The Return(s) of Superstar Scientists

Michela Giorcelli, Yujing Huang, Bo Li*

April 30, 2026

Abstract

This paper studies the long-run effects of high-skill return migration on scientific productivity and institutional development, using evidence from US-trained Chinese scientists during the Great Return Wave of the 1950s. Exploiting quasi-random variation in return opportunities generated by Cold War–era US exit restrictions, we compare the careers and research trajectories of scholars who returned to the People's Republic of China with those who remained in the United States. Combining newly digitized historical archives with multilingual bibliometric data, we trace individual scientific output, research agendas, and institutional spillovers over several decades. We show that return migration had persistent effects on domestic knowledge formation, field development, and intergenerational talent formation, highlighting the central role of human capital and know-how transfers in early-stage nation building.

*Contact information: Giorcelli: University of California, Los Angeles, NBER, CEPR, CESifo, IZA, and J-PAL, 9262 Bunche Hall, 315 Portola Plaza, Los Angeles, CA 90095, USA; email: mgiorcelli@econ.ucla.edu. Huang: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles CA 90095, USA; email: yujingh@g.ucla.edu. Li: Guanghua School of Management, Peking University; email: boli@gsm.pku.edu.cn. Fonda Hu and Leon Wang provided excellent research assistance. We thank seminar participants at UCLA GEM seminar and 11th Monash-Warwick-Zurich-CEPR Text-as-Data Workshop for helpful comments and discussion.

1 Introduction

Human capital lies at the core of scientific knowledge creation and long-run economic growth (Romer, 1990, 1994). The allocation of top scientific talent across countries and institutions shapes where frontier ideas are produced, how research agendas evolve, and which innovations ultimately diffuse. A large literature on high-skill migration has shown that this allocation matters: the emigration of elite scientists can generate persistent losses for sending countries, while inflows of foreign talent can substantially boost innovation in receiving economies (Waldinger, 2010, 2012, 2016; Borjas and Doran, 2012; Moser et al., 2014). Yet far less is known about the consequences of return migration. When superstar scientists move back to their home countries, they may fundamentally reshape local knowledge systems, affecting not only their own productivity, but also the direction of research, the diffusion of frontier ideas, and the formation of domestic scientific capacity.

Despite the prevalence and importance of this phenomenon, empirical evidence on the causal and long-run effects of high-skill return migration remains scarce. The production and diffusion of scientific knowledge are inherently endogenous: top scientists sort into frontier fields and more productive universities or sectors, while political and institutional environments shape where ideas are generated and transmitted. Moreover, it is difficult to track individual scientists across countries, languages, and publication systems and to systematically measure their research output and agendas, particularly over long horizons.

This project examines how foreign education shaped the career trajectories of individual scholars and their long-run contributions to institutional scientific development, using evidence from Chinese scholars who studied in the United States and faced sharply divergent return opportunities during the Great Return Wave of the 1950s. Following the establishment of the People's Republic of China in 1949, the new government launched a nationwide campaign to modernize and industrialize the economy and actively mobilized overseas-educated Chinese to "return and serve the motherland and the people." One key example is Qian Xuesen (Tsien, Hsue-shen), a Caltech-trained aerospace engineer who returned to China in 1955 and went on to become a leading architect of the country's missile and nuclear weapons programs. These return efforts emphasized advanced scientific and technical training as a critical input for nation-building and institutional capacity formation.

Our first contribution is a comprehensive new dataset, collected and digitized from multiple historical archives, that reconstructs the universe of US-trained Chinese scholars between 1929 and 1955 and tracks their educational trajectories, return decisions, institutional placements, and scientific output over several decades. We combine annual directories of Chinese students in the United States, doctoral dissertation records, declassified U.S.

government documents on exit permit decisions, biographical archives of returnees, and multilingual bibliometric databases (OpenAlex and CNKI). These sources allow us to construct scholar-level panels of publications, citations, coauthorship networks, leverage Natural Language Processing (NLP) methods to measure research agendas spanning both English- and Chinese-language outlets, and to link individual scholars to the evolution of domestic academic institutions and disciplines in China.

Our identification strategy exploits the fact that return migration during the Great Wave was disrupted by sharp and historically contingent political constraints. U.S. security policies and foreign travel controls during the early Cold War, most notably the *1951 Prevention of Departure of Aliens with Scientific and Technical Training*, selectively restricted exit for Chinese scholars in fields deemed to have military or industrial value. As a result, when diplomatic relations between the United States and China deteriorated, some scholars were able to return promptly, while others faced exogenous delays or were prevented from returning despite willingness to do so. In our empirical analysis, we compare the outcomes of these groups over 40 years. These groups had statistically similar educational backgrounds, fields of specialization, and pre-return trajectories, and were exposed to comparable research environments prior to the policy shock.

At the individual level, we examine differences in the trajectories of academic interests, influence, and scientific productivity between returnees and stayers, with particular attention to whether returnees adjusted their research agendas in response to state-led industrial priorities, ideological campaigns, or political incentives, and whether they contributed to China's domestic knowledge base in distinct ways. At the institutional level, we study the broader effects of return migration as a form of high-skill know-how transfer, investigating whether the presence of US-trained returnees catalyzed the emergence of new disciplines, accelerated institutional capacity building, and promoted intergenerational knowledge transmission through mentorship and department formation. By assessing both short- and long-run spillovers in scientific productivity, talent formation, and innovation, we further examine whether fields or institutions that initially benefited from returnees continued to exhibit stronger outcomes decades later, even in the presence of major political disruptions such as the Cultural Revolution.

This project contributes to understanding how cross-national flows of talent, often disrupted by geopolitical events, shape not only the individuals involved, but also the broader developmental trajectories of institutions and scientific ecosystems. By examining this historical episode of return migration and knowledge transfer under political duress, the study offers insights into present-day debates on international education, scientific collaboration, and national security. In particular, recent restrictions on student visas, heightened scrutiny

of Chinese researchers, and growing tensions in US–China relations exactly echo the mid-20th century dynamics we analyze. In this way, the project brings historical analysis to bear on issues of contemporary relevance, shedding light on how state policies and geopolitical rivalries continue to shape the global circulation of knowledge.

Contribution to the Literature. This paper contributes to the literature on the determinants of knowledge production and spillovers by providing new evidence on how return migration of foreign-trained scientists reshapes home-country knowledge systems. Prior work has shown that international migration affects innovation through brain drain, diaspora networks, and displacement shocks, largely focusing on frontier economies and outward mobility (Aghion and Howitt, 1992; Borjas and Doran, 2012; Docquier and Rapoport, 2012; Moser et al., 2014; Kerr et al., 2016; Waldinger, 2012, 2016; Iaria et al., 2018). Much less is known about the effects of return migration on domestic research capacity, particularly in economies with weak preexisting scientific institutions and at an early stage of industrialization and scientific development. This paper fills that gap by studying how returnees generate persistent increases in knowledge production by rebuilding departments, introducing frontier research practices, and creating local spillovers that extend beyond individual productivity.

Second, the paper advances the literature on upper-tail human capital and the production of "useful knowledge" by documenting a multiplier role for elite scientists in early-stage research environments. While growth models emphasize aggregate human capital and idea accumulation (Romer, 1990, 1994; Mokyr, 2005a,b), and recent work highlights the importance of scientific elites (Squicciarini and Voigtländer, 2015; Giorcelli, 2019; Bianchi and Giorcelli, 2020; Giorcelli and Li, 2021), this paper shows that top scientists matter disproportionately because they shape institutional learning. Returnees act as organizational anchors: they train junior scholars, establish research norms, and attract complementary talent. These mechanisms amplify the long-run impact of upper-tail human capital in settings where marginal additions to expertise relax binding constraints on experimentation and mentoring.

Finally, the paper contributes to the literature on institutions, politics, and nation building through foreign education by linking return migration to the endogenous direction of research. Existing work demonstrates that political incentives and institutional constraints shape scientific agendas (Jia et al., 2019, 2024; Acemoglu et al., 2024), while foreign education can transmit political and organizational norms that affect long-run development (Spilimbergo, 2009; Kung and Wang, 2025). This paper bridges these literatures by showing how foreign-trained returnees mediate the interaction between state priorities and scientific capacity. Returnees transmit technical and organizational knowledge acquired abroad, en-

abling states to redirect research toward strategic objectives while simultaneously expanding domestic research capability. The findings clarify how international human capital flows contribute to institutional development and long-run state capacity.

2 The Great Return Wave

2.1 Calling on Overseas-Educated Chinese for Nation-Building

Following the establishment of the People's Republic of China (PRC) in 1949, the new government confronted an economy that was overwhelmingly agrarian, technologically backward, and institutionally underdeveloped (Lardy, 1995). With limited domestic capacity in modern science and engineering, the state identified overseas-educated Chinese as a critical source of advanced human capital for industrialization and nation-building.

Beginning in late 1949, the PRC launched a coordinated campaign encouraging foreign-trained Chinese to return, commonly framed as a call to "serve the motherland and the people." The government established the Committee on Returned Students Affairs in December 1949 to facilitate return, providing travel subsidies, reception services, and centralized job placement in universities, research institutes, and state-owned enterprises (Zhang and Mo, 2018). These policies reflected a deliberate strategy to rapidly internalize frontier knowledge acquired abroad.

At the time, the stock of overseas Chinese students was sizable. In 1949, approximately 5,541 Chinese students were studying abroad, the product of earlier state-sponsored and private efforts to train elite students overseas (China Institute in America, 1954; Li, 2000). About 70 percent were enrolled in the United States, with the remainder primarily in the United Kingdom and Japan, and more than two-thirds pursued degrees in STEM fields.

2.2 Cold War Tensions and Barriers to Return

This initial momentum was disrupted by the rapid deterioration of relations between the PRC and the United States. After 1949, the United States refused to recognize the Communist government, severing diplomatic ties and leaving Chinese nationals without consular protection. The outbreak of the Korean War in 1950 further intensified security concerns and anti-communist sentiment.

Against this backdrop, U.S. authorities increasingly viewed Chinese students, especially those with advanced scientific and technical training, as potential national security risks. In October 1951, the United States implemented administrative restrictions on the departure of aliens with scientific and technical training. In particular, U.S. immigration authorities

formally prohibited Chinese students and scholars from departing the United States without authorization, with violations punishable by fines of up to $5,000, imprisonment of up to five years, or both, reflecting security concerns over the transfer of sensitive scientific and technical knowledge to Communist China (Li, 2000). Declassified intelligence erports emphasized the risk that returnees could transfer knowledge relevant to military, nuclear, and industrial programs (see Appendix for more details).

These policies generated substantial frictions to return. Archival records document interrogations by federal authorities, prolonged surveillance, repeated denials of exit permits, and pressure to adjust immigration status or naturalize (Han, 1993). While some students eventually returned after extended delays, others were prevented from leaving altogether despite documented intentions to do so. As a result, return outcomes during the early 1950s were shaped less by individual choice than by historically contingent political constraints.

2.3 The Geneva Conference

By the mid-1950s, the status of Chinese students in the United States became entangled with broader diplomatic negotiations. During the Geneva Talks of 1954–1955, Chinese students were discussed alongside American prisoners held in China, effectively becoming bargaining instruments in bilateral negotiations (Han, 1993). At this stage, many students were no longer viewed by U.S. authorities as strategically indispensable and had become administratively and politically costly.

