler im Ausland, 1936).

Hall, Bronwyn H., Adam Jaffe, and Manuel Trajtenberg, “Market Value and Patent Citations: A First Look,” NBER Working Paper No. 7741, 2000.

Hunt, Jennifer, “Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa,” *Journal of Labor Economics,* 29, No. 3 (2011), 417-457.

Hunt, Jennifer, and Marjolaine Gauthier-Loiselle, “How Much Does Immigration Boost Innovation?” *American Economic Journal: Macroeconomics*, 2, No. 2 (2010), 31-56.

Hounshell, David H. and John Kenly Smith, Jr. Science and Corporate Strategy, Du Pont R&D, 1902-1980. Cambridge University Press, Cambridge U.K. 1988.

Imbens, Guido and Joshua Angrist, *"Identification and Estimation of Local Average Treatment Effects", Econometrica*, 62, no. 2 (1994), 467-475

Kalender der Deutschen Universitäten und Hochschulen (Leipzig: J.A. Barth Verlag, 1933).

Kerr, William R., and William F. Lincoln, “The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention,” *Journal of Labor Economic*s, 28, No. 3 (2010), 473-508.

Kröner, Peter, *Vor Fünfzig Jahren - Die Emigration deutschsprachiger Wissenschaftler 1933 – 1939,* edited by Gesellschaft für Wissenschaftsgeschichte Münster (Wolfenbüttel: Heckners Verlag, 1983).

Kürschners Deutscher Gelehrtenkalender (Berlin: W.D. Gruyter Verlag, 1931).

Lai, Ronald, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming, 2011, "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)", http://hdl.handle.net/1902.1/15705 UNF:5:9kQaFvALs6qcuoy9Yd8uOw== V1 [Version]

Möller, Horst, “Wissenschaft in der Emigration – Quantitative und geographische Aspekte” *Berichte zur Wissenschaftsgeschichte,* 7 (1984), 1-9.

Moser, Petra, “How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World Fairs,” *American Economic Review,* 95, No. 4 (September 2005), 1214-1236.

Moser, Petra, “Innovation without Patents – Evidence from World’s Fairs,” *Journal of Law and Economics* (April 2012).

Moser, Petra and Alessandra Voena, “Compulsory Licensing: Evidence from the Trading-with-the-Enemy-Act,” *American Economic Review,* 102, No. 1 (2012), 396-427.

Moser, Petra, Joerg Ohmstedt, and Paul W. Rhode, “Patents Citations and the Size of Patented Inventions—Evidence from Hybrid Corn,” 2013. http://ssrn.com/abstract=1888191.

Sachar, Howard, *A History of the Jews in America* (New York: Knopf Publishing Group, 1992).

Stephan, Paula E., and Sharon G. Levin, “Exceptional contributions to US science by the foreign-born and foreign-educated,” *Population Research and Policy Review*, 20 (2001), 59-79.

Strauss, Herbert A., Werner Röder, Belinda Rosenblatt, and Hannah Caplan, *International Biographical Dictionary of Central European Émigrés 1933-1945. Vol II: Arts, Sciences and Literature* (New York: K.G. Saur, 1983.

Trajtenberg, Manuel, “A Penny for Your Quotes: Patent Citations and the Value of Innovations.” *The RAND Journal of Economics*, Spring 1990, 21(1), 172-187.

Waldinger, Fabian, “Quality Matters: The Expulsion of Professors and the Consequences for PhD Student Outcomes in Nazi Germany,” *Journal of Political Economy*, 118, No. 4 (2010), 787-831.

Waldinger, Fabian, “Peer Effects in Science – Evidence from the Dismissal of Scientists in Nazi Germany,” *The Review of Economic Studies*, 79, No. 2 (2012), 838-861.

Waldinger, Fabian, “Bombs, Brains, and Science: The Role of Human and Physical Capital for the Creation of Scientific Knowledge,” mimeo, University of Warwick, 2013

TABLE 1– SUMMARY STATISTICS: U.S. PATENTS BY DOMESTIC INVENTORS ACROSS USPTO CLASSES

	(1) All Classes	(2) Classes with 1920-70 patents by U.S. émigrés	(3) Classes without 1920-70 patents by U.S. émigrés	(4) Classes with pre-1933 patents by dismissed	(5) Classes without pre-1933 patents by dismissed
Patents by U.S. inventors 1920-70	2,073,771	771,377	1,302,394	619,308	1,454,463
Number of classes	166	60	106	48	118
Mean class age in 1932	87.23	84.6	88.7	87.4	87.3
P-value of equality of means test		0.085		0.929	
Mean # of foreign patents in 1932	0.93	0.92	0.93	0.70	1.01
P-value of equality of means test		0.942		0.216	
Mean patents per class and year 1920-70	244.95	252.08	240.92	252.99	241.69
Mean patents per class and year 1920-32	193.39	149.25	218.38	157.50	207.99
Mean patents per class and year 1933-70	262.59	287.26	248.63	285.65	253.21

Notes: Data include patent – main class combinations of U.S. inventors in classes with 1920-1970 patents by German university chemists. Patents by U.S. inventors in these classes were collected from www.uspto.gov. Dismissed and émigré professors are identified from the List of Displaced German Scholars (1937), Deichman (2001), Kröner (1983), and Straus (1983).

TABLE 2 – ORDINARY LEAST SQUARES REGRESSIONS
DEPENDENT VARIABLE IS PATENTS PER CLASS AND YEAR BY U.S. INVENTORS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Émigré class * Post	105.222***	91.712***	84.803***	75.439***				
	(22.203)	(19.212)	(18.950)	(19.326)				
# émigré patents * Post					5.848*	4.992*	4.527**	3.991**
					(3.058)	(2.561)	(2.182)	(1.956)
# foreign patents	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Quadratic class age	No	No	Yes	Yes	No	No	Yes	Yes
Patent pools	No	No	No	Yes	No	No	No	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
R-squared	0.783	0.845	0.849	0.851	0.779	0.842	0.846	0.848

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 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 group. 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 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 3 – FIRST STAGE AND REDUCED FORM

DEPENDENT VARIABLES ARE ÉMIGRÉ CLASS*POST (COLS 1-2), # OF ÉMIGRÉ PATENTS * POST (COLS 3-4), AND PATENTS PER CLASS AND YEAR BY U.S. INVENTORS (COLS 5-8)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	First Stage				Reduced Form			
	Émigré class * Post		# Émigré patents * Post					
Dismissed class * Post	0.370***	0.339***			80.821***	57.752***		
	(0.081)	(0.079)			(23.155)	(19.436)		
# dismissed patents * Post			1.384***	1.303***			35.595***	22.330***
			(0.442)	(0.435)			(6.547)	(6.339)
# foreign patents	No	Yes	No	Yes	No	Yes	No	Yes
Quadratic class age	No	Yes	No	Yes	No	Yes	No	Yes
Patent pools	No	Yes	No	Yes	No	Yes	No	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
R-squared	0.801	0.809	0.770	0.773	0.779	0.849	0.782	0.849
F-statistic	20.80	18.25	9.79	8.99				

Standard errors clustered at the class level in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: In first stage regresssions (columns 1-4), the dependent variables are *Émigré class * Post* (columns 1 and 2) and *# émigré patents * Post* (columns 3 and 4). *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. *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. 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 class age* is the 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. In reduced form regressions (columns 5-8) the dependent variable measures patents by U.S. inventors per USPTO class and year, excluding patents by émigrés.