The talks culminated in a joint announcement on September 10, 1955, affirming the right of nationals residing in the other country to return home if they wished. In principle, this agreement removed formal barriers to departure. In practice, return flows remained limited. U.S. records indicate that more than half of applicants withdrew their requests, reflecting persistent administrative obstacles, uncertainty, and the passage of time (Han, 1993). In parallel, the Refugee Relief Act of 1953 facilitated adjustment to permanent residence for many Chinese nationals, further reducing incentives to return.

By the late 1950s, return migration had largely ceased. Chinese departures from the United States fell sharply, reaching only two in fiscal year 1958 (Kung, 1973). Archival evidence summarized by Wang (2015) indicates that, despite encouragement or, in some cases, force to stay from the U.S. side and subsequent political contestation from China side, the majority of Chinese students and scientists trained abroad ultimately remained in the United States. Of an estimated cohort of roughly 5,000 individuals, about 4,000 stayed, including Chen-Ning Yang and Tsung-Dao Lee, who later received the 1957 Nobel Prize in Physics, while approximately 1,200 successfully contested U.S. restrictions and returned to

China during the 1950s. Also, internal reports from Beijing Municipal Archives document that 1,954 Chinese students and scientists returned to mainland China between 1949 and 1958, of whom 1,244 (roughly 64 percent) returned from the United States. A majority of these returnees specialized in the natural sciences, broadly defined to include engineering, with the remainder concentrated in the social sciences or other fields. Combined with domestic political turmoil in China, these developments effectively ended the Great Return Wave until diplomatic normalization and the resumption of overseas study after 1978. The cohort shaped by this episode provides the institutional setting for our empirical analysis.

3 Data

In this section, we describe the data that we collected and digitized from several historical archives. Additional details can be found in the Appendix.

3.1 Directory of Chinese Students in the United States

To construct a comprehensive dataset of Chinese students who pursued higher education in the United States during the first half of the 20th century, we digitize and extract individual-level information from a series of historical directories compiled by Chinese student organizations and educational institutions. These directories, published annually or biennially, were produced by associations such as the Chinese Students' Alliance, the Chinese Students Handbook Company, the Chinese Students' Christian Association, and later the China Institute in America. The directories span from 1911 to 1953, including the *Directory of the Chinese Students in the United States of America*, *The handbook of Chinese students in U.S.A.*, and the *Directory of Chinese Students in Colleges and Universities in the United States*. While the directories vary in format, we harmonize the data to create a unified structure that captures detailed information on students' names, universities attended, degrees obtained, and fields of study.

3.2 Dictionary of Chinese Scholars Who Studied Abroad and Returned to China

We manually collected and digitized individual-level data from three retrospective biographical compilations documenting Chinese scholars who studied abroad and returned to China between the 1840s and early 1960s. The first source, the *Dictionary of Chinese Scholars Who Returned from North America in the 1950s*, contains detailed records on 1,806 individuals,

including their educational background, return dates, and post-return career trajectories. The second source, *the Comprehensive Survey of Scientific and Technical Personnel Studying Abroad and Returning to China, 1840–1949*, covers 3,500 early returnees and draws on biographical books, archival records, and institutional lists. The third source, *Scientific and Engineering Manpower in Communist China, 1949–1963*, profiles 1,200 prominent scientists and engineers using a broad range of official publications, yearbooks, and scientific periodicals. Together, these sources provide high-coverage, individual-level data on the timing, fields of study, and institutional affiliations of early returnees.

We use this dataset to identify individuals who pursued graduate or postgraduate studies abroad and returned to China in the late 1940s and 1950s, either voluntarily or under geopolitical pressure. In addition to biographical details, the data include information on disciplinary specialization, institutional roles, and leadership positions within Chinese universities and research institutes. This historical sample allows us to characterize the early cohort of returnee scientists and compare them to individuals who remained abroad.

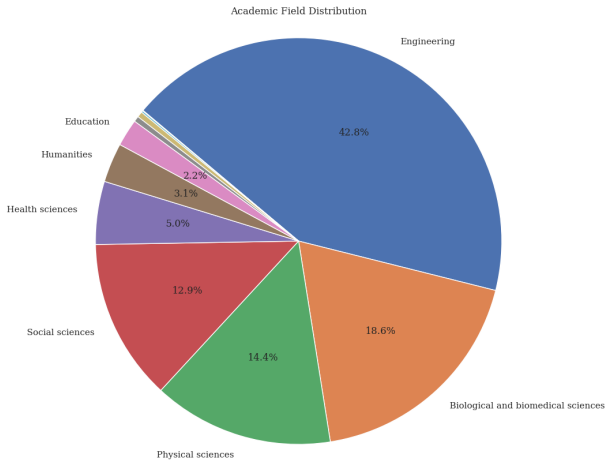


Figure 1: Distribution of study for the 6,068 returnees.

3.3 Measures of Individual-Level Scientific Productivity and Research Agendas

We construct annual, individual-level measures of scholarly output by combining bibliometric, patent, and archival sources. Our measures capture not only the *volume* and *influence* of output, but also the *orientation* of research agendas, following methods developed in recent work on the economics of science and innovation.

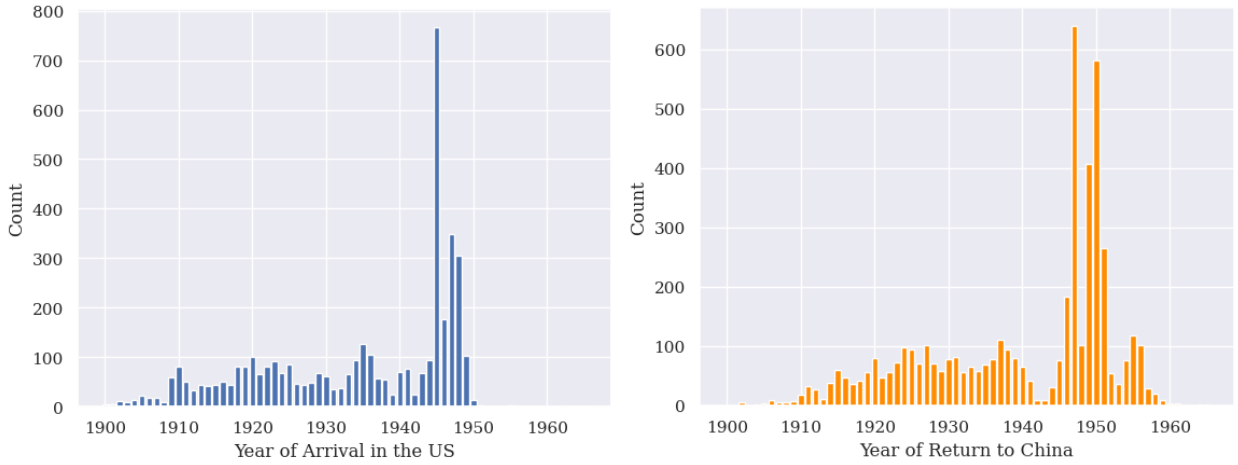


Figure 2: Year of Arrival in the US and Year of Return to China for the 6,068 returnees.

Publications. We aggregate publications from two complementary databases. OpenAlex, the successor of the discontinued Microsoft Academic Graph, provides global coverage with standardized metadata and citation linkages (as in Moser et al., 2025), while the China National Knowledge Infrastructure (CNKI) captures nearly all Chinese-language academic literature. For CNKI, we follow the construction in Acemoglu et al. (2024), collecting titles, abstracts, authors, publication years, venues, and citation counts. For each scholar, we query under three naming conventions (Chinese characters, Hanyu Pinyin, Wade–Giles) to address historical variation in name usage. We then apply a series of filters to manually disambiguate records using information on institutional affiliations, research fields, and years of birth and death, ensuring that each publication is uniquely and correctly assigned to a scholar in our database and that spurious matches are excluded. From these data we extract: (i) the total number of publications; (ii) field-normalized citation counts to adjust for disciplinary differences (see, e.g., Moser et al., 2014); (iii) the share of coauthors based in China, to capture domestic integration; and (iv) the share of publications in state-prioritized research areas, identified from contemporary industrial policy documents.

Field Classification and Research-Agenda Vectors. To characterize the *content* and *evolution* of research, we construct scholar-level research-agenda vectors using multilingual text embeddings. We adapt methods from Bello et al. (2025) and Koffi (2025), combining English- and Chinese-language bibliometric sources to assemble a comprehensive corpus of publication titles and abstracts for each scholar–year. We first harmonize author identities, publication years, and venues across databases, and preprocess text by lowercasing, de-accenting, removing stopwords, and tokenizing. We then represent research content using state-of-the-art multilingual sentence embeddings. Specifically, we embed each paper's title

and abstract using `SentenceTransformers`, which map texts in different languages into a shared semantic space. For each scholar i in year t, we define an *agenda vector* as the mean embedding of all publications authored in that year. This vector provides a continuous, high-dimensional representation of the scientist's research orientation. For robustness, we compare results to alternative representations based on WORD2VEC and latent semantic analysis (LSA) applied to TF–IDF–weighted text.

To define research fields, we adopt an unsupervised clustering approach inspired by Moser et al. (2025). We represent scientific output using text-based embeddings and apply k-means clustering to group research outputs into fields, selecting the number of clusters using standard diagnostics (elbow, silhouette, and gap statistics). Each paper is assigned uniquely to a field, enabling the construction of scholar–year field shares and dominant research fields. In additional robustness analyses, we implement alternative clustering procedures based on nonlinear dimension reduction (UMAP) and density-based clustering (HDBSCAN).

Patents. To capture *applied* output, we match scholars to patent filings in domestic and international patent databases, following the linkage approach in Moser (2016) and Moser et al. (2025). Patents complement publication-based measures by reflecting contributions to technology transfer and industrial applications.

Outcome Variables. We construct the following dependent variables for each scholar–year:

1. **Productivity/Impact:** (a) publications per year; (b) five-year forward citations; (c) field–year normalized citation percentiles; (d) top-venue share (field-specific journal/conference lists); (e) h-index within a fixed ten-year window around the return event.

2. **Agenda Outcomes (semantic):** (a) *stability*, the cosine similarity of an author's agenda vector between t and $t-1$; (b) *convergence to destination*, the similarity to the mean vector of colleagues at the Chinese host institution; (c) *retention of pre-return lines*, the similarity to the mean vector of pre-return coauthors; (d) *frontier alignment*, the similarity to the top-venue centroid within field–year (e.g., Iaria et al., 2018); (e) *breadth/focus*, the entropy or Herfindahl of field shares across works; (f) *novelty*, the share of bigrams or terms not previously observed in the author's publication history (e.g., Moscona, 2024).

Together, these measures allow us to compare not only the productivity of stayers versus returnees, but also how their research agendas shift, persist, or reorient in response to the

institutional and political environments in China.

Dual-Frontier Construction. A distinctive feature of our setting is that returnees publish in a domestic corpus (CNKI) while stayers publish in an international corpus (OpenAlex), and the two corpora point to different research frontiers. To leverage this, we construct *two* versions of every Kelly-style outcome. For each scholar's paper p in subfield s and year t, we compute backward similarity to papers in $s \times [t-5, t-1]$ and forward similarity to papers in $s \times [t+1, t+10]$ using semantic embeddings (LaBSE multilingual, with sbert-chinese as a robustness check), and then form $r \log q_p$ following Kelly et al. (2021), residualized by year. We do this within each of two reference universes:

1. *Domestic frontier* (CNKI reference): the 106,000-paper stratified CNKI sample, with subfields defined on FORD subcategories. This measures how each paper relates to the leading Chinese-language papers in its subfield.

2. *Global frontier* (OA reference): OpenAlex's 3.5-million-paper stratified reference, with each Chinese-language paper nearest-neighbor assigned to one of 244 OA subfields via centroid cosine similarity. This measures how each paper relates to the leading international papers in the same field.

Because the same scholar's papers receive both scores, we can read the gap directly: a paper that is highly aligned with the Chinese frontier (high CNKI-reference alignment) but distant from the global frontier (low OA-reference alignment) is locally important but internationally peripheral. The complementary case, alignment with the global frontier, captures *catching-up* to international science. Table 1 previews the unconditional gap; Section 6 traces how it evolves around return.

Table 1: Frontier alignment of scholar publications: unconditional means by reference universe

Comparison	Mean	SD	N
OA scholars vs OA frontier	0.391	0.152	31,500
CNKI scholars vs CNKI frontier	0.603	0.124	9,092
CNKI scholars vs OA frontier (catching-up)	0.359	0.131	9,395

Notes: Mean LaBSE cosine similarity between matched scholar publications and the breakthrough centroid in the corresponding subfield-year, reported separately for the three reference universes used in the paper. Row 1: OpenAlex publications by matched scholars compared to the OpenAlex breakthrough centroid. Row 2: CNKI publications by matched scholars compared to the CNKI breakthrough centroid (within-corpus alignment). Row 3: CNKI publications compared to the OpenAlex breakthrough centroid, the catching-up measure used in Section 6. SD denotes the cross-paper standard deviation. N is the number of paper-year observations with non-missing alignment in each row.

Validating the Embedding Space. Before using these vectors as outcome measures, we verify that they recover the substantive structure of science across both corpora. Figure 3 shows three side-by-side t-SNE projections at the field (macro-domain) level: the left panel uses only OpenAlex matched scholar papers (76,326 papers, sampled to 30,000 for tractability), the middle panel uses the full set of CNKI matched scholar papers (16,481 papers, all shown), and the right panel pools both corpora (92,807 papers, sampled to 30,000). Figure 4 repeats the exercise at the OpenAlex subfield level. In all three projections, papers from the same field or subfield cluster tightly, and substantively related categories (e.g. atomic and condensed-matter physics; electrical and mechanical engineering) occupy adjacent rather than disjoint regions of the space. The clustering structure is recovered separately within each corpus and is preserved when the two are pooled, supporting the use of LaBSE distances as the substrate for the frontier-alignment outcomes that follow.

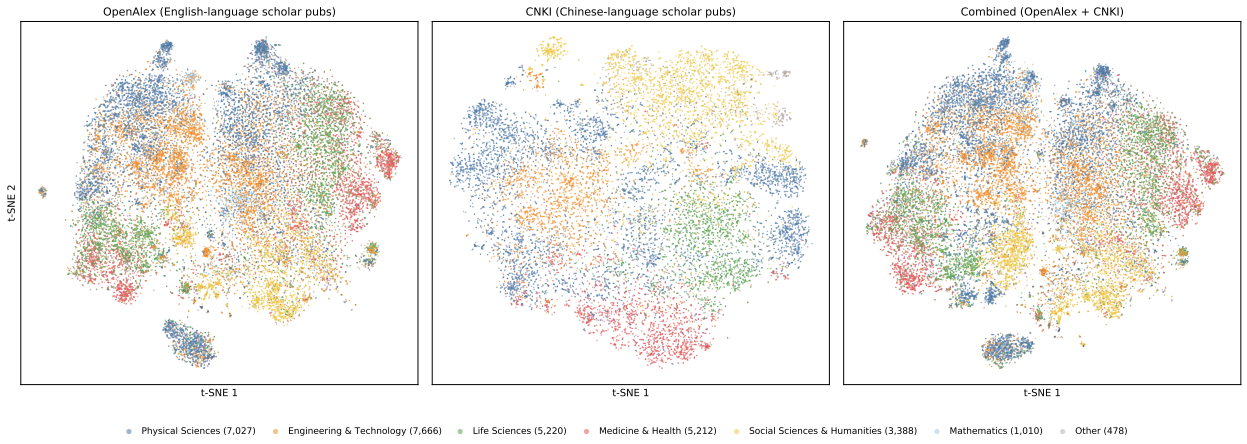


Figure 3: t-SNE map by macro-domain: OpenAlex, CNKI, combined

Notes: t-SNE projections of LaBSE paper embeddings, computed separately per corpus. Left: 30,000 OA matched scholar papers (stratified by field). Middle: all 16,481 CNKI matched papers. Right: pooled OA+CNKI, 30,000 stratified. Colors: macro-domain (CNKI FORD majors mapped to OA fields). Legend counts refer to the combined panel.

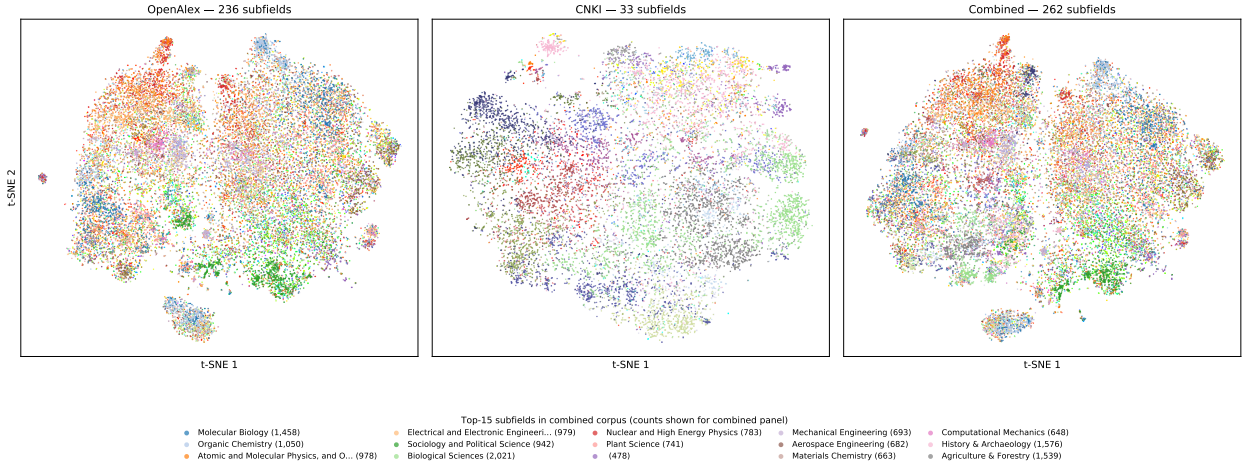


Figure 4: t-SNE map by OpenAlex subfield: OpenAlex, CNKI, combined

Notes: Same projections as Figure 3; points colored by OA subfield (every subfield assigned a distinct color; panel header reports the count). Legend lists the top-15 subfields in the combined corpus; remaining subfields are colored but unlabeled. CNKI subfields are mapped to OA subfields via centroid cosine similarity.

4 Identification Strategy

We begin with a simple cross-sectional OLS comparing the productivity of returnees and stayers up to 1976,[1]

$$y_i = \beta \, \text{Return}_i + X_i'\gamma + \mu_{f(i)} + \mu_{c(i)} + \varepsilon_i, \tag{1}$$

where y_i is individual i's scientific productivity accumulated by 1976 (e.g., total publications, citations, domestic coauthorship share, China-oriented topics), Return_i indicates returning to the PRC by 1957,[2] X_i includes pre-departure characteristics (e.g., gender, age, institutions attended), and $\mu_{f(i)}$ and $\mu_{c(i)}$ are field and cohort fixed effects (defined by decade of initial entry into U.S. study). This specification is equivalent to comparing average career outcomes of returnees and stayers with similar pre-departure individual characteristics within field–cohort cells. Although it does not address the endogeneity of return decisions, it provides a transparent baseline and motivates the need for our causal identification strategies.

Causal Identification via Permit/Denial. To address selection, we exploit quasi-random U.S. exit decisions. Among the 434 students who applied to depart between 1951–54, 310 were *permitted* while 124 were *denied* for security reasons (with a small number later leaving). Restricting the sample to these applicants, we estimate a two-stage least squares specification that uses permit status as an instrument for actual return:

$$\text{First stage:} \quad D_i = \pi_1 \, \text{Permitted}_i + \alpha_{f(i)} + \alpha_{c(i)} + \eta_i, \tag{2}$$

$$\text{Second stage:} \quad y_i = \beta_{\text{LATE}} \hat{D}_i + \alpha_{f(i)} + \alpha_{c(i)} + \xi_i, \tag{3}$$

where D_i is an indicator for being in China by 1957, y_i is lifetime productivity measured up to 1976, and $\alpha_{f(i)}$ and $\alpha_{c(i)}$ are field and cohort fixed effects. This IV design identifies β_{LATE}, the causal effect of return for compliers; that is, students whose return status was determined by U.S. security decisions.

Validation Within the Denied Group. As a robustness check, we focus on the 124 students who were initially denied exit between 1951–54. After restrictions were relaxed, some of these students eventually departed while others remained in the United States. Be-

[1] The Cultural Revolution ended in 1976, and the PRC began to normalize relations with the United States shortly thereafter. In 1978 the government launched the "reform and opening" policy, and by 1978–79 the first new cohorts of Chinese students were again sent abroad, mainly to the United States and Japan.

[2] We define returnees as those who departed for China by 1957 and stayers as those who remained in the United States thereafter. This cutoff ensures that students whose departure was delayed by U.S. restrictions but who ultimately returned are not misclassified as stayers. After 1957, very few students returned, owing to the domestic political turmoil of the late 1950s and 1960s.

Table 2: Chinese Students in the United States (1946–1954)

Item	**Number (approx.)**
1. Total number of Chinese who arrived in U.S. on student visas or as students on other visas	
1946–50	4,800
1951–54	1,500
Total	6,300
2. Departed from U.S. before June 1951 (Korean War)	800
3. Applied to depart 1951–April 1954 (restrictions in effect)	434
4. Permitted to depart 1951–54	310
5. Temporarily denied permission to depart because of U.S. security regulations 1951–April 1954	124
6. Of those in 5 above, number who wished to depart when canvassed June 1954	62
7. Of those in 6 above:	
a. Departed U.S.	15
b. Notified, but not yet departed	12
c. Not notified they could leave	35
8. Additional students applying for permission to depart since June 30, 1954 (cases in process)	36
9. Total departed U.S. 1946–54	1,100
10. Remaining in U.S. as of December 1954	5,200

Source: U.S. Department of State, "Memorandum for the President: Status of Chinese Students Who Have Been Refused Permission to Leave the United States," March 3, 1955, declassified memorandum.

cause all individuals in this group demonstrated a willingness to return by applying for exit, differences in realized return primarily reflect the external U.S. constraint rather than underlying preferences. We therefore re-estimate the baseline cross-sectional regression within this denied subsample:

$$y_i = \beta \, \text{Return}_i + X_i'\gamma + \mu_{f(i)} + \mu_{c(i)} + \varepsilon_i, \tag{4}$$

where y_i is lifetime productivity up to 1976 and Return_i indicates whether the student ultimately returned by 1957. This design focuses on a more homogeneous subset, reducing concerns about self-selection and providing an additional validation of the IV results.

Dynamics of Return. Finally, we examine how research agendas and productivity evolved *conditional on actual return*. To do so, we construct matched samples of returnees and stayers that allow us to trace event-time dynamics around the year of return. Each returnee is matched to one (or m) stayers by field, PhD cohort, and institution, with placebo $\tau = 0$ assigned to the matched stayers in the same calendar year. These designs do not use the permit/denial instrument because event-time alignment requires the realized year of return, but instead provide descriptive evidence on the mechanisms through which return reshaped careers.