TABLE 4 - INSTRUMENTAL VARIABLE REGRESSIONS
DEPENDENT VARIABLE IS PATENTS PER CLASS AND YEAR BY U.S. INVENTORS

	(1)	(2)	(3)	(4)
Émigré class * Post	218.707***	170.136***		
	(60.614)	(57.992)		
# émigré patents * Post			25.717***	17.137**
			(8.750)	(6.909)
# 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
Observations	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: 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 5 – SUMMARY STATISTICS: U.S. PATENTS BY DOMESTIC INVENTORS WHO WERE ACTIVE PRIOR TO 1933

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
		Fraction of patents in research fields of émigrés			Fraction of pre-1933 patents in research fields of dismissed chemists		
	All Inventors	<50%	50%	>50%	<50%	50%	>50%
Total inventors active before 1933	210,410	144,647	7,842	57,921	155,261	4,719	50,430
Annual probability of patenting 1920-70	0.035	0.034	0.050	0.036	0.035	0.067	0.034
Annual probability of patenting 1920-32	0.098	0.098	0.120	0.097	0.098	0.161	0.094
Annual probability of patenting 1933-70	0.014	0.013	0.026	0.015	0.013	0.036	0.013
Patents per inventor and year 1920-70	0.043	0.042	0.055	0.045	0.043	0.084	0.042
Patents per inventor and year 1920-32	0.112	0.111	0.132	0.111	0.111	0.184	0.107
Patents per inventor and year 1933-70	0.020	0.018	0.029	0.023	0.019	0.050	0.019

Notes: Data include 210,410 U.S. patentees with at least one patent between 1920 and 1932. We constructed data on patents per year of these patentees through a search algorithm, which identified patents by individual inventors per class and year, using Google's *Patent Grant Optical Character Recognition (OCR) Text (1920-1979)* database. The Appendix includes a detailed description of the search algorithm and the process of data cleaning.

TABLE 6- ORDINARY LEAST SQUARES AND INSTRUMENTAL VARIABLES, DEPENDENT VARIABLE IS PATENTING BY U.S. INVENTORS THAT WERE ACTIVE BEFORE 1933

	(1)	(2)	(3)	(4)
	OLS (Linear Probability)		IV	
Share of patents in émigré classes * Post	0.002**	-0.007***	-0.001	-0.022***
	(0.001)	(0.002)	(0.002)	(0.006)
Quadratic time to first patent	No	Yes	No	Yes
Quadratic time since first patent	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Inventor fixed effects	Yes	Yes	Yes	Yes
Observations	10,730,910	10,730,910	10,730,910	10,730,910
R-squared	0.045	0.147	-	-

Standard errors clustered at the level of an inventor's main class

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: The dependent variable equals one if inventor *i* obtains at least one patent in year *t,* and 0 otherwise. The sample includes all domestic U.S. patentees with at least one patent between 1920 and 1932. *Share of patents in émigré classes* measures the total share of patents by a U.S. inventor that are in the 60 research fields of émigrés. The variable *Post* equals 1 for years after the dismissals. *Quadratic time to first patent* is a second-degree polynomial for years until an inventor patents for the first time in any of the 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 the 166 classes.

TABLE 7 – FIRST STAGE AND REDUCED FORM

	(1)	(2)	(3)	(4)
	First Stage		Reduced Form	
Share of pre-1933 patents in dismissed classes * Post	0.403***	0.402***	-0.0003	-0.009***
	(0.086)	(0.086)	(0.001)	(0.002)
Quadratic time to first patent	No	Yes	No	Yes
Quadratic time since first patent	No	Yes	No	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Inventor fixed effects	Yes	Yes	Yes	Yes
Observations	10,730,910	10,730,910	10,730,910	10,730,910
F-statistic	21.83	21.65		
R-squared	0.434	0.434	0.045	0.147

Standard errors in parentheses clustered at the level of an inventor's main class of patenting

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Notes: In columns 1-2 the dependent variable *Share of patents in émigré classes * Post* measures the total share of patents by a U.S. inventor that are in the 60 research fields of émigrés. In columns 3-4, the dependent variable equals 1 if inventor *i* obtains at least one patent in year t, and 0 otherwise. *Share of pre-1933 patents in dismissed classes* measures the share of a domestic U.S. inventor's pre-1933 patents that are in 48 classes with pre-1933 patents of dismissed chemists.

TABLE 8 – SUMMARY STATISTICS ON ENTRY OF NEW PATENTEES ACROSS RESEARCH FIELDS

	(1) All Classes	(2) Classes with 1920-70 patents by U.S. émigrés	(3) Classes w/o 1920-70 patents by U.S. émigrés	(4) Classes with pre-33 patents by dismissed	(5) Classes w/o pre-33 patents by Dismissed
Number of classes	166	60	106	48	118
Panel A: Entrants into research fields:					
Total entrants into classes 1920-1970	1,396,318	499,417	896,901	404,927	991,391
Mean entrants per class and year 1920-70	164.9	163.2	165.9	165.4	164.7
Mean entrants per class and year 1920-32	153.8	116.1	175.1	121.6	166.8
Mean entrants per class and year 1933-70	168.8	179.3	162.8	180.4	164.0
Panel B: Entrants into patenting:					
Total entrants (no prior patents) 1920-1970	964,526	327,224	637,302	268,084	696,442
Mean entrants (no prior patents) per class and year 1920-70	113.9	106.9	117.9	109.5	115.7
Mean entrants (no prior patents) per class and year 1920-32	125.0	92.0	143.8	97.2	136.3
Mean entrants (no prior patents) per class and year 1933-70	110.1	112.1	109.0	113.7	108.7

Notes: Entrants are patentees who patent for the first time in one of 166 research fields, defined at the level of USPTO technology classes. To collect these data we developed an algorithm that matches inventors across classes and years, and assigns a unique identifier to each inventor. See the Data Appendix for a detailed description. We apply this algorithm to the full text of 1,365,689 U.S. patent documents in Google's *Patent Grant Optical Character Recognition (OCR) Text (1920-1979)* database across 166 technology classes between 1920 and 1970. This yields 1,396,318 entrants who patented for the first time in class *c* and 964,526 entrants without prior patents in the 166 technology classes. Dismissed and émigré professors are identified from the List of Displaced German Scholars (1937), Deichman (2001), Kröner (1983), and Straus (1983).

TABLE 9 - ORDINARY LEAST SQUARES AND INSTRUMENTAL VARIABLES REGRESSIONS
DEPENDENT VARIABLE IS NUMBER OF ENTRANTS PER YEAR

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS				Instrumental Variables			
	Entrants into field		Entrants into patenting		Entrants into field		Entrants into patenting	
Émigré class * Post	73.799***	58.181***	53.434***	43.967***	162.287***	142.119***	116.707***	109.466***
	(15.674)	(14.715)	(12.522)	(12.261)	(44.195)	(45.982)	(34.565)	(37.863)
# foreign patents	No	Yes	No	Yes	No	Yes	No	Yes
Quadratic class age	No	Yes	No	Yes	No	Yes	No	Yes
Patent pools	No	Yes	No	Yes	No	Yes	No	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
R-squared	0.781	0.835	0.763	0.805	0.767	0.824	0.750	0.792