For each returnee i, define $\tau = 0$ as the first year employed in China. We estimate the individual-level scientific productivity change upon return:

$$y_{i\tau} = \alpha_i + \lambda_\tau + \theta_{f(i),\tau} + \sum_{k \neq -1} \delta_k \mathbf{1}\{\tau = k\} \cdot \text{Returnee}_i + u_{i\tau}, \tag{5}$$

with $y_{i\tau} \in \{\text{pubs}, \ \text{cit_pctile}, \ \text{topvenue_share}\}$. These stacked difference-in-differences specifications document how return affected productivity trajectories relative to matched stayers with placebo event years. Similarly, we estimate the shift of research agenda:

$$\text{Sim}_{i\tau}^{(\cdot)} = \sum_{k \neq -1} \beta_k \mathbf{1}\{\tau = k\} \cdot \text{Returnee}_i + \gamma_i + \lambda_\tau + \theta_{f(i),\tau} + \varepsilon_{i\tau}, \tag{6}$$

where $\text{Sim}^{(\cdot)}$ denotes semantic outcomes such as stability, convergence, retention, and frontier alignment. Pre-trends ($k < 0$) test for parallel paths, while post-coefficients trace the dynamic reorientation of research agendas after return.

Spillovers to Domestic Scientific Production. Beyond individual careers, returned scholars plausibly reshaped China's domestic scientific system. During the early 1950s, many

U.S.-trained returnees entered fields that were nascent or effectively nonexistent in China, including modern physics, electronics, aeronautics, and applied engineering. Although returnees typically submitted preference lists for institutional placement, final assignments were determined by the central government under the planned economy. These assignments reflected national development priorities, with returnees often directed toward strategically important sectors such as defense, heavy industry, and applied sciences. After 1957, China experienced a prolonged period of international scientific isolation. With very limited access to Western frontier knowledge until diplomatic normalization and the "reform and opening" period beginning in 1978, domestic researchers faced severe constraints on exposure to global scientific advances. As a result, U.S.-trained returnees became a primary conduit through which advanced research methods, frontier topics, and tacit scientific knowledge entered the domestic academic system.

To quantify these institution-level effects, we exploit the staggered arrival of returnees across academic fields. We estimate field-level difference-in-differences specifications of the form

$$y_{jt} = \alpha_j + \vartheta_t + \delta \, \text{Post}_{jt} \times \text{Returnee}_j + \gamma X_{jt} + \varepsilon_{jt}, \tag{7}$$

where y_{jt} denotes scientific outcomes in field j at time t, such as the number of active scholars, publications, or citations. Returnee_j is an indicator for whether field j received U.S.-trained returnees (or, alternatively, the cumulative number of returnees in that field), and Post_{jt} equals one in years after the arrival of the first returnee. Field fixed effects α_j absorb time-invariant differences across disciplines, while year fixed effects ϑ_t control for aggregate shocks to the scientific system. Section 6 implements both the binary version of (7) and a continuous-cumret robustness specification.

Summary. Our empirical strategy combines four complementary designs. First, we exploit quasi-random permit/denial decisions as an instrument to identify the causal effect of return on scientific productivity. Second, we validate this approach within the denied subsample, where realized return is shaped by external constraints rather than preferences. Third, we use matched event studies to trace dynamic effects on agendas and productivity, and explore the mechanisms through which return reshaped individual careers and knowledge transmission. Fourth, we exploit the staggered arrival of returnees across academic fields to estimate the spillovers to domestic scientific production during a period of international isolation.

5 Effects on Individual Scholars

We restrict the analysis sample to scholars with return_year $\in [1940, 1957]$ and to all U.S.-based stayers, and we keep observations from age 20 onward. The baseline specification is

$$y_{it} = \alpha_i + \lambda_t + \beta \left(\text{Returnee}_i \times \text{Post}_{it}\right) + \varepsilon_{it}, \tag{8}$$

with person and year fixed effects and standard errors clustered at the person level. We estimate (8) on six outcomes: log(publications + 1) and log(citations + 1), in each case computed three ways—using OpenAlex publications only, CNKI publications only, and the sum across both corpora. OpenAlex captures internationally circulating, English-language scholarship and CNKI captures Chinese-language journal articles, so this decomposition separates effects on international publishing from effects on domestic publishing.

Table 3 reports the results. Returnees publish 4.5% less in OpenAlex after return (column 1) and 6.2% more in CNKI (column 2), with the corpus sum showing a small net increase of 1.1% (column 3). Citation impact moves the same way: a 15.8% drop on the OpenAlex side (column 4) is partly offset by a 6.8% rise on the CNKI side (column 5), leaving citations 10.3% lower in total (column 6). Single-corpus panel comparisons that observe only one of the two channels would conclude that return reduces both productivity and impact. Once both corpora are observed jointly, the productivity decline reverses sign and the impact decline shrinks substantially.

Table 3: Effect of return on publications and citations

	log(publications + 1)			log(citations + 1)		
	(1)	(2)	(3)	(4)	(5)	(6)
	OpenAlex	CNKI	Total	OpenAlex	CNKI	Total
Returnee × Post	−0.046***	0.060***	0.011**	−0.172***	0.066***	−0.109***
	(0.005)	(0.002)	(0.005)	(0.012)	(0.004)	(0.012)
Observations	274,232	274,232	274,232	274,232	274,232	274,232
Mean of dep. var.	0.162	0.028	0.190	3.402	0.132	3.534
Person fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Two-way fixed-effects estimates of equation (8). Sample: scholars with return year in [1940, 1957] and U.S.-based stayers, observations from age 20 onward. Standard errors clustered at the person level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figures 5 and 6 trace the dynamics behind the average effects in Table 3. We use 1,388

returnee-stayer pairs matched within subfield on first-publication year and pre-return mean output, all with return years in $[1940, 1957]$, and estimate

$$y_{it} = \alpha_i + \lambda_t + \sum_{k \neq -1} \beta_k \, \mathbf{1}\{\tau_{it} = k\} \cdot \text{Returnee}_i + u_{it}, \tag{9}$$

where τ_{it} is event time relative to the returnee's return year and the matched stayer is centered on the same calendar year as its returnee. Coefficients are normalized to the mean over $\tau \in [-5, -2]$ and standard errors are clustered at the matched-pair level.

The OpenAlex panel of Figure 5 shows flat pre-trends and a steady decline after return, reaching -0.05 log points by $\tau = +15$. The CNKI panel rises sharply from $\tau = 0$, plateaus near $+0.10$ by $\tau = +5$, and remains elevated thereafter. Combined, returnees publish slightly more papers than matched stayers within the first ten years of return, with the gap closing during the Cultural Revolution (1966–1976) when scientific publishing in China contracted. Citations follow the same dynamic structure with larger magnitudes (Figure 6).

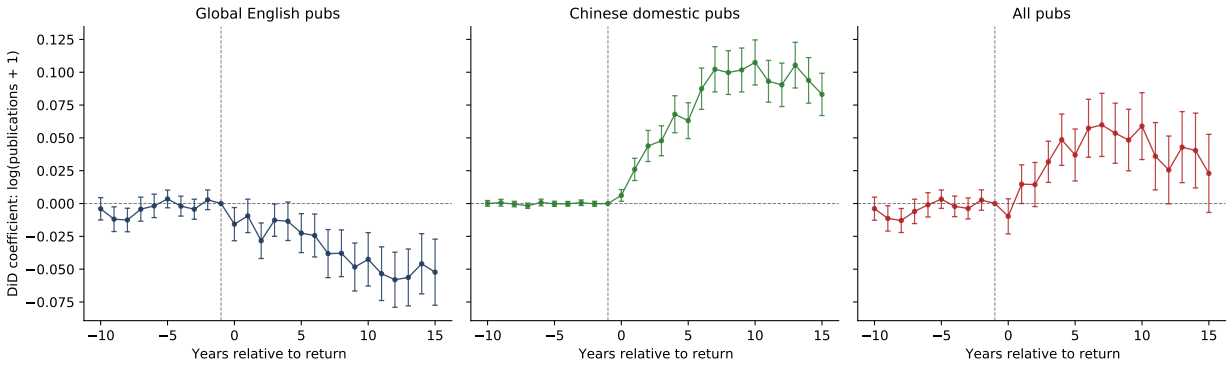


Figure 5: Event study, log(publications + 1)

Notes: Coefficients $\hat{\beta}_k$ from (9). 1,388 matched returnee-stayer pairs with return years in $[1940, 1957]$. Reference period $\tau = -1$. Bars: 95% confidence intervals from pair-clustered standard errors.

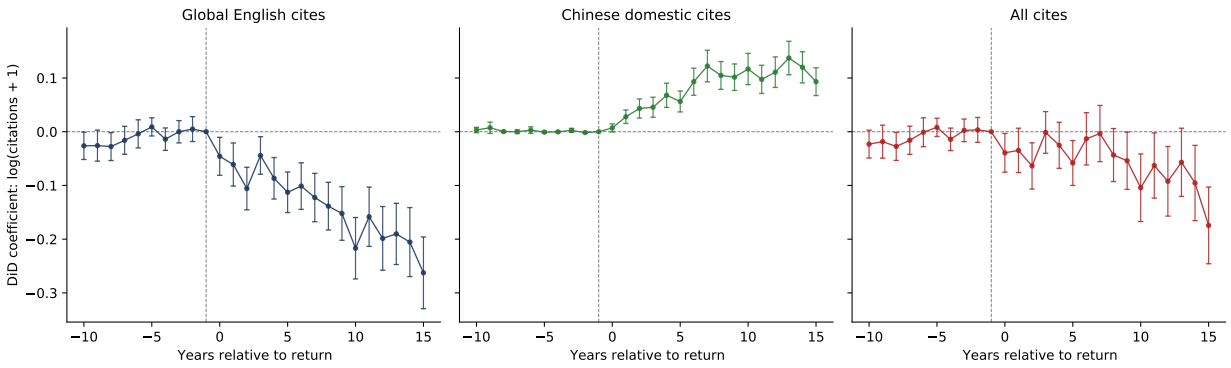


Figure 6: Event study, log(citations + 1)

Notes: Same specification as Figure 5, with log(citations + 1) as the dependent variable.

The cross-corpus reallocation could mask a quality decline if returnees published their best work in one channel and weaker work in the other, or if the average quality of their output fell after return. We re-estimate (8) replacing the count outcomes with semantic-similarity measures of paper-level frontier alignment (Kelly et al., 2021), where the reference frontier is the centroid of high-impact papers in the same subfield-year. Table 4 reports the estimates. The alignment of OpenAlex publications to the global frontier, where returnees and matched stayers are directly comparable, is essentially unchanged after return (−0.015, s.e. 0.010). The alignment of CNKI publications to the global frontier—the catching-up measure—is positive but only marginally significant (+0.061, s.e. 0.038). Returnees reallocate their publishing channel without measurably changing the alignment of their output to the international frontier.

Table 4: Effect of return on paper-level frontier alignment

	(1) OA pubs vs OA frontier	(2) CNKI pubs vs CNKI frontier	(3) CNKI pubs vs OA frontier
Returnee × Post	−0.015	−0.006	0.061*
	(0.010)	(0.051)	(0.038)
Observations	20,948	3,913	4,014
Mean of dep. var.	0.395	0.612	0.372
Person fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes

Notes: Two-way fixed-effects estimates of equation (8) on paper-level frontier-alignment outcomes from Kelly et al. (2021). The dependent variable in column 1 is the LaBSE cosine similarity between OpenAlex publications and the OpenAlex breakthrough centroid in the corresponding subfield-year; column 2 is the analogous within-CNKI alignment; column 3 is the alignment of CNKI publications to the OpenAlex global frontier (the catching-up measure). Standard errors clustered at the person level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

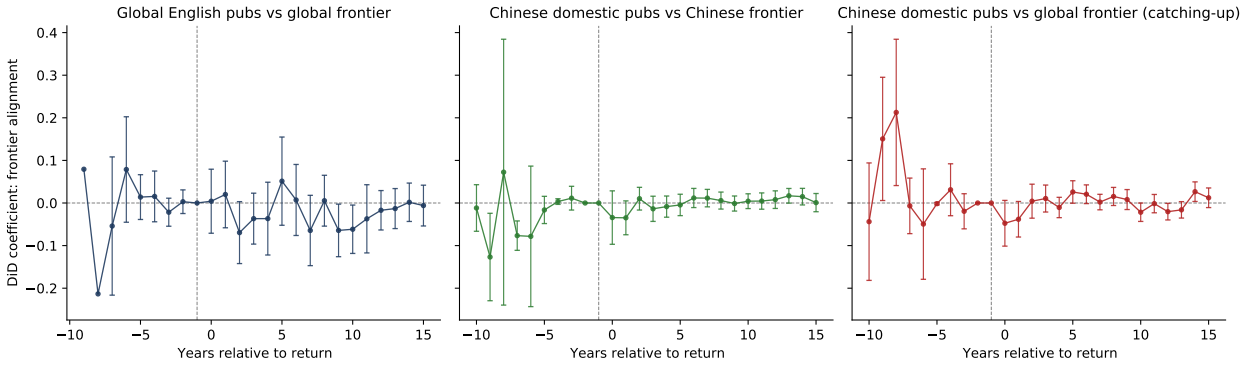


Figure 7: Frontier alignment around return

Notes: Left panel: matched DiD on the OpenAlex-side global-frontier alignment. Middle and right: within-returnee event-time means for CNKI-side alignment to the CNKI and OpenAlex frontiers, normalized to the pre-period mean over $\tau \in [-5, -2]$. The matched stayer comparison is undefined for CNKI outcomes because U.S.-based stayers do not publish in CNKI. 1,388 matched pairs.

6 Effects on Domestic Scientific Production

Returnees do not publish less in total or produce lower-quality work after return. We turn next to whether their arrival pulled the broader Chinese scientific system closer to the in-

ternational frontier. We implement equation (10) at the subfield-year level: Returnee_j is an indicator for whether subfield j ever received a 1940–1957 returnee, and Post_{jt} equals one in years on or after the first 1940–1957 returnee's arrival in j. Subfields without a 1940–1957 arrival enter as never-treated controls. The pre-1940 returnee stock $\text{cumret}_{j,t}^{\text{pre-1940}}$ enters as a covariate X_{jt}:

$$y_{jt} = \alpha_j + \vartheta_t + \delta\,\text{Post}_{jt} \times \text{Returnee}_j + \gamma\,\text{cumret}_{j,t}^{\text{pre-1940}} + \varepsilon_{jt}, \tag{10}$$

where y_{jt} is the mean frontier alignment of subfield-j publications in year t, computed separately for OpenAlex and CNKI publications against the OpenAlex global frontier. Standard errors are clustered at the subfield level. As a continuous robustness specification, we replace the indicator with the cumulative count of 1940–1957 returnees who had published in j by year t, denoted $\text{cumret}_{j,t}^{\text{1940-57}}$:

$$y_{jt} = \alpha_j + \lambda_t + \gamma\,\text{cumret}_{j,t}^{\text{pre-1940}} + \delta\,\text{cumret}_{j,t}^{\text{1940-57}} + \varepsilon_{jt}, \tag{11}$$

with both regressors and outcome standardized so that δ is interpretable per standard deviation of cumulative-returnee exposure.

Table 5 reports the estimates. The binary specification of (10) produces a positive coefficient of 0.019 on the alignment of OpenAlex publications to the global frontier (column 1, $t = 1.40$) and a smaller, indistinguishable-from-zero coefficient of 0.011 on the alignment of CNKI publications (column 2). The continuous specification of (11) delivers similar magnitudes: a one-standard-deviation increase in cumulative 1940–1957 returnees raises OpenAlex-side alignment by 0.020 standard deviations (column 3) but does not move the CNKI-side alignment (column 4). The point estimates are economically modest at the aggregate panel level; restricting to subfields whose first 1940–1957 returnee actually arrived during the cohort window and to the event-time window $[-10, 15]$ around that arrival, the OpenAlex coefficient on $\text{cumret}_{j,t}^{\text{1940-57}}$ rises to 0.069 ($t = 1.68$, $p = 0.09$). The spillover therefore operates primarily through the international-publishing arm of returnees' own work and the work of their domestic collaborators who appear in OpenAlex; it does not show up in measurable changes to the global-frontier alignment of Chinese-language publications at the subfield level.

Table 5: Subfield-level spillover: 1940–1957 returnees and frontier alignment

	Eq. (10): Binary Post×Returnee		Eq. (11): Continuous (per SD)	
	(1)	(2)	(3)	(4)
	OA vs OA fr.	CNKI vs OA fr.	OA vs OA fr.	CNKI vs OA fr.
Treatment	0.019	0.011	0.020	−0.005
	(0.014)	(0.021)	(0.014)	(0.020)
Pre-1940 cumret control	Yes	Yes	Yes	Yes
Subfield fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Subfields	211	176	211	176
Observations	5,427	2,427	5,427	2,427

Notes: Subfield-year regressions. Columns 1–2 estimate equation (10): $\text{Returnee}_j = 1$ if subfield j ever received a 1940–1957 returnee, $\text{Post}_{jt} = 1$ in years on or after the first such arrival. Subfields with no 1940–1957 arrival serve as never-treated controls. Columns 3–4 estimate equation (11) with the cumulative count of 1940–1957 returnees in subfield j by year t, standardized; the outcome is also standardized so that the coefficient is per standard deviation. Columns 1 and 3 take the mean alignment of OpenAlex publications to the OpenAlex global frontier as the outcome; columns 2 and 4 take the mean alignment of CNKI publications to the same frontier. All specifications include subfield and year fixed effects and partial out the pre-1940 returnee stock. Standard errors clustered at the subfield level (columns 1–2) or analytic OLS (columns 3–4) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The aggregate spillover masks heterogeneity across subfields. Following the historical record on the Twelve-Year Plan for Science and Technology (1956), we isolate six strategic subfields explicitly prioritized by the Chinese government during the Great Return Wave: Aerospace Engineering, Nuclear and High Energy Physics, Atomic and Molecular Physics, Condensed Matter Physics, Electrical and Electronic Engineering, and Mechanical Engineering. Figure 8 traces mean frontier alignment by period (1940–1954, 1955–1964, 1965–1976) for each subfield, separately for OpenAlex and CNKI publications.

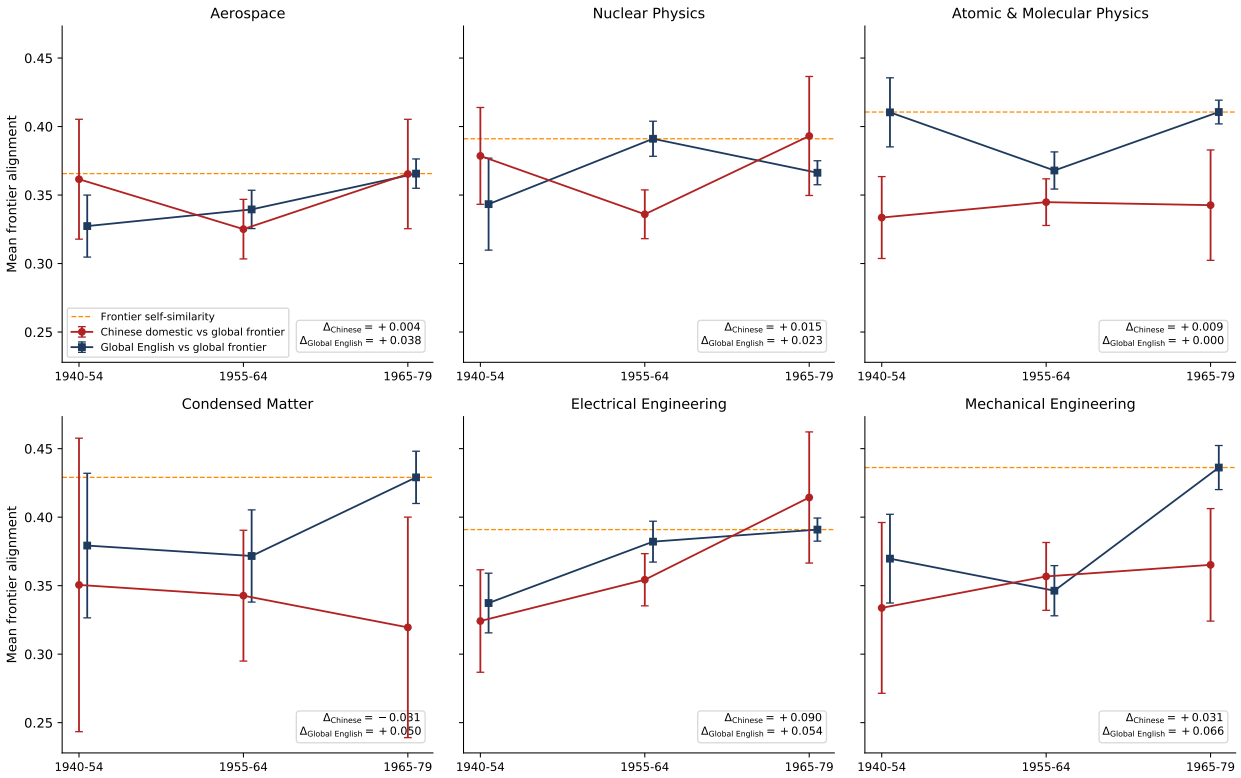


Figure 8: Convergence in direction across strategic subfields

Notes: Mean alignment of subfield-period publications to the OpenAlex global frontier. Red: CNKI publications. Blue: OpenAlex publications. Dashed orange: maximum OpenAlex alignment across periods (frontier self-similarity benchmark). Δ_{CN} and Δ_{OA} are the corresponding changes between 1940–1954 and 1965–1976.

The OpenAlex-side alignment rises in all six subfields between 1940–1954 and 1965–1976, with gains ranging from near zero in Atomic and Molecular Physics to +0.066 in Mechanical Engineering. The CNKI-side rises in five of six subfields, with the largest gain in Electrical and Electronic Engineering ($\Delta_{\text{CN}} = +0.090$); only Condensed Matter Physics shows a small decline. The two engineering subfields (Electrical, Mechanical) account for the bulk of the catching-up signal in CNKI-side publishing, consistent with the planning regime concentrating returnee placement and domestic-publishing infrastructure in fields directly relevant to industrialization.

To visualize where in the semantic space of science Chinese domestic output concentrated relative to the international frontier, we project paper-level LaBSE embeddings to two dimensions using t-SNE and bin them into a hexagonal density grid. Figure 9 shows four panels: panel (a) maps the global English-language frontier; panels (b)–(d) map the corresponding Chinese-language output for the three sub-periods 1940–1954, 1955–1964, and 1965–1976.