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

Notes: In columns 1-2 and 5-6, the dependent variable is number of new patentees per year in class *c* without prior patents in class *c*. In columns 3-4 and 7-8, the dependent variable is number of new patentees per year in class *c* with neither prior patents in class *c* nor prior patents in any other of the 166 classes. *Émigré class* equals 1 for classes that include at least one U.S. patent by an émigré. Classes without émigré patents form the control group. We instrument with *Pre-1933 Dismissed class * Post* for *Émigré class * Post. Pre-1933 Dismissed class* equals 1 for classes that include at least one pre-1933 U.S. patent by a dismissed chemist. The dummy variable *Post* equals 1 for years after the dismissals. First stage regressions are reported in column 2 of Table 3. *# 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 10 – PATENTING OF CO-INVENTORS, AND CO-INVENTORS OF CO-INVENTORS OF ÉMIGRÉ CHEMISTS

	Patents in 166 technology classes		
	Patents assigned only to 60 émigré classes	Patents assigned only to 106 non- émigré classes	Patents assigned to both émigré and non- émigré classes
Panel A: Co-inventors of senior émigrés:			
1920-1932	4	4	0
1933-1970	469	24	75
1920-1970	473	28	75
Panel B: Co-inventors of co-inventors of senior émigrés:			
1920-1932	48	59	24
1933-1970	1,103	162	264
1920-1970	1,151	221	288

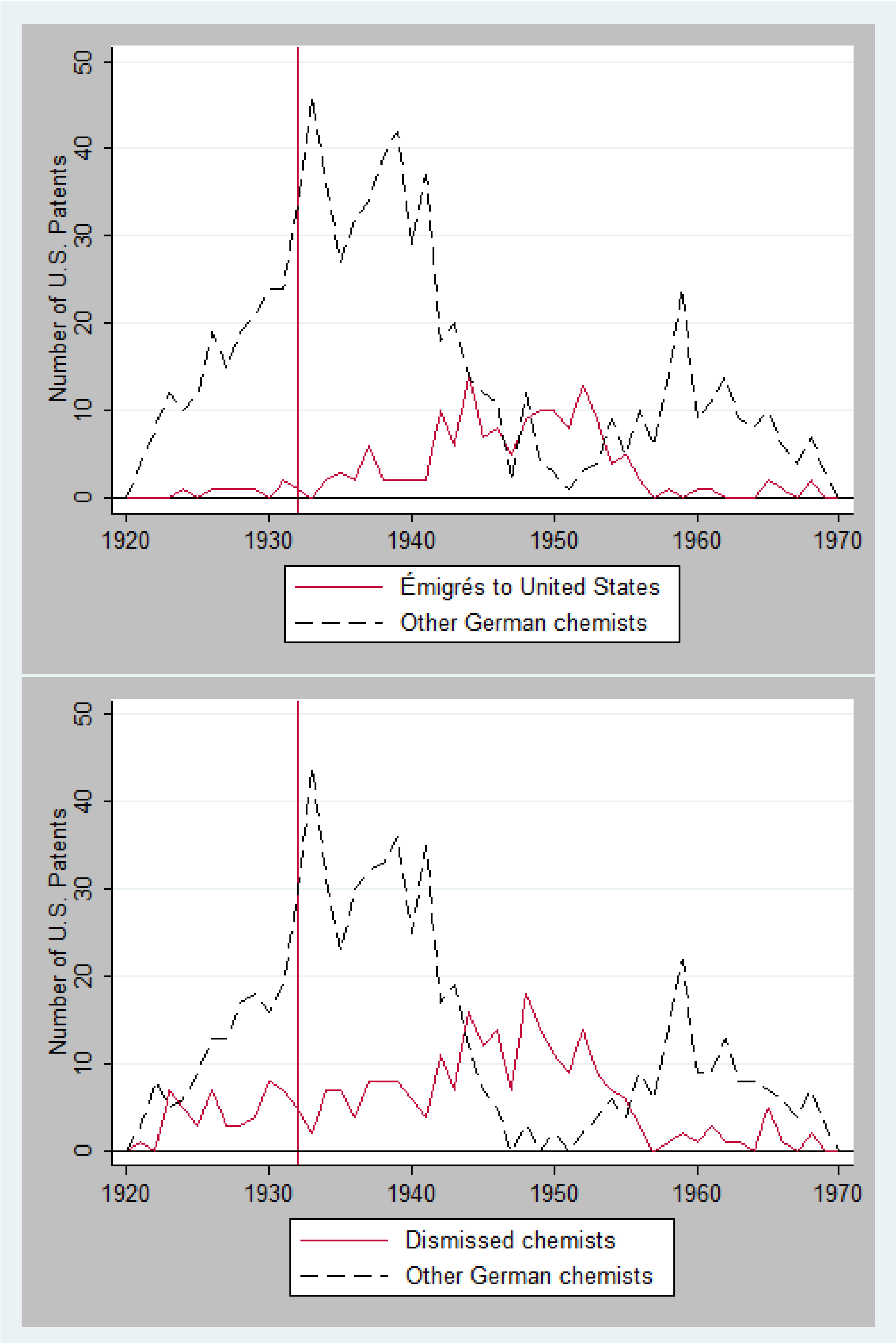
Notes: Data for Panel A include include co-inventors of senior émigrés; which we identified from the list of inventors on patent grants. Data for Panel B include co-inventors of co-inventors (2nd degree co-inventors) of senior émigrés. Data on 1920-1970 patents of co-inventors were hand-collected from Google Patents (www.patents.google.com). Data on 1920-1970 patents of co-inventors of co-inventors were collected with an algoritm using the inventor data.

TABLE 11 – PATENTING OF YOUNG ÉMIGRÉ CHEMISTS

	Patents in 166 technology classes		
	Patents assigned only to 60 émigré classes	Patents assigned only to 106 non- émigré classes	Patents assigned to both émigré and non- émigré classes
1920-1932	6	0	0
1933-1970	113	22	34
1920-1970	119	22	34

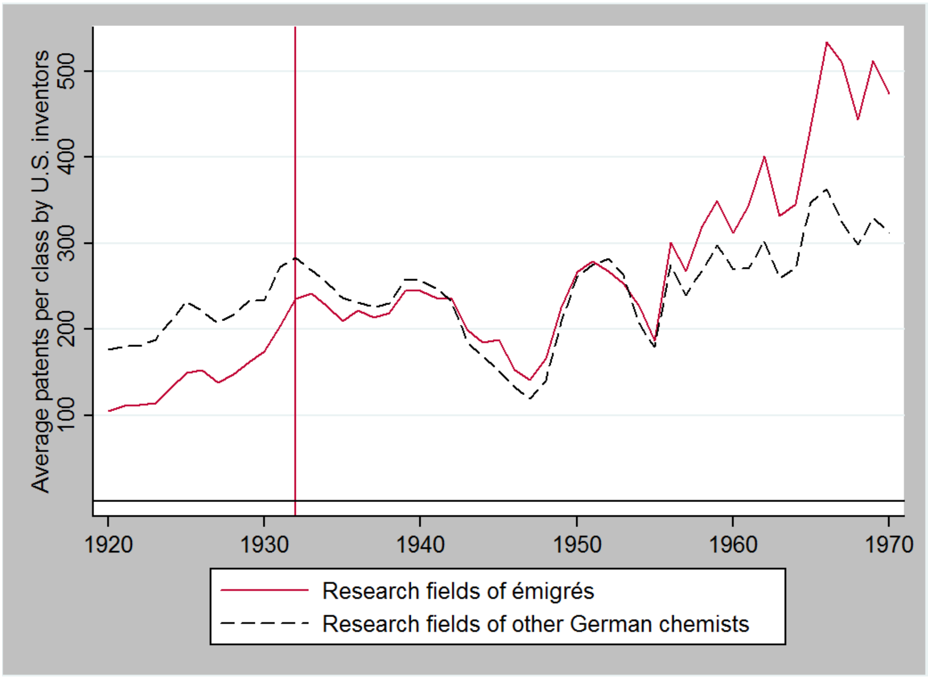
Notes: Data include young émigré chemists as listed in Straus (1983). Data on 1920-1970 patents were collected from Google Patents (www.patents.google.com).