The Chinese mass migrates over time toward the regions of the embedding space that are densest at the frontier, with the largest visible shift between 1955–1964 and 1965–1976 in the engineering and applied-physics regions of the map.

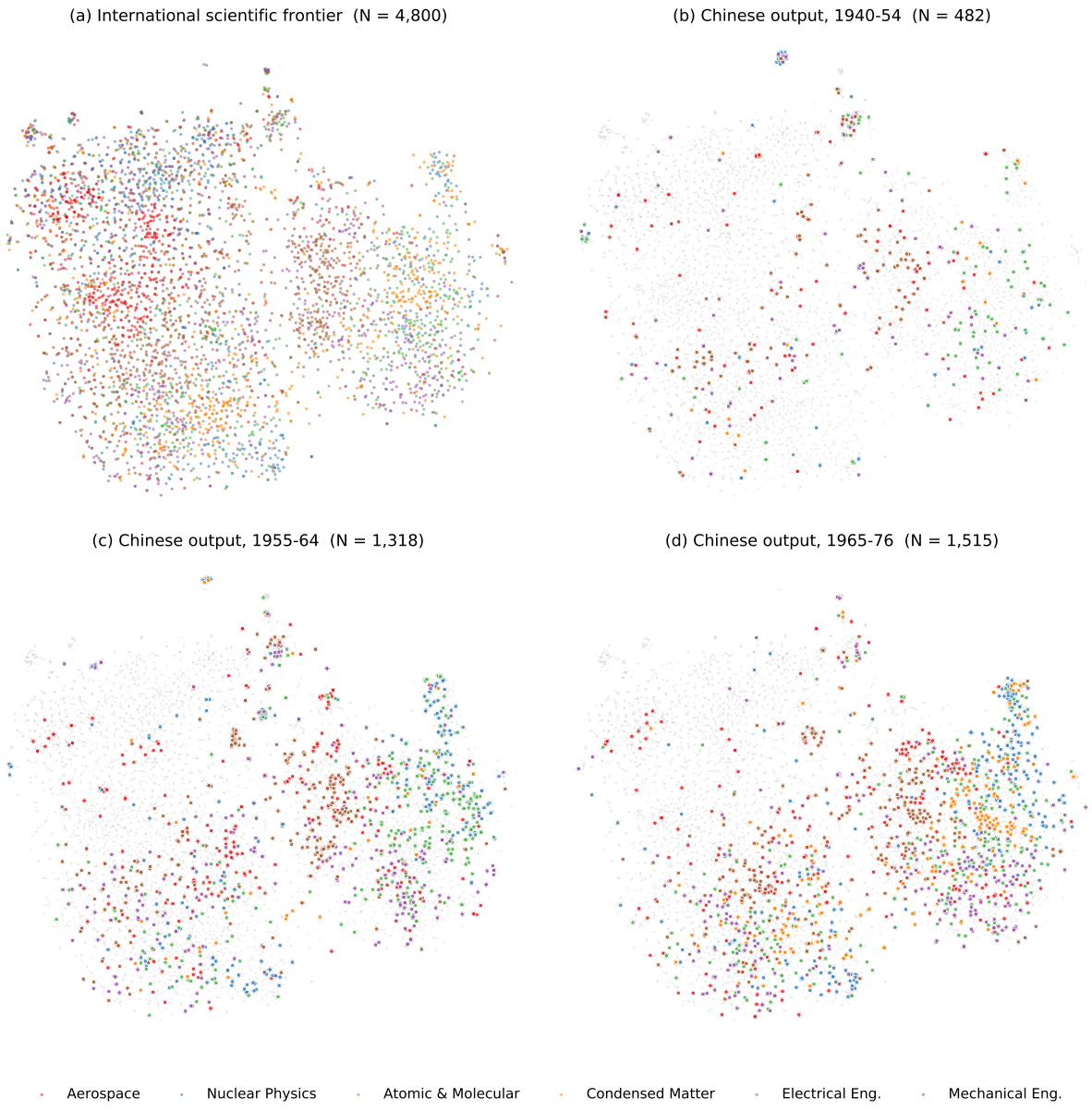


Figure 9: Volume of scientific output in the semantic space of strategic subfields

Notes: Two-dimensional t-SNE projection of LaBSE paper embeddings restricted to the six strategic subfields. Panel (a): the global English-language frontier (top-cited OpenAlex papers, all years). Panels (b)–(d): hexagonal density of Chinese-language CNKI publications during 1940–1954, 1955–1964, and 1965–1976. Color intensity in each panel reflects the local density of papers in that two-dimensional region. The Chinese-output mass converges over time toward the densest regions of the international frontier, particularly between 1955–1964 and 1965–1976 in the engineering and applied-physics regions of the map.

7 Conclusion

The Great Return Wave of 1940–1957 brought more than 1,300 U.S.-trained Chinese scientists back to a country in the middle of building a planned research system. We track their publications across the international (OpenAlex) and domestic (CNKI) bibliographies through 1976 and measure each paper's alignment to the international research frontier using multilingual semantic embeddings. The cross-corpus, person-level design lets us decompose effects on scientific output that single-corpus analyses cannot separate.

Three findings stand out. First, return reallocates publishing channels rather than reducing total scholarly output: returnees publish 4.5 percent less in OpenAlex and 6.2 percent more in CNKI, leaving the corpus sum slightly positive. Citations fall by 15.8 percent on the international side and rise by 6.8 percent on the domestic side, for a combined decline of 10.3 percent. Second, the reallocation does not lower paper-level alignment to the international frontier on either side, so the channel switch does not come at a measurable quality cost. Third, the spillover beyond the individual scientist is concentrated where it would be expected under a directed research regime. Subfields that received a 1940–1957 returnee experience a positive post-arrival shift in international-frontier alignment, and the effect is most pronounced in the engineering subfields explicitly named in the Twelve-Year Plan for Science and Technology of 1956.

These results speak to a longstanding question in the economic history of socialist science: whether centrally planned research systems can absorb foreign frontier knowledge, and where in the field-space the absorption occurs. Soviet-era emulation strategies described by Lardy (1995) and the science-policy template the early PRC inherited from them (Wang, 2015) predicted convergence in disciplines tied to industrial and military priorities, and convergence stalls at the periphery. Our subfield estimates match that prediction: the catching-up signal loads on engineering and applied physics, not on basic physics or the humanities. The pattern is consistent with a broader literature documenting how political institutions shape the direction, rather than only the level, of innovation (Acemoglu et al., 2021; Moscona, 2024; Moser et al., 2025), and complements work tracing the long-run effects of mass cross-border knowledge transfers in market economies (Moser et al., 2014; Giorcelli, 2019). Returnees in our setting were not a marginal input that the planner could deploy uniformly; they were the binding constraint on absorptive capacity in the priority subfields, and the data trace exactly that pattern.

The empirical strategy used here is portable, and three extensions are natural. The first is to follow the 1978 reopening cohort, where the same individuals are observed in a market environment and the channel-substitution mechanism should weaken if the planner's

preferences were what concentrated catch-up in engineering. The second is to apply the dual-frontier measure to other socialist science systems with bilingual bibliographies (the post-1949 USSR, the GDR, post-1949 Cuba), where the predicted heterogeneity across subfields differs by planning regime. The third is to map the coauthor network that returnees built around themselves. The cross-corpus panel records every domestic collaborator on every CNKI paper and every international collaborator on every OpenAlex paper, so each returnee appears simultaneously as the seed of a Chinese coauthor cluster and as a persisting node in the U.S. network they trained in. Tracing how those two networks evolve, whether the domestic cluster widens as it absorbs the returnee, whether the international tie persists or atrophies under closure, and whether bridging scientists drive the subfield-level catching-up we document, is the natural next step, and the embedding-based outcomes used here generalize directly to coauthor-pair similarity. Whether returnees pull a domestic system toward the frontier, and where in the field-space they do so, appears to depend less on the scientists themselves than on the institutional environment that decides where to put them and on the networks that form around them once they arrive.

References

Acemoglu, Daron, David Y Yang, and Jie Zhou, "Political Pressure and the Direction of Research: Evidence from Chinas Academia," *NBER Working Paper*, 2021, (29402).

_ , David Y. Yang, and Jie Zhou, "Power and the Direction of Research: Evidence from China's Academia," 2024. Draft manuscript.

Aghion, Philippe and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–351.

Bello, Piera, Alessandra Casarico, and Debora Nozza, "Research Similarity and Women in Academia," *The Economic Journal*, 2025, p. ueaf113.

Bianchi, Nicola and Michela Giorcelli, "Scientific education and innovation: from technical diplomas to university STEM degrees," *Journal of the European Economic Association*, 2020, *18* (5), 2608–2646.

Borjas, George J and Kirk B Doran, "The collapse of the Soviet Union and the productivity of American mathematicians," *The Quarterly Journal of Economics*, 2012, *127* (3), 1143–1203.

China Institute in America, *A Survey of Chinese Students in American Universities and Colleges in the Past One Hundred Years: In Commemoration of the One Hundredth Anniversary of the Graduation of the First Chinese from an American University, Yung Wing, B.A., Yale 1854*, New York: China Institute in America, 1954. Preliminary report. Jointly sponsored by the National Tsing Hua University Research Fellowship Fund and the China Institute in America.

Docquier, Frédéric and Hillel Rapoport, "Globalization, brain drain, and development," *Journal of economic literature*, 2012, *50* (3), 681–730.

Giorcelli, Michela, "The long-term effects of management and technology transfers," *American Economic Review*, 2019, *109* (1), 121–152.

_ and Bo Li, "Technology transfer and early industrial development: evidence from the Sino-Soviet alliance," Technical Report, National Bureau of Economic Research 2021.

Han, Yelong, "An untold story: American policy toward Chinese students in the United States, 1949—1955," *Journal of American-East Asian Relations*, 1993, pp. 77–99.

Iaria, Alessandro, Carlo Schwarz, and Fabian Waldinger, "Frontier Knowledge and Scientific Production: Evidence from the Collapse of International Science," *Quarterly Journal of Economics*, 2018, *133* (2), 927–991.

Jia, Ruixue, Huihua Nie, and Wei Xiao, "Power and publications in Chinese academia," *Journal of Comparative Economics*, 2019, *47* (4), 792–805.

_ , Margaret E Roberts, Ye Wang, and Eddie Yang, "The impact of US–China tensions on US science: Evidence from the NIH investigations," *Proceedings of the National Academy of Sciences*, 2024, *121* (19), e2301436121.

Kelly, Bryan, Dimitris Papanikolaou, Amit Seru, and Matt Taddy, "Measuring Technological Innovation over the Long Run," *American Economic Review: Insights*, 2021, *3* (3), 303–320.

Kerr, Sari Pekkala, William Kerr, Çağlar Özden, and Christopher Parsons, "Global talent flows," *Journal of Economic Perspectives*, 2016, *30* (4), 83–106.

Koffi, Marlene, "Innovative Ideas and Gender (In) Equality," *American Economic Review*, 2025, *115* (7), 2207–2236.

Kung, S. W., *Chinese in American Life: Some Aspects of Their History, Status, Problems, and Contributions*, Westport, CT: Greenwood Press, 1973. Reprint of the edition originally published by University of Washington Press, Seattle.

Lardy, Nicholas R., *Emulating the Soviet Model: Economic Development in China, 1949–1957*, Cambridge, UK: Cambridge University Press, 1995.