FIGURE 1 – U.S. PATENTS PER YEAR BY GERMAN CHEMISTS



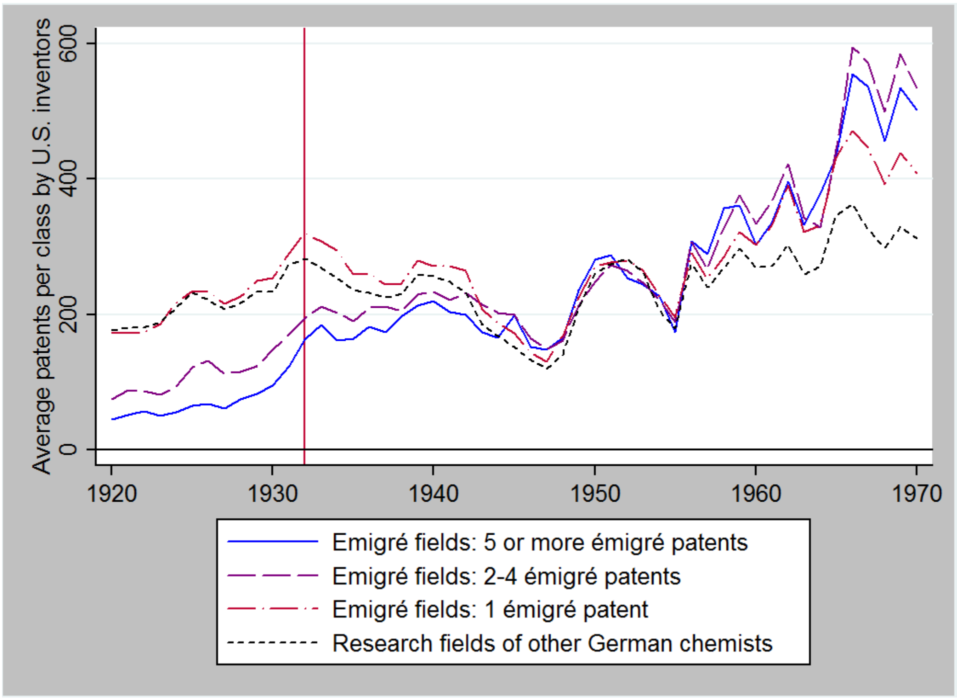
Notes: Data cover 946 U.S. patents by 535 professors and lecturers of chemistry at German and Austrian universities. 1933 is the year of the first dismissals. The top panel shows patent issues per year for chemists who emigrated to the United States; these data include 157 U.S. patents by *émigrés to the United States*. The bottom panel presents patent issues per year for *dismissed chemists*; these data include 282 U.S. patents. We collected U.S. patents per years for émigrés and dismissed chemists from Google Patents (www.patents.google.com).

FIGURE 2A –U.S. PATENTS PER CLASS AND YEAR BY DOMESTIC U.S. INVENTORS IN RESEARCH FIELDS OF ÉMIGRÉS AND OTHER GERMAN CHEMISTS



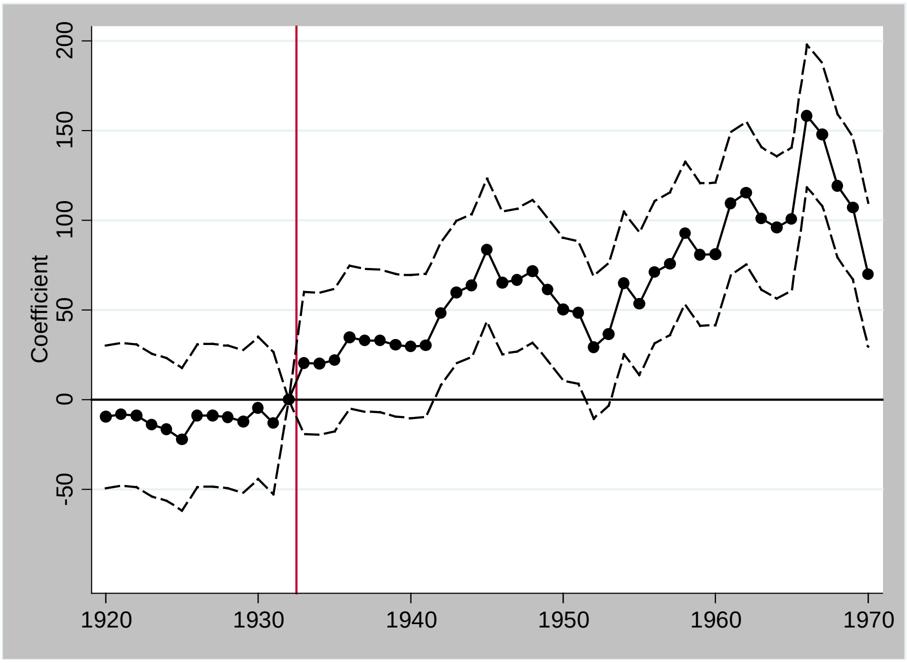
Notes: Data cover 2,073,771 patent – main class combinations by U.S. inventors across 166 research fields defined at the level of USPTO classes. *Research fields of émigrés* cover 60 classes that include at least one patent between 1920 and 1970 by a German or Austrian émigré to the United States. *Research fields of other German chemists* cover 106 USPTO classes that include at least one patent between 1920 and 1970 by another German chemist but include no patents by émigrés.

FIGURE 2B –U.S. PATENTS PER CLASS AND YEAR BY DOMESTIC U.S. INVENTORS IN RESEARCH FIELDS OF ÉMIGRÉS AND OTHER GERMAN CHEMISTS



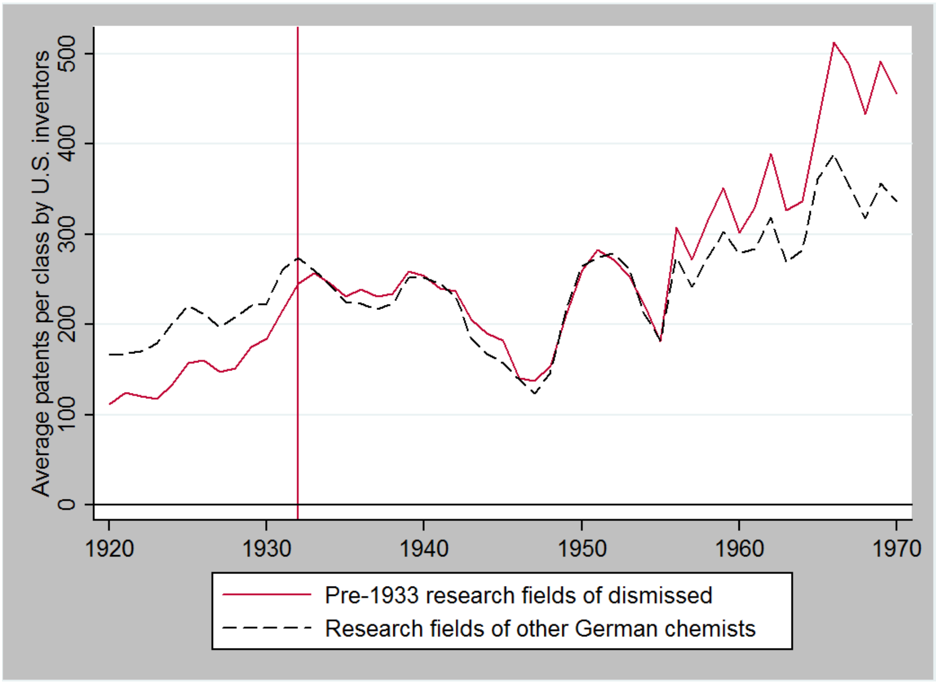
Notes: Data cover 2,073,771 patent – main class combinations across 166 research fields defined at the level of USPTO classes. *Émigré fields: 5 or more émigré patents* include classes that include 5 or more patents between 1920 and 1970 by German or Austrian émigrés to the United States. *Émigré fields: 2-4 émigré patents* include classes that include 2 to 4 émigré patents. *Émigré fields: 1 émigré patent* include classes that include 1 émigré patent. *Research fields of other German chemists* cover 106 USPTO classes that include at least one patent between 1920 and 1970 by another German chemist but include no patents by émigrés.

FIGURE 3 – YEAR-SPECIFIC OLS ESTIMATES
U.S. PATENTS PER YEAR IN RESEARCH FIELDS OF ÉMIGRÉS



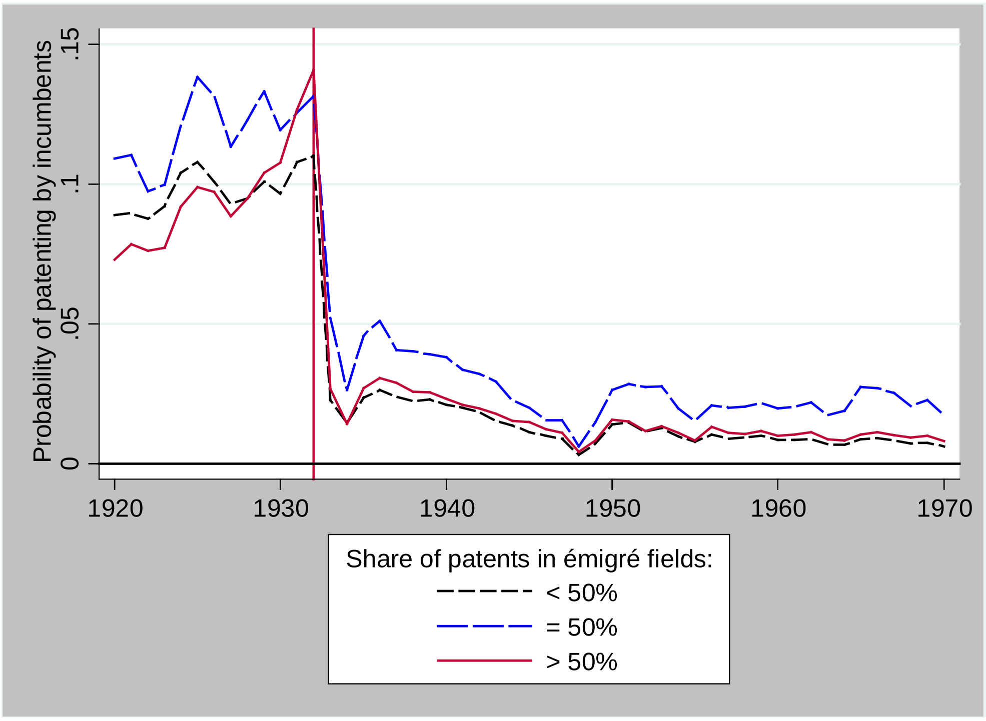
Notes: Coefficients $ß_t$ in the regression Patents by U.S. inventors$_{c,t}$ = $\alpha_0 + \sum_{t=1920}^{1970} ß_t \textit{émigré class}_c \cdot year_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$, where the dependent variable measures U.S. patents issued to U.S. inventors per class and year, and the variable émigré class$_c$ equals 1 for research fields of émigrés. The variable $year_t$ represents an indicator variable for each year between 1920 and 1970, and 1932 is the excluded category. 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. Patents by émigré chemists are excluded from the counts of U.S. inventors. Standard errors are clustered at the level of research fields.

FIGURE 4 –PATENTS BY DOMESTIC INVENTORS IN RESEARCH FIELDS
IN WHICH DISMISSED CHEMISTS WERE ACTIVE BEFORE 1933



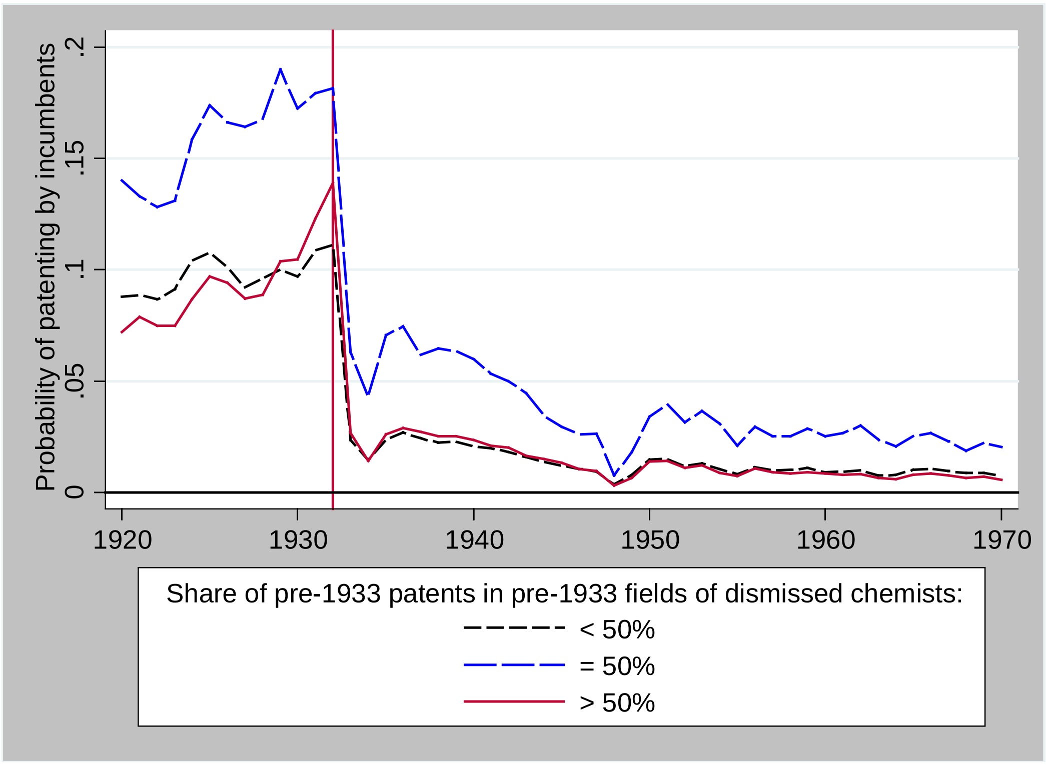
Notes: Data cover 2,073,771 patent – main class combinations by U.S. inventors across 166 research fields defined at the level of USPTO classes. *Pre-1933 research fields of dismissed chemists* cover 48 classes that include at least one patent between 1920 and 1932 by a dismissed chemist. *Research fields of other German chemists* cover 118 USPTO technology classes that include at least one patent by another German chemist, but include no pre-1933 patents by dismissed chemists.