Li, Tao, *A Historical Record of Overseas Education in China since 1949*, Beijing: Higher Education Press, 2000.

li Yuan, Tung, *A Guide to Doctoral Dissertations by Chinese Students in America, 1950–1960*, New York: Chinese-American Council for Education, 1961.

Mokyr, Joel, "The intellectual origins of modern economic growth," *The Journal of Economic History*, 2005, *65* (2), 285–351.

_ , "Long-term economic growth and the history of technology," in "Handbook of economic growth," Vol. 1, Elsevier, 2005, pp. 1113–1180.

Moscona, Jacob, "Flowers of Invention," 2024. Draft manuscript.

Moser, Petra, "Patents and Innovation in Economic History," *Annual Review of Economics*, 2016, *8* (Volume 8, 2016), 241–258.

_ , Alessandra Voena, and Fabian Waldinger, "German Jewish Émigrés and U.S. Invention," *American Economic Review*, 2014, *104* (10), 3222–3255.

_ , Sahar Parsa, and Shmuel San, "Immigration and Innovation: Lessons from the Quota Acts," 2025. Working paper.

Romer, Paul M, "Endogenous technological change," *Journal of political Economy*, 1990, *98* (5, Part 2), S71–S102.

_ , "The origins of endogenous growth," *Journal of Economic perspectives*, 1994, *8* (1), 3–22.

sing Kung, James Kai and Alina Wang, "Overseas students, ideology, and the fall of dynastic China," *Journal of Economic Growth*, 2025, pp. 1–40.

Spilimbergo, Antonio, "Democracy and foreign education," *American economic review*, 2009, *99* (1), 528–543.

Squicciarini, Mara P and Nico Voigtländer, "Human capital and industrialization: Evidence from the age of enlightenment," *The Quarterly Journal of Economics*, 2015, *130* (4), 1825–1883.

Waldinger, Fabian, "Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany," *Journal of political economy*, 2010, *118* (4), 787–831.

_ , "Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany," *The review of economic studies*, 2012, *79* (2), 838–861.

_ , "Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge," *Review of Economics and Statistics*, 2016, *98* (5), 811–831.

Wang, Zuoyue, "The Cold War and the Reshaping of Transnational Science in China," in Naomi Oreskes and John Krige, eds., *Science and Technology in the Global Cold War*, Cambridge, MA: MIT Press, 2015, pp. 343–369.

Zhang, Jun and Yueyun Mo, "The Historical Evolution of China's Overseas Students' Return Policies since the Founding of the People's Republic of China," *Party History Research and Teaching*, 2018, (5), 60–69.

Appendix

1 Data Sources

Directories of Chinese Scholars in the United States

Across the early twentieth century, different organizations voluntarily took on the responsibility of creating a record of Chinese students at American universities. These organizations oftentimes used these directories for the two-pronged purpose of fostering community ties among Chinese students and amplify marketing efforts. As the responsibility for creating directories for Chinese students was informally passed from organization to organization, the methods of collection began to change as well. Due to the lack of a centralized database, statistics on Chinese students were often a collation of information gathered from different methods, including outreach to universities, personal interviews, and records from local student clubs. The prevalence of name changes and ambiguity around classifications of origin made it difficult to maintain consistent records over the years.

Table 6: **Chinese Student Directories in the United States (1911–1952)**

Organization	**Years Covered**	**Variables Available**	**Data Collection Methods**
Chinese Students' Alliance in the U.S.A.	1911–12, 1914–15, 1921, 1929–33	Name, School, Major/Field of Study, Degree Pursued, Mailing Address	Compiled and edited by a handbook committee and local club officers, with support from alumni and professional societies.
Chinese Students' Christian Association	1935–40	Name, School, Major/Field of Study, Mailing Address	Aggregated registrar records, correspondence with individual students, collaboration with student groups, institutional catalogs, and bulletin mailings.
The China Institute in America	1941–45, 1949–52	Name, School, Degree Program, Graduation Year, Student Status (undergrad/grad), Major Field	Data obtained from schools (winter semester returns), Chinese student groups, and independent graduate surveys. Limitations noted: incomplete due to mid-year enrollment changes, name inconsistencies, and exclusions based on origin.

The first step of our data collection focused on the retrieval of all directories of Chinese students in the United States, which served as a time stamp of their activity abroad. We then constructed a dataset of all scholars by concatenating the directories.

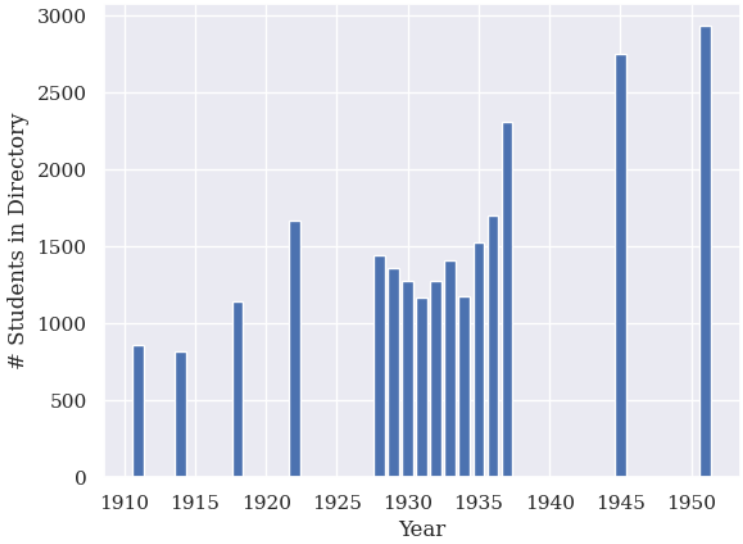


Figure 10: Annual Count of Students in Directory

year	No. of Entries	Truncated Names
1911	861	387
1914	821	593
1929	1357	586
1930	1281	375
1931	1168	433
1932	1281	374
1933	1413	420
1935	1702	275
1937	2312	682
1945	2750	1078
1951	2934	292

Table 7: Count of Students by Directory

Below are a few challenges that we experienced in trying to merge data across disparate sources:

- truncated or anglicized names (more common for first names), which created significant misalignment between the dataset of returnees and the dataset of Chinese scholars in the US
- heterogeneity in academic schema, making it difficult to group scholars under field of study
- unclear timelines across directories; although we could not find overlapping directories across organizations, we also did not find any formal documents corroborating our claim that each organization operated dis-jointly
- multi-year gaps in directories left huge holes in our data, making it difficult to corroborate timelines of certain scholars whose time abroad fell within one of these periods

Returnees

We compile a panel dataset of Chinese scholars who studied abroad and returned to China using three primary archival sources covering 1840–1963. The Comprehensive Survey of Scientific and Technical Personnel Studying Abroad and Returning to China, 1840–1949 profiles 3,500 scientific and technical personnel, documenting pre-departure domestic education, foreign study details, and post-return employment; the data, originally structured in tabular form by field, were digitized and cleaned for misspellings and syntactical inconsistencies. The Scientific and Engineering Manpower in Communist China, 1949–1963 contains information on 1,200 prominent scientists and engineers, including birth year, educational background, specialization, political status, and Chinese Academy of Sciences (CAS) membership, compiled from a wide range of official and published materials. Finally, the Dictionary of Chinese Scholars Who Returned from North America in the 1950s covers 1,806 returnees, with detailed variables on education, migration dates, post-return careers, political roles, and notable life events, based on interviews, organizational records, personal communications, and archival materials. For the Dictionary, we parsed archival records using GPT-4 API calls to classify variables related to time abroad and post-return activities. Table 8 summarizes the scope, variable coverage, and data collection methods for each source.

Table 8: **Primary Sources on Returnee Scholars to China (1840–1963)**

Source	**Time Period Covered**	**Variables Available**	**Data Collection Methods**
A Comprehensive Survey of Scientific and Technical Personnel Studying Abroad and Returning to China, 1840–1949	1840–1949	Name, Domestic School (before going abroad), Year of Departure, Country of Study Abroad, Foreign Institution, Academic Discipline, Major, Year of Return, Domestic Work Location (after return)	Profiles 3,500 scientific and technical personnel in mainland China that had studied abroad. Compiled from biographical books, publications, and archival materials on students who studied abroad, particularly those in science and engineering.
Scientific and Engineering Manpower in Communist China, 1949–1963	1949–1963	Name, Year of Birth, Educational Background (highest degree held), Academic Position (as of 1964), Special Field (basic training), Social and Political Status (as of 1964), CAS Membership Indicator	Profiles 1,200 prominent scientists and engineers in mainland China. Data compiled from Chinese official documents and books, periodicals and newspapers, statistical yearbooks, scientific journals, leadership speeches, and translated materials.
Dictionary of Chinese Scholars Who Returned from North America in the 1950s	1950s	Name, Year of Birth, Year of Death, Domestic Education (Affiliation, Degree, Year, Location), Foreign Education (U.S. Affiliation, Degree, Year, Location), Date of Arrival in the U.S., Date of Return to China, Post-return Employment and Destination, Employment in China before Studying Abroad, Academic Discipline, Academic Specialization, Political Roles, CAS Academician Status, Whether Detained, Whether Died During Cultural Revolution	Profiles 1,806 returnees. Based on five types of sources: (1) Interviews, (2) information from acquaintances, (3) records from organizations, (4) public materials (e.g., memoirs), (5) archival letters/photos.

Publication Data

To measure scholar productivity and characterize research specialization, we assemble a comprehensive publication panel by merging English- and Chinese-language bibliometric sources. Our construction proceeds in two steps: (i) for each scholar, we recover a broad candidate set of publications using multiple name variants; (ii) we remove false matches using a text-based classifier and structured metadata checks, yielding a clean scholar–year corpus of titles and abstracts.

Publications. For each scholar, we collect publications from OpenAlex and CNKI. From OpenAlex we retain `work_id`, title, abstract, publication year, venue (journal/proceedings), document type, authorship roster (including `author_id` and author order), author affiliations (institution identifiers and countries), concepts/keywords (with hierarchical levels), `referenced_works` (outgoing citations), and `cited_by_count`. From CNKI we add Chinese-language works with titles, abstracts, venues, publication years, and author affiliations. Table 9 summarizes the core fields retained from OpenAlex.

Table 9: Core Publication Fields Retained

Entity	**Fields**
Work	id, title, abstract, year, venue, type, concepts (id, level), referenced_works, cited_by_count
Authorship	work_id, author_id, author_position, affiliation_id, affiliation_country
Author	author_id, canonical name, aliases, ORCID (if any)
Institution	inst_id, name, country, historical name variants

Name Variants and Candidate Retrieval. Scholars frequently published under multiple romanization systems. To maximize recall, we query OpenAlex under three naming conventions: Chinese characters (when available), Hanyu Pinyin, and Wade–Giles. We store all returned candidates and retain an `author_alias` table mapping each observed name variant to a canonical scholar identifier. Table 10 reports the distribution of match counts by naming convention.