FIGURE 5 –PATENTING PER YEAR BY INCUMBENT INVENTORS IN RESEARCH FIELDS OF ÉMIGRÉS COMPARED WITH FIELDS OF OTHER GERMAN CHEMISTS



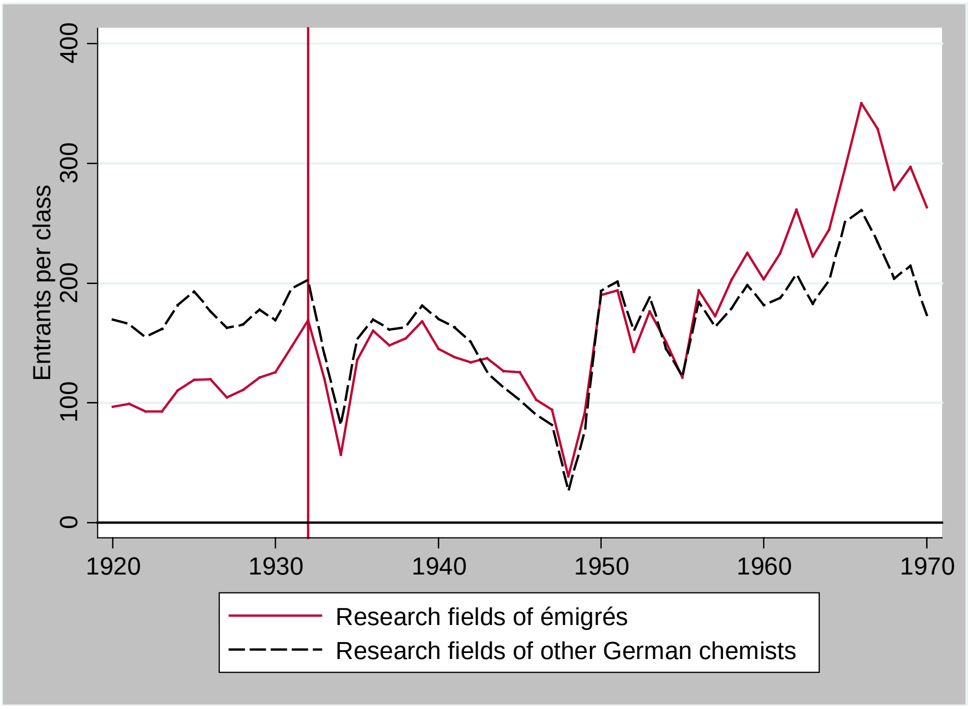
Notes: *Probability of patenting by incumbents* measures the average probability of patenting per year by 210,410 inventors who patented at least one invention before 1933. *Share of patents in émigré fields* measures the share of all patents (1920-1970) by an individual inventor that are in a class with at least one patent by an émigré.

FIGURE 6 –PATENTING PER YEAR BY INCUMBENT INVENTORS IN PRE-1932 FIELDS OF DISMISSED CHEMISTS COMPARED WITH FIELDS OF OTHER GERMAN CHEMISTS



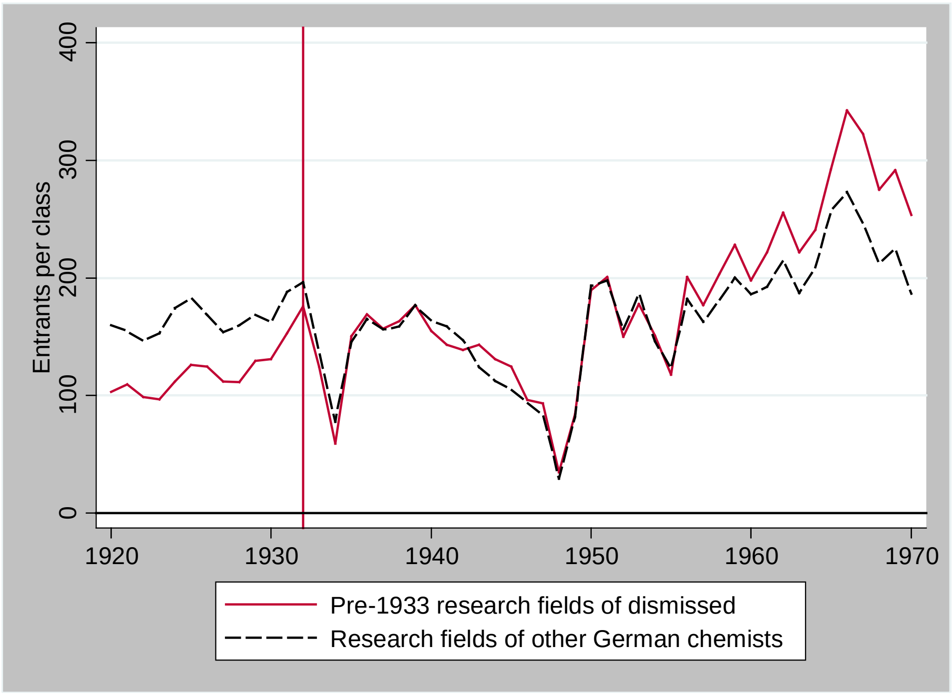
Notes: *Probability of patenting by incumbents* measures the average probability of patenting per year by 210,410 inventors who patented at least one invention before 1933. *Share of patents in pre-1933 fields of dismissed chemists* measures the share of pre-1933 patents (1920-1932) by an individual inventor that are in a class with at least one pre-1933 patent by a dismissed chemist.

FIGURE 7 –ENTRY OF U.S. PATENTEES INTO RESEARCH FIELDS OF ÉMIGRÉS COMPARED WITH FIELDS OF OTHER GERMAN CHEMISTS



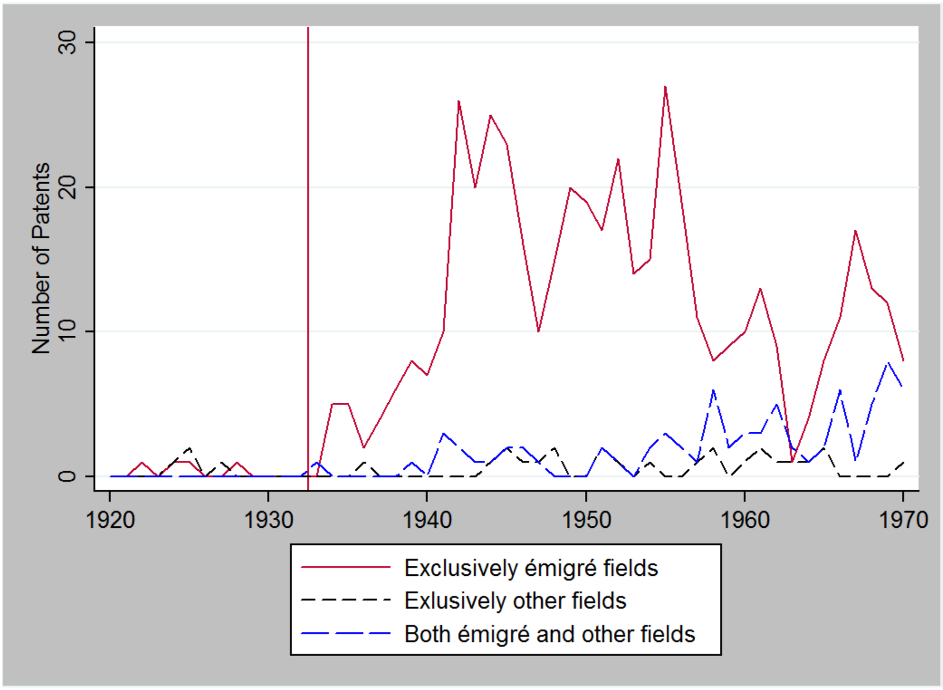
Notes: *Entrants per class* measures the number of new researchers that entered the average research field in year *t*. Entry into a research field is defined by the first patent of an inventor in a patent class. *Research fields of émigrés* consist of 60 USPTO classes that include at least one patent between 1920 and 1970 by a German or Austrian émigré to the United States. *Research fields of other German chemists* cover 106 USPTO classes that include at least one patent between 1920 and 1970 by another German chemist but include no patents by émigrés.

FIGURE 8 – ENTRY OF U.S. PATENTEES INTO RESEARCH FIELDS WITH PRE-1933 PATENTS OF DISMISSED COMPARED WITH RESEARCH FIELDS OF OTHER GERMAN CHEMISTS



Notes: *Entrants per class* measures the number of new researchers that entered the average research field in year *t*. Entry into a research field is defined by the first patent of an inventor in a USPTO patent class. *Pre-1933 research fields of dismissed* consist of 48 USPTO classes that include at least one patent between 1920 and 1932 by a dismissed German or Austrian chemist. *Research fields of other German chemists* cover 118 USPTO classes that include at least one patent by a German chemist but no pre-1933 patent by a dismissed chemist.

FIGURE 9 – PATENTS BY CO-INVENTORS OF ÉMIGRÉS



Notes: Total patents by 47 co-inventors of émigrés. *Exclusively émigré fields* measures the number of patents per year that were exclusively assigned to émigré fields. *Exclusively other fields* measures the number of patents per year that were exclusively assigned to other fields. *Both émigré and other* fields measures the number of patents per year that were assigned to both émigré and other fields. We collected the U.S. patents of co-inventors from Google patents (www.patents.google.com).

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.

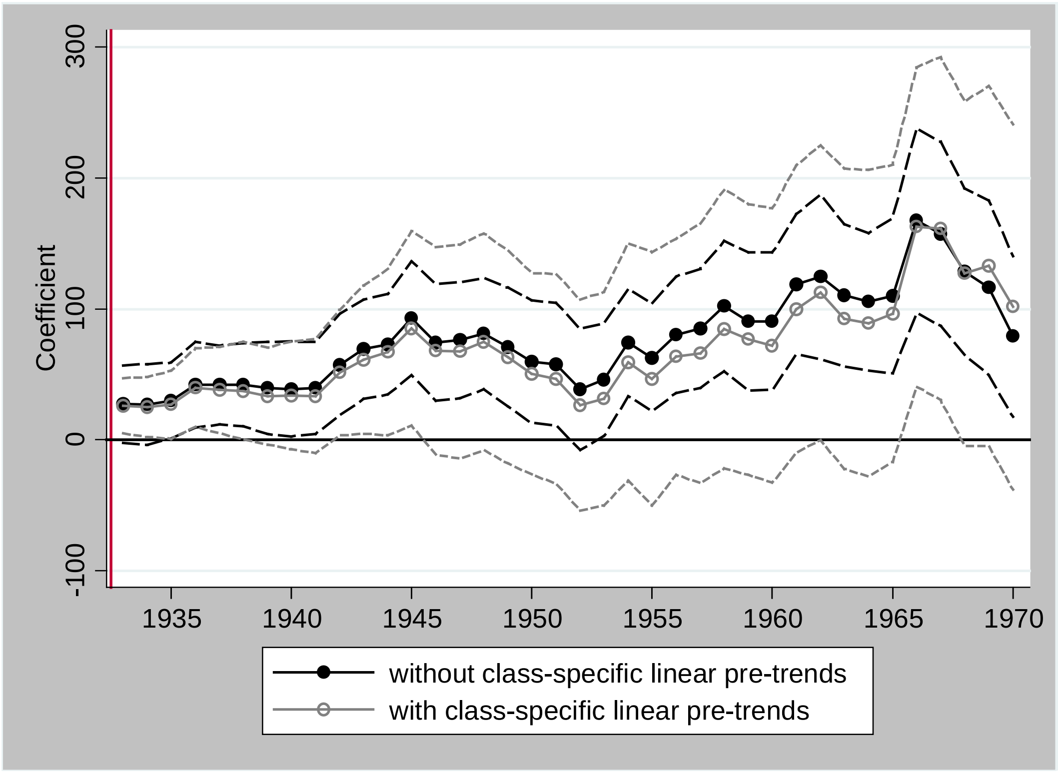
Ruggles, Steven, and Matthew Sobek et. al., Integrated Public Use Microdata Series: Version 2.0 (Minneapolis: Historical Census Projects, University of Minnesota, 1997).

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.