Table 10: OpenAlex Match Counts by Naming Convention

Format	**0 matches**	**1 match**	**>1 matches**
Chinese characters	77	26	47
Pinyin	27	20	97
Wade–Giles	113	12	31

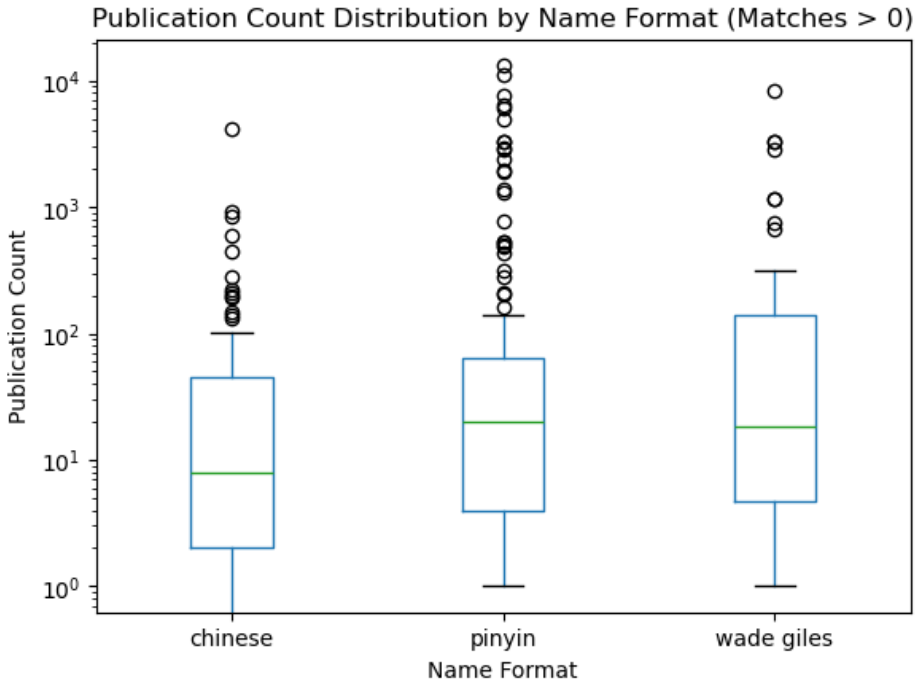


Figure 11: Publication Counts for Matched Scholars

Filtering False Matches. High recall necessarily produces false positives, particularly for common surnames and short romanized names. We therefore filter candidate publications using a supervised text-based classifier. Specifically, we use a BERT-style model to score whether a candidate work plausibly belongs to a given scholar based on the title/abstract embedding, coauthor overlap, affiliation consistency, field proximity, and temporal plausibility. We complement the classifier with rule-based exclusions (e.g., incompatible affiliations, implausible publication years) and manual adjudication for high-stakes cases (e.g., CAS fellows, department chairs). The resulting output is a clean scholar–year corpus of publications with harmonized metadata across OpenAlex and CNKI.

Dissertations. We complement publication data with a historical record of dissertations completed by Chinese scholars in the United States (Yuan, 1961. We cross-reference dissertation authors to our scholar roster and use dissertation fields, institutions, and years as additional anchors for identity verification and cohort construction.

Author Disambiguation and Matching. Our disambiguation protocol combines (i) rule-based romanization normalization (segmenting given/family names; removing hyphens/tones; harmonizing common variants), (ii) network evidence from coauthor and affiliation co-occurrence, and (iii) manual adjudication for ambiguous cases. We audit matching accuracy through pairwise spot checks, with targeted review of the most prolific scholars and the most common names.

Citation Construction. To avoid snapshot bias in citation counts, we construct annual incoming citations by scanning `referenced_works` in later publications and assigning citations to the year they occur. We then compute forward citation windows (e.g., five-year citations) and field-normalize citation outcomes by publication year and field.

Field Definition Sensitivity. We assess robustness to alternative field constructions by varying the number of clusters k (e.g., 300–800), swapping the text representation (BERT embeddings versus LSA or Word2Vec), and re-estimating field assignments. We report field stability metrics (e.g., adjusted Rand index across specifications) and show that the main results are robust across alternative fieldings.

Venue Lists and Top-Venue Share. We construct field-specific top-venue lists (journals and conference proceedings) and allow venue rankings to evolve over time. We report robustness to alternative top-N thresholds and to excluding venues with changing scope or editorial standards.

2 U.S. Exit Restrictions on Chinese Students during the Early Cold War

This subsection documents the policy rationale and implementation of U.S. exit restrictions on Chinese students during the early Cold War. Drawing on declassified CIA, State Department, and FBI materials, we show that U.S. authorities treated the overseas training of Chinese students, particularly in technical and scientific fields, as a potential national security risk and acted deliberately to restrict their return to the People's Republic of China (PRC).

Fields Deemed Strategically Sensitive. Across intelligence and diplomatic documents, a consistent set of disciplines was identified as posing heightened risk if U.S.-trained Chinese students returned to China. These fields were closely linked to weapons development, military-industrial capacity, and advanced infrastructure.

Table 11: Fields Targeted by U.S. Security Concerns

Category	**Fields Mentioned or Inferred**
Nuclear Science	Nuclear physics, nuclear engineering
Weapons and Missile Technology	Aeronautics, aerospace engineering, guidance systems
Advanced Engineering	Mechanical, electrical, civil engineering
Electronics and Computing	Electronics, communications, radar, control systems
Chemical and Biological Sciences	Chemistry, chemical engineering, biomedicine
Metallurgy and Materials	Metallurgy, materials science
Mathematics and Physics	Mathematics, applied physics
Geophysics	Geophysics (e.g., military mapping and exploration)

Sources: CIA, *Science and Technology in Communist China* (1966); National Intelligence Estimate 13-2-62, *Chinese Communist Advanced Weapons Capabilities* (1962); State Department and FBI memoranda.

Government Actions. Declassified State Department and FBI documents from the early 1950s show that U.S. authorities deliberately detained or blocked the departure of technically trained Chinese students. These actions were motivated by three main considerations:

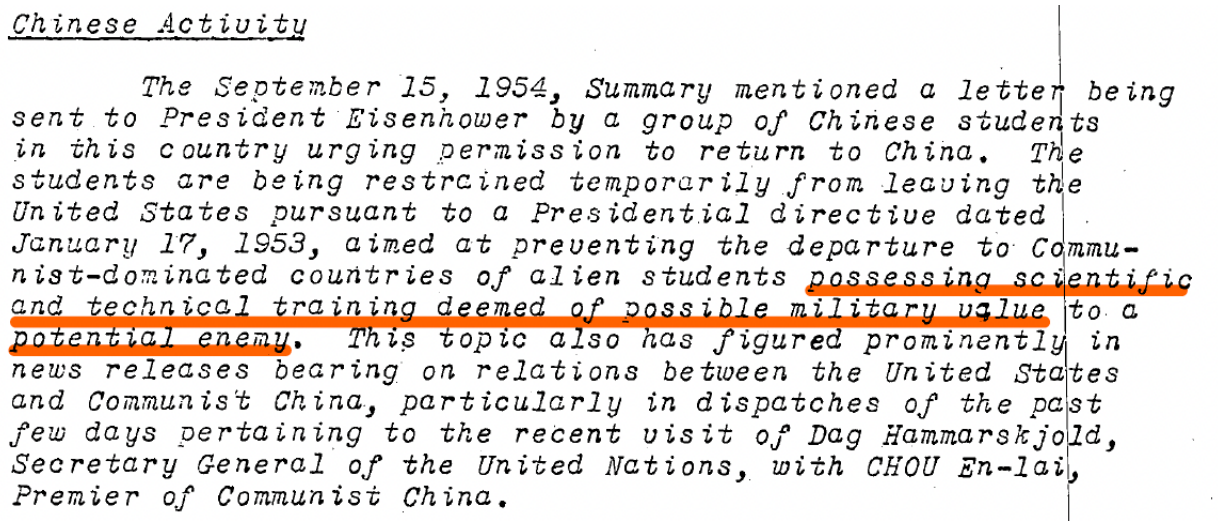
- preventing the transfer of U.S.-acquired technical knowledge to China's weapons and industrial programs;

- using detained students as leverage in negotiations over American prisoners held in China;
- responding to intelligence assessments warning that returnees trained in sensitive disciplines posed a national security risk.

Surveillance and Enforcement Criteria. Students were subject to heightened monitoring and scrutiny if they:

- pursued advanced training in strategically sensitive fields;
- were suspected of affiliation with the Chinese Communist Party or participation in pro-PRC student organizations;
- expressed intentions or made observable preparations to return to mainland China;
- accessed or conducted research linked to U.S. military or classified projects.

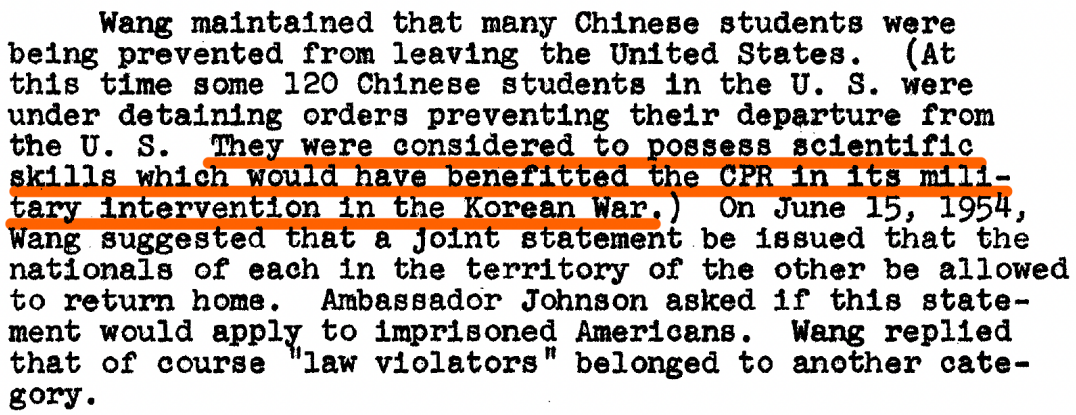
Strategic Motivation. The overarching objective of these restrictions was to limit the repatriation of U.S.-acquired technical expertise that could accelerate China's military and industrial modernization. U.S. authorities explicitly linked student return to concerns over nuclear weapons development, missile technology, and long-run strategic competition during the early Cold War.



Chinese Activity

The September 15, 1954, Summary mentioned a letter being sent to President Eisenhower by a group of Chinese students in this country urging permission to return to China. The students are being restrained temporarily from leaving the United States pursuant to a Presidential directive dated January 17, 1953, aimed at preventing the departure to Communist-dominated countries of alien students possessing scientific and technical training deemed of possible military value to a potential enemy. This topic also has figured prominently in news releases bearing on relations between the United States and Communist China, particularly in dispatches of the past few days pertaining to the recent visit of Dag Hammarskjold, Secretary General of the United Nations, with CHOU En-lai, Premier of Communist China.

Figure 12: FBI memorandum expressing concern over Chinese students "possessing scientific and technical training deemed of possible military value to a potential enemy." J. Edgar Hoover to Harold E. Stassen, January 19, 1955.



Wang maintained that many Chinese students were being prevented from leaving the United States. (At this time some 120 Chinese students in the U. S. were under detaining orders preventing their departure from the U. S. They were considered to possess scientific skills which would have benefitted the CPR in its military intervention in the Korean War.) On June 15, 1954, Wang suggested that a joint statement be issued that the nationals of each in the territory of the other be allowed to return home. Ambassador Johnson asked if this statement would apply to imprisoned Americans. Wang replied that of course "law violators" belonged to another category.

Figure 13: State Department analysis of negotiations with Chinese representatives, June 1954–September 1955, concerning the detention and release of Chinese nationals in the United States.

Implications for Identification. This historical record underscores that U.S. exit restrictions were (i) sharply targeted toward specific technical fields, (ii) implemented through administrative discretion rather than individual behavior, and (iii) motivated by geopolitical considerations external to scholars' productivity or preferences. These features support our use of permit and denial decisions as plausibly exogenous variation in return migration and motivate field-based heterogeneity analyses in the main text.