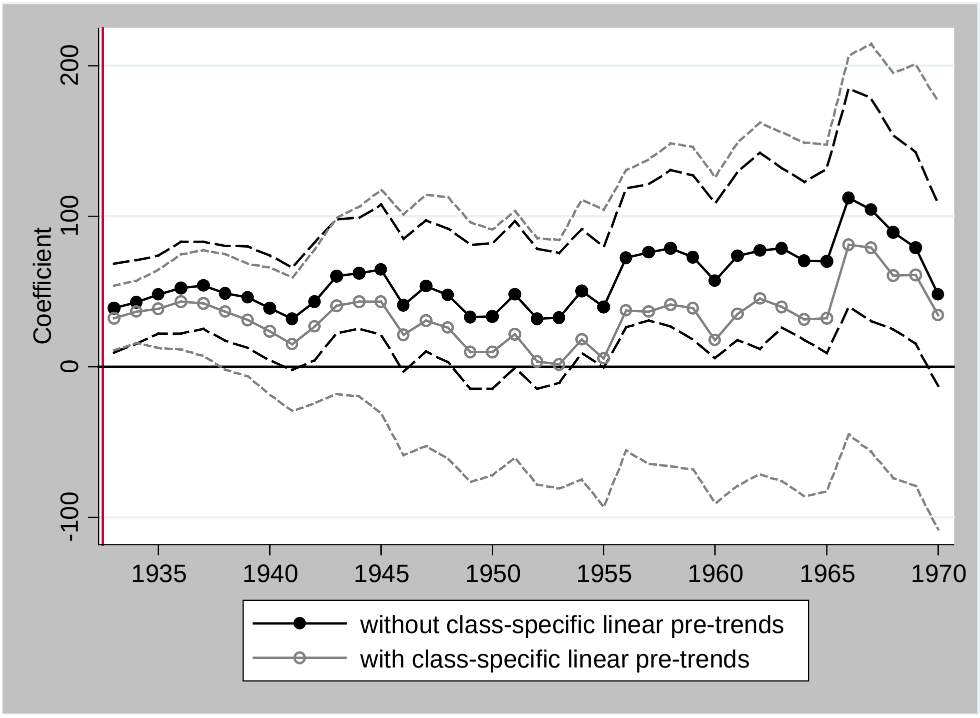
Appendix Figures

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



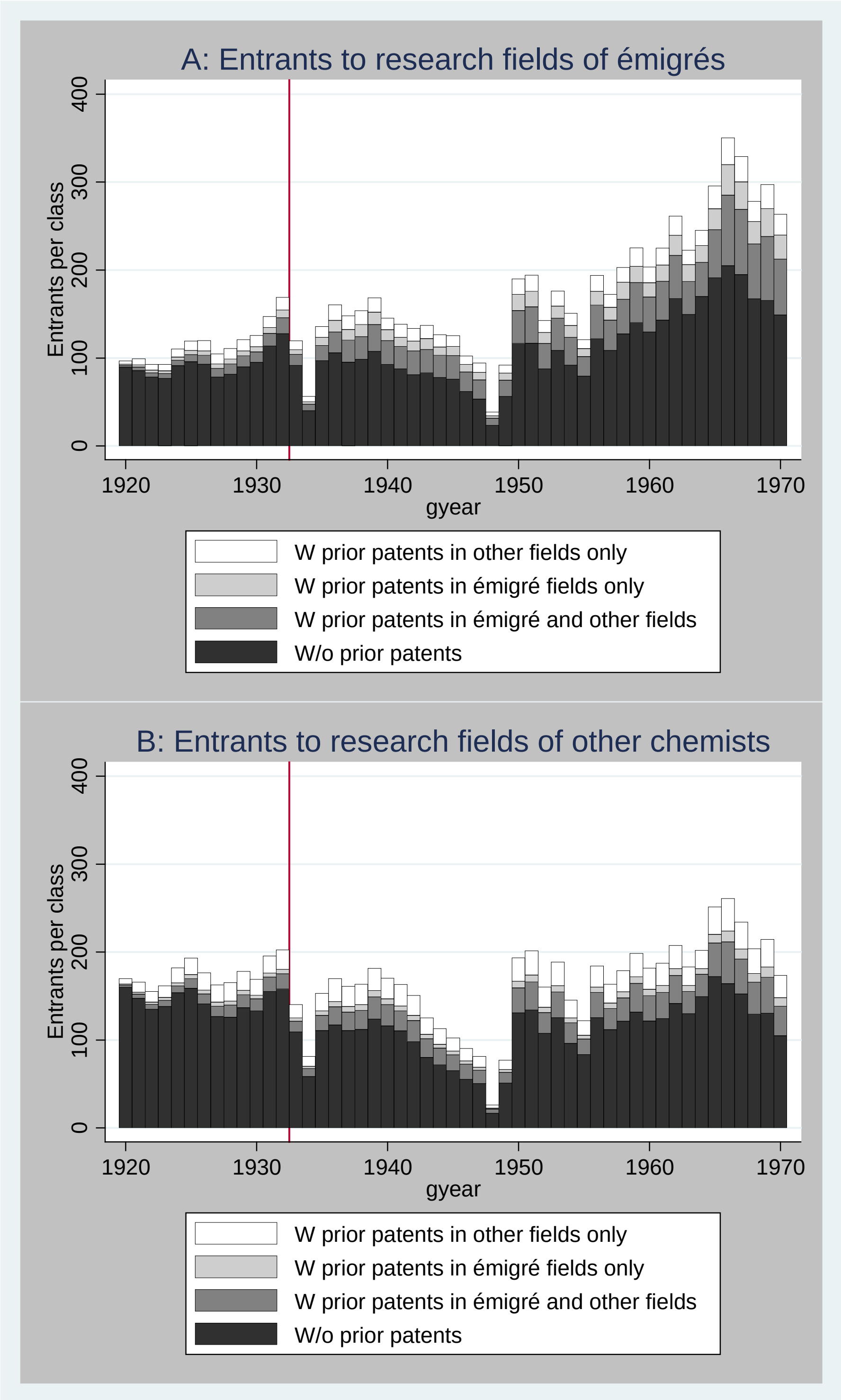
Notes: Time-varying estimates without class-specific linear pre-trends estimate $ß_t$ in the regression *Patents by U.S. inventors*$_{c,t} = \alpha_0 + \sum_{t=1933}^{1970} ß_t \, \acute{e}migr\acute{e}\ class_c \cdot year_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$ where $year_t$ is a 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*$_{c,t} = \alpha_0 + \sum_{\tau=1933}^{1970} ß_\tau \, \acute{e}migr\acute{e}\ class_c \cdot year_\tau + \eta_c \cdot t + \delta_t + f_c + \upsilon_{c,t}$. 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)



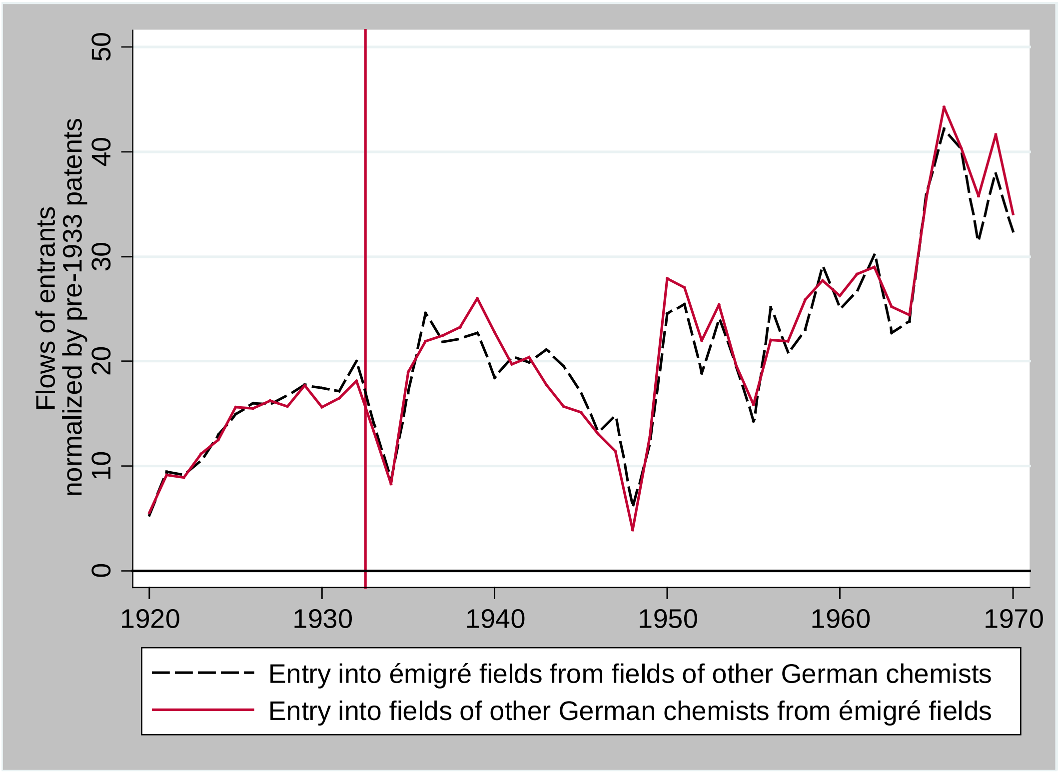
Notes: Time-varying estimates without class-specific linear pre-trends estimate $ß_t$ in the regression *Patents by U.S. inventors*$_{c,t}$ = $\alpha_0 + \sum_{t=1933}^{1970} ß_t\, pre-1933\ dismissed\ class_c \cdot year_t + \gamma' X_{c,t} + \delta_t + f_c + \varepsilon_{c,t}$ where $year_t$ is a set 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 Citations-weighted dependent variable		IV	
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